

Antti Ryhänen

**DATA-ANALYTIIKAN HYÖDYNTÄMI-
NEN URHEILUJOUKKUEEN SUORITUS-
KYVYN KEHITTÄMISESSÄ**

Tekniikan ja luonnontieteiden tiedekunta
Kandidaatintyö
Elokuu 2020

TIIVISTELMÄ

Antti Ryhänen: Data-analytiikan hyödyntäminen urheilujoukkueen suorituskyvyn kehittämisessä

Kandidaatin työ

Tampereen yliopisto

Tietojohtaminen

Elokuu 2020

Tässä kandidaatintyössä tutkittiin data-analytiikan hyödyntämistä joukkueurheilussa. Tutkimuksen ongelmana oli löytää keinoja urheilutapahtumista kerättävän datan hyödyntämiseen joukkueurheilussa niin, että siitä saadaan jalostettua joukkueelle lisää pelillistä suorituskykyä. Tutkimuksessa keskityttiin kolmeen pääasiaan, jotka olivat datan kerääminen, datan jalostaminen informaatioksi ja informaation hyödyntäminen päätöksenteossa.

Tämä kandidaatintyö on toteutettu kirjallisuuskatsauksena. Tutkimuksessa on hyödynnetty aiemmin julkaistua pääosin vertaisarvioitua aineistoa, sekä muuta aiheeseen liittyvää kirjallisuutta.

Aihealuetta lähestyttiin urheilujoukkueen tiedontarpeiden selvittämällä, jonka jälkeen käytiin läpi, millaista dataa on saatavilla, miten dataa voidaan kerätä ja mitä haasteita keräämisessä on. Jalostamisvaiheessa tarkasteltiin aluksi laajemmin urheilun data-analytiikan kokonaisuutta ja tämän jälkeen pohdittiin datan jalostamisessa käytettävää tekniikkaa. Informaation hyödyntämisluvussa tutkittiin, miten jalostettu informaatio saadaan muunnettua suorituskyvyksi urheilutuottelussa. Data-analytiikan hyödyntäminen urheilussa kasvaa ja kehittyy jatkuvasti, joten tutkimus aihealueesta on hyvin merkityksellistä.

Tutkimuksessa havaittiin, että dataa on saatavilla urheilusta valtavia määriä ja datan tehokkaaseen hyödyntämiseen vaaditaan automatisoituja, suurien ja nopeasti liikkuvien datavirtojen käsittelyyn soveltuvia järjestelmiä. Dataa on myös saatavilla useista eri lähteistä, jolloin sen käsittely vaikeutuu ennestään. Lisäksi tutkimuksessa havaittiin, että urheilussa käytettävä data-analytiikka tulee kehittymään vielä valtavasti tulevaisuudessa tarjoten myös datatieteille loistavan alustan toimia ja kehittyä.

Avainsanat: data-analytiikka, tiedon louhinta, massadata, esineiden internet, tietojärjestelmät, joukkueurheilu, suorituskyky

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Tausta.....	1
1.2 Tutkimusongelma.....	3
1.3 Työn rakenne	4
2. TUTKIMUKSEN TOTEUTUS	5
2.1 Tutkimusmenetelmä.....	5
2.2 Tutkimusaineisto	6
3. JOUKKUEURHEILUDATA.....	8
3.1 Tiedon tarpeet joukkueurheilussa.....	8
3.2 Datan saatavuus	9
3.3 Datan kerääminen.....	11
3.4 Datan keräämisen haasteet	12
4. DATAN JALOSTAMINEN JOUKKUEURHEILUSSA	14
4.1 Data-analytiikka joukkueurheilussa	14
4.2 Datan esikäsittely.....	16
4.3 Tiedon louhinta	17
4.4 Datan visualisointi	19
5. SUORITUSKYKYÄ INFORMAATIOSTA	21
5.1 Joukkueen suorituskykyyn vaikuttavat tekijät	21
5.2 Päätöksenteko harjoittelussa	22
5.3 Päätöksenteko pelitilanteissa	23
5.4 Modernit analytiikkajärjestelmät joukkueurheilussa	24
6. JOHTOPÄÄTÖKSET	25
6.1 Yhteenveto.....	25
6.2 Tulosten arviointi ja jatkotutkimusmahdollisuudet.....	26
LÄHTEET	28

1. JOHDANTO

Tässä tutkimuksessa tarkastellaan, miten data-analytiikkaa voidaan hyödyntää joukkueurheilussa, niin että sen avulla voidaan saada urheilujoukkueelle lisää pelillistä suorituskyykyä suhteessa muihin sarjan joukkueisiin. Tutkimuksessa joukkueen suorituskyvyllä tarkoitetaan sitä, kuinka hyvin joukkue pystyy suoriutumaan urheilutapahtumassa ja saavuttamaan asetetut tavoitteet. Urheilujoukkueen voidaan sanoa suoriutuneen hyvin, jos se pystyy voittamaan ottelun. Urheilutapahtumalla tutkimuksessa viitataan otteluun, jossa pelataan vastustajajoukkuetta vastaan tai joukkueen sisäiseen harjoitukseen.

Johdantoluvussa tarkastellaan tutkimuksen taustoja ja tutkimusongelmaa päätutkimuskysymyksen ja alatutkimuskysymysten avulla. Lisäksi luvussa käydään läpi lyhyesti tutkimuksen rakennetta.

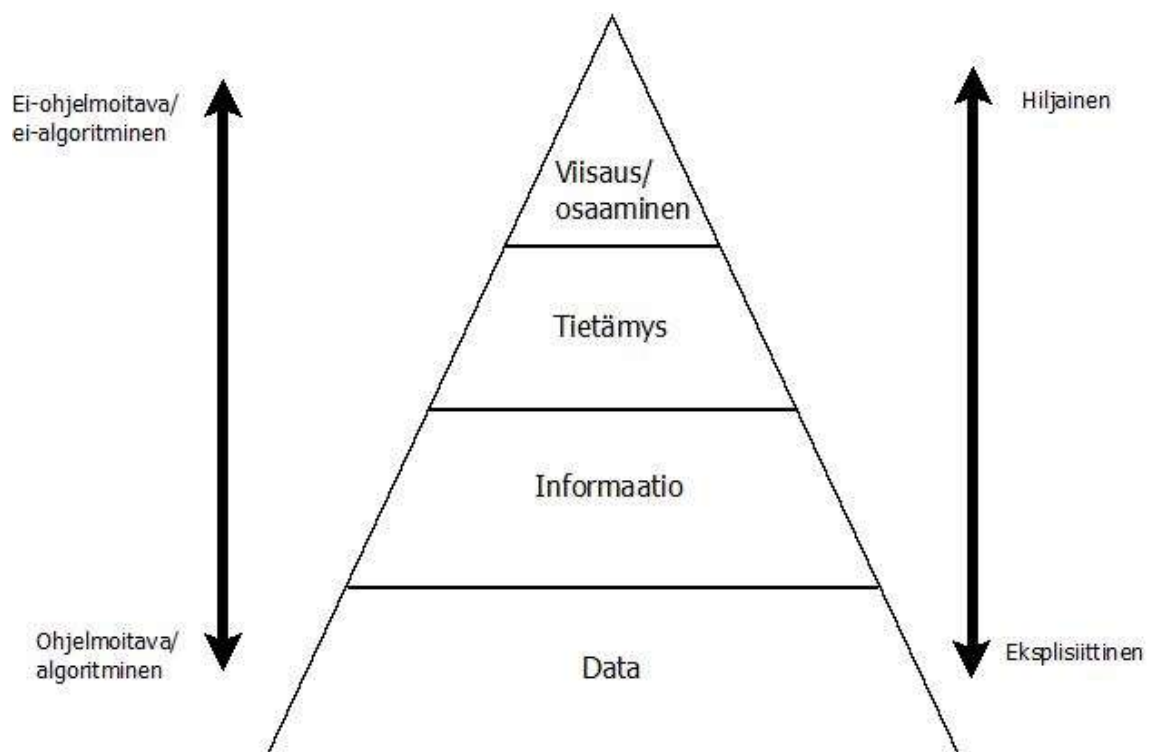
1.1 Tausta

Ihmiset näkevät ja kokevat jokapäiväisessä elämässään jatkuvasti urheilua eri muodoissa. Kilpaurheilua voidaan seurata televisiosta tai paikan päällä urheilutapahtumissa. Lisäksi media tuottaa jatkuvasti uutta sisältöä eri urheilulajeista ja -tapahtumista ympäri maailmaa. Urheilua voidaankin kuvata alana, jossa työn suorittaminen ja suorittamisen tulos ovat kaikkien saatavilla ja näin ollen yleistä tietoa (Miller 2015). Mikäli tunnettu urheilija tai urheilujoukkue on suoriutunut urheilutapahtumassa erityisen hyvin tai huonosti, ihmiset saavat modernissa maailmassa median tai muiden lähteiden kautta tiedon tapahtuneesta yleensä kohtalaisen nopeasti.

Globaalisti urheilualan markkina-arvo oli vuonna 2018 noin \$488,5 miljardia ja sen uskotaan kasvavan vuoteen 2022 mennessä noin \$614,1 miljardiin (ResearchAndMarkets 2019). Urheilu on siis kohtalaisen nopeasti kasvava ala, joka tarjoaa jatkuvasti uusia liiketoimintamahdollisuuksia. Kilpaurheilujoukkueiden tulot koostuvat pääosin medialta, liiketoimintakumppaneilta, katsojilta, sponsoreilta, mainostajilta, sekä TV-yhtiöiltä (Gratton 2012; Miller 2015). Tulorakenteesta voidaan päätellä, että menestyvässä urheiluliiketoiminnassa olennaista on viihdearvo, jonka urheilija tai joukkue pystyy sidosryhmilleen tarjoamaan. Miller (2015) mainitsee kirjassaan, että joukkue hyötyy voittamisesta enemmän kuin häviämisestä, mutta mikäli joukkue voittaa jatkuvasti kaikki ottelunsa,

viihdearvo voi alkaa laskemaan. Varsinkin PohjoisAmerikassa suurimmilla joukkueurheilusarjoilla on käytössä esimerkiksi palkkakatto, mikä mahdollistaa sen, että yksittäiset joukkueet eivät voi olla materiaailtaan yliveraisia. Tällä voidaan mahdollistaa viihdearvon parempi säilyminen otteluissa.

Urheilussa käytettävä analytiikka perustuu yleisesti hyvin paljon numeroihin, joiden avulla voidaan kuvata urheilijoiden tai joukkueiden suorituskykyä tilastoissa (Barlow 2015; Link 2018). Joukkueurheilussa ottelun lopputulos voidaankin kuvata numeroiden avulla, mutta numeroiden avulla voidaan kuvata myös muita ottelun sisällä tapahtuvaa, kuten esimerkiksi joukkueen pelivälineen hallussapitoaikaa tai pelaajien tekemiä rikkeitä. Näitä tilastoja analysoimalla voidaankin kuvata hyvin joukkueen tai joukkueen yksittäisen pelaajan suorituskykyä ja onnistumista urheiluottelussa. Tässä tutkimuksessa aihetta tarkastellaan datan prosessoinnin avulla, eli miten datasta saadaan jalostettua informaatiota, tietämystä ja osaamista. Alla olevassa kuvassa 1 on esitelty tiedon tasot.



Kuva 1: Tiedon tasot. (Mukailtu Laihonen et al. (2013); Rowley (2007))

Kuvassa 1 on esitelty tiedon tasot, jotka pätevät myös urheilussa käytettävään analytiikkaan. Datalla tarkoitetaan ei rakenteellisia asioita, kuten urheilussa numeroita tilastoissa. Informaatio on puolestaan enemmän rakenteellista, joka muodostuu dataa jalostamalla, kuten urheilussa ottelukoosteet tai tilastoista jalostettua dataa. Tietämys alkaa

olemaan enemmän inhimillistä ja se perustuu yleensä kokemukseen, urheilussa tietämystä voidaan oppia esimerkiksi informatiivisten suoritusohjeiden tai esimerkin kautta. Viisautena tai osaamisena voidaan pitää hyvin sisäistettynä tietona, joka urheilussa voidaan saavuttaa toistojen avulla ja sitä voidaan hyödyntää otteluissa suorituskykyä nostavasti. (Laihonen et al. 2013; Rowley 2007.)

1.2 Tutkimusongelma

Tutkimuksen tarkoituksena on löytää joukkueurheilussa käytettäviä työkaluja ja menetelmiä, joiden avulla kilpaurheilujoukkue pystyy hyödyntämään urheilutapahtumista saatavilla olevaa dataa parantaakseen omaa suorituskykyään, sekä menestymään joukkueena paremmin. Näin ollen tutkimuksen näkökulma painottuu enimmäkseen tekniselle osa-alueelle, vaikka urheilu onkin alana hyvin inhimillinen. Joukkueen pelillisen suorituskyvyn voidaan nähdä vaikuttavan huomattavasti suhteisiin organisaation taloudellisesti tärkeiden sidosryhmien, kuten sponsoreiden-, televisioyhtiöiden- ja fanien kanssa (Carvalho et al. 2019). Tästä voidaan päätellä, että kilpaurheilujoukkueen pelillinen suorituskyky on hyvin olennainen osa urheiluorganisaation menestymistä liiketoiminnassaan. Tutkimusongelma on esitettyä päätutkimuskysymyksen muodossa alla:

- Miten urheilutapahtumista kerättyä dataa voidaan hyödyntää niin, että sen avulla saadaan tuotettua joukkueelle lisää pelillistä suorituskykyä?

Päätutkimuskysymyksen tueksi tutkimuksessa on laadittu muutamia alatutkimuskysymyksiä, jotka jakavat aihealueen pienempiin osa-alueisiin. Alatutkimuskysymysten tarkoituksena on helpottaa ja kohdentaa tiedonhakua, sekä auttaa aikataulutuksessa. Alatutkimuskysymyksistä voidaan huomata, että kysymykset noudattavat hyvin pitkälti perinteisiä tiedonhallinnan malleja lähtien tiedontarpeiden tunnistamisesta ja päättyen päätöksentekijöiden ratkaisuihin. Tutkimuksen alatutkimuskysymykset ovat listattuna alla:

- Millaista urheiludataa joukkueissa tarvitaan?
- Millaista dataa joukkueurheilutapahtumista on saatavilla?
- Miten ja millaisilla työkaluilla urheiludataa voidaan kerätä joukkueurheilutapahtumista?
- Miten kerättyä urheiludataa voidaan jalostaa?
- Miten informaatiota voidaan hyödyntää joukkueen suorituskyvyn kehittämisessä?

Tutkimuksessa käsitellään historiallisen ja reaaliaikaisen datan hyödyntämistä päätöksenteossa. Tutkimuksen näkökulma on tekninen ja tutkimuksessa keskitytään erilaisiin keinoihin kerätä ja jalostaa dataa. Lisäksi tutkimuksessa tarkastellaan, miten jalostettua dataa voidaan hyödyntää joukkueen suorituskyvyn kehittämisessä. Aihealueesta on suljettu pois useita oleellisesti toimintaan vaikuttavia ulkoisia sidosryhmiä, kuten fanit ja katsojat, media, viranomaiset, sponsorit ja TV-yhtiöt.

1.3 Työn rakenne

Työn rakenne noudattaa tiedonhallinnan prosessin vaiheita, joissa kuvan 1 mukaisesti datasta jalostetaan tietoa ja osaamista, jonka avulla urheilujoukkue pystyy saavuttamaan pelillistä suorituskykyä. Luvussa 2 käydään läpi tutkimuksen toteutusta, pitäen sisällään alaotsikot tutkimusmenetelmän- ja tutkimusaineiston esittelyille. Luvussa 3 tarkastellaan joukkueurheilusta kerättävää dataa. Luku pitää sisällään tiedontarpeiden määrittelyn, tarjolla olevan datan saatavuuden, sen keräämisen ja keräämiseen liittyviä haasteita. Kuvan 1 ”tiedon tasot” pyramidin mukaisesti luvussa 3 käsitellään pyramidin alinta tasoa.

Luvussa 4 tutkitaan, miten kerätystä datasta voidaan joukkueurheilussa jalostaa informaatiota, jota voidaan hyödyntää päätöksenteon tukena. Aliluvussa 4.1 tarkastellaan urheilussa käytettävän data-analytiikan prosessia yleisesti, jonka jälkeen siirrytään tarkastelemaan datan jalostamista tarkemmin esikäsittelyn, tiedon louhinnan ja datan visualisoinnin avulla. Viidennessä luvussa tarkastellaan urheilujoukkueen päätöksentekoprosessia, niissä tilanteissa, kun päätöksen tukena käytetään data-analytiikkaa. Tutkimuksen viimeisessä luvussa, eli luvussa 6 vedetään yhteen tutkimuksen tulokset, sekä pohditaan jatkotutkimusmahdollisuuksia.

2. TUTKIMUKSEN TOTEUTUS

Tässä luvussa esitellään tutkimuksen toteutusta. Luku pitää sisällään alaluvut tutkimusmenetelmän esittelylle ja käytetyn tutkimusaineiston esittelylle.

2.1 Tutkimusmenetelmä

Tämä tutkimus on toteutettu kirjallisuuskatsauksena, tietolähteinä on käytetty pääosin vertaisarvioituja lähteitä, kuten tieteellisiä artikkeleita, kirjoja tai tutkimuksia. Basole & Saupe (2016) mukaan data-analytiikan hyödyntäminen urheilussa lisääntyy jatkuvasti maailmalla ja teknologiat muuttuvat jatkuvasti. Tästä syystä aineiston valinnassa on otettu erityisesti huomioon myös aineiston julkaisuajankohta.

Tiedonhaussa käytetyin työkalu on ollut Tampereen yliopiston tarjoama palvelu Andor, joka näyttää aineiston saatavuuden eri tietokannoissa. Andor:in pohjalta tietoa on haettu myös useista muista tietokannoista, kuten esimerkiksi ProQuest ja Springer Link. Näiden rinnalla on hyödynnetty myös googlen tarjoamaa Google Scholar tietokantapalvelua, mutta sen kautta pääsy osaan aineistosta on rajattua. Tietokantahakujen lisäksi aineistoa on kerätty myös aikaisemmissa tutkimuksissa käytettyjä lähteitä tutkimalla.

Tietokannoista tietoa on haettu yhdistelemällä erilaisia yksittäisiä hakusanoja, joista käytetyimpiä ovat olleet seuraavat: "Data analytics", "Sports analytics", "Data mining", "Sports data", "Data visualization", "big data" ja "internet of things". Yksittäisillä hakusanoilla tietoa hakiessa tietokannat antavat tuloksia paljon, joten hakulausekkeita on luotu yhdistelemällä hakusanoja ja yleensä lisäämällä hakulausekkeeseen "sports". Hakutuloksia on rajattu myös esimerkiksi julkaisupäivän ja saatavuuden mukaan.

Taulukko 1: Hakutulokset, andor

Hakunimike tai -lauseke	Hakutulokset (kpl)	Vertaisarvioidut lehdet (kpl)	Mainittu otsikossa (kpl)
"Data analytics"	428 840	39 887	21 881
"Sports analytics"	2 981	439	369
"Sports data"	6 648	1 848	306
"Data analytics" AND Sports	17 983	2 417	48
"Data mining" AND Sports	15 976	6 685	90
"Sports analytics" AND "big data"	333	87	7

Taulukossa 1 on esiteltyä andor:ista muutamilla eri hakusanoilla ja -lausekkeilla tulneiden hakutuloksien määriä. Kaikki taulukossa näkyvät hakutulokset ovat rajattuja niin, että ne ovat saatavilla verkossa. Vaikka taulukon luvut näyttävät kohtalaisen suurilta, tuloksista vain pieni osa on soveltuvia tutkimuksessa käytettäväksi. Valitut lähteet ovat pääosin vertaisarvioituja artikkeleja tai aihealueeseen soveltuvia kirjoja. Päätökset aineiston soveltuvuudesta on tehty pintapuolisella tarkastelulla ja pikalukemisen avulla. Näin tekstin asianmukaisuudesta on saatu kohtalaisen tarkka kuva. Alaluvussa 2.2 esitellään muutamia oleellisia lähteitä, joita on käytetty tutkimuksen tukena.

2.2 Tutkimusaineisto

Tutkimusaineistoa läpikäydessä lähteisiin tarkempi tutustuminen on suoritettu pikalukemisena, tekstin silmäilyä ja mahdollisen sisällysluettelon, tiivistelmän, sekä yhteenvedon tarkasteluna. Näiden avulla aineistosta saa hyvän kokonaiskuvan kohtalaisen nopeasti. Aineistoa luettaessa on hyödynnetty Microsoft OneNote sovellusta muistiinpanojen tekemisessä. Tutkimusta tehdessä aineistoa on arvioitu kriittisesti verraten sitä muuhun aiheita käsittelevään kirjallisuuteen. Alla on esiteltyä lyhyesti muutamia tutkimuksen kannalta olennaisia lähteitä.

Miller (2015) esittelee kirjassaan "Sports Analytics and Data Science: Winning the Game with Methods and Models" data-analytiikan työkaluja voittavan urheilujoukkueen rakentamiseen. Kirja antaa hyvän kokonaiskuvan datatieteen hyödyntämisestä urheilussa, mutta kirjan loppupään luvut, eli varsinkin luvut 11, 12 ja osio A (Data Science

Methods) ovat olleet erityisen hyödyllisiä tutkimusta tehdessä. Hyödyllisiä näistä luvuista tekee niiden soveltuvuus aihealueeseen ja hyvät esimerkit.

Brefeld et al. (2018) avaavat hyvin esimerkein varustellussa teoksessaan "Machine Learning and Data Mining for Sports Analytics" hupputason jalkapallossa, -jäähkiekossa, -koripallossa, -baseballissa ja -Amerikkalaisessa jalkapallossa käytettäviä koneoppimisen ja tiedon louhinnan menetelmiä. Teoksen avulla saa myös hyvän kuvan siitä, millaista tiedon visualisointia näissä huippu-urheilusarjoissa hyödynnetään.

Ofoghi et al. (2013) esittävät tekstissään "Data Mining in Elite Sports: A Review and a Framework" kattavasti huippu-urheilussa käytettäviä tiedon louhinnan menetelmiä. Tekstissä käydään läpi data-analytiikan tavoitteita ja tiedon louhinnan työkaluja, kuten klusterointia, luokittelua, suhteiden mallinnusta ja assosiaatiosääntöjä hyvin esimerkein.

Azar et al. (2018) kirjoittavat tekstissään "Sports analytics and the big-data era" urheilussa käytettävän analytiikan toiminnallisuuksista massadatan aikakaudella. Tekstissä esitellään myös tutkimuksessa hyödynnetty urheiluanalytiikan viitekehys.

Yllä olevien lähteiden lisäksi tutkimuksessa on hyödynnetty myös muuta aineistoa tiettyjen asiakokonaisuuksien hahmottamiseksi ja aineistojen vertailun mahdollistamiseksi. Urheilua käsittelevien lähteiden lisäksi tutkimuksessa on hyödynnetty myös ei urheiluun liittyvää aineistoa data-analytiikan ja muiden aiheeseen liittyvien käsitteiden selvittämiseen.

3. JOUKKUEURHEILUDATA

Tieto voidaan määritellä eri tasoihin esimerkiksi sen tyyhin mukaisesti ja perinteisesti jako voidaan tehdä datan, informaation ja tietämyksen välillä. Kun tietoa jalostetaan datasta informaatioksi ja tiedoksi, sen arvo kasvaa. (Rowley 2007.) Modernissa maailmassa ja myös urheilussa dataa on aina enemmän kuin ehditään analysoida, mutta dataa on kuitenkin helppo kerätä moderneilla työkaluilla ja sitä on edullista säilyttää (Miller 2015). Tästä syystä automatisoidut datan keräämis- ja analysointijärjestelmät ovat ja tulevat jatkossakin olemaan erittäin hyödyllisiä.

Tässä luvussa tarkastellaan, miten tiedon alinta tasoa, eli dataa saadaan kerättyä joukkueurheilusta ja mitä tekijöitä tulee ottaa huomioon datan keräämisessä. Dataa kerätessä on olennaista ottaa huomioon, millaista dataa tarvitaan ja millaista dataa on saatavilla. Tästä syystä alaluvussa 3.1 käsitellään joukkueurheilun tiedontarpeita ja alaluvussa 3.2 käsitellään urheiludatan saatavuutta. Varsinaista datan keräämistä käsitellään alaluvussa 3.3 ja tämän jälkeen tarkastellaan vielä datan keräämisen haasteita alaluvussa 3.4.

3.1 Tiedon tarpeet joukkueurheilussa

Kilpaurheilujoukkueet hyödyntävät dataa otteluiden analysoinnissa, pelistrategian- ja harjoittelun suunnittelemisessa, loukkaantumisten ehkäisemisessä, suorituskyvyn hallinnassa ja ennustamisessa, sekä joukkueen rakentamisessa (Brefeld et al. 2019; Link 2018). Kuten perinteisemmissäkin tiedonhallinnan prosesseissa, tiedon tarpeiden tunnistaminen on myös urheilussa ja varsinkin kilpaurheilussa olennaista, koska dataa on saatavilla valtavia määriä. Schumaker et al. (2010) jakaa tiedontarpeet urheilussa viiteen portaittaiseen luokkaan, joista ensimmäisessä urheiludataa ei juurikaan hyödynnetä, toisessa asiantuntijat tekevät ennustavia päätöksiä tilannekohtaisesti inhimillisen ajattelun perusteella. Kolmannella tasolla asiantuntijat tekevät ennustavia päätöksiä perustuen aikaisempaan näyttöön, eli historialliseen dataan. Neljännessä vaiheessa hyödynnetään laajemmin tilastoja päätöksentekoprosessissa ja viidennessä vaiheessa hyödynnetään tiedon louhinnan (engl. data mining) työkaluja päätöksentekoprosessin apuna. (Schumaker et al. 2010.) Modernia data-analytiikkaa hyödynnettäessä päätöksenteko on pääosin kolmannessa, neljännessä tai viidennessä luokassa, sillä niissä suorituskyvyn kehittämiseksi hyödynnetään urheilutapahtumista kerättyä tai kerättävää dataa.

Tiedontarpeita asettavia päätöksentekijöitä joukkueurheilussa on esimerkiksi organisaation johto, valmennus ja joukkueen pelaajat (Azar et al. 2018). Kuten normaalimmasakin liike-elämässä, urheiluorganisaation organisaation ylin johto on vastuussa strategisesta, eli pitkän aikavälin päätöksenteosta, kuten valmentajien ja pelaajien hankinnasta. Valmennus puolestaan valitsee pelaavat pelaajat, on vastuussa harjoittelusta ja ottelukohtaisista strategioista. Yksittäiset urheilijat puolestaan tekevät päätöksiä oman suorituskykynsä kehittämistä yhdessä valmennuksen kanssa. (Bar-Eli et al. 2011.) Joukkueurheilussa suorituskyvyn kannalta olennaisia taktisen ja operatiivisen tason päätöksiä tekevät siis pääosin joukkueen valmennus ja yksittäiset pelaajat.

Urheilun suorituskykyä tarkastelevassa analytiikassa olennainen tarkastelun kohde on KPI-mittarit (Key Performance Indicators), jotka korreloivat hyvin paljon ottelun lopputuloksen kanssa, mutta eivät usein näy kovin selvästi urheiluotteluita satunnaisesti seuraaville katsojille (Alamar 2013; Kniffin et al. 2017). Englanninkielisen termin mukaisesti KPI tarkoittaa urheilussa yksittäisen urheilijan tai joukkueen tärkeimpiä suorituskykyyn vaikuttavia tekijöitä. Näitä tekijöitä on eri urheilulajeissa valtavasti ja niiden mittaamiseen tarvitaan usein paljon erilaisia työkaluja. Kretzer et al. (2015) listaavat tutkimuksessaan 8 urheilujoukkueen suorituskyvyn kannalta olennaista KPI-mittaria joukkueurheilussa, jotka ovat pelaajien väliset etäisyydet kentällä, eri pelipaikoilla pelaavien pelaajien suorituskykyvaatimukset, pelikentän eri osa-alueiden tapahtumat ja vaatimukset, pelaajien väliset optimaaliset kulmat eri pelitilanteissa, pelaajien varjossa olo aika, eli milloin pelaaja ei ole vapaana kentällä, pelivälineen hallussapitoaika, pelaajien nopeus eri tilanteissa ja pelaajien fyysiset ominaisuudet, kuten sydämen syketaaso tai laktaattitasot ottelussa.

3.2 Datan saatavuus

Urheilussa suorituskykyä kuvaavaa dataa voi syntyä urheilijoiden suorituksista ja ratkaisuksista, valmentajien tai johdon päätöksistä ja muista ottelun sisäisistä tapahtumista (Schumaker et al. 2010). Kerätty data voi olla määrällistä tai laadullista ja sitä on saatavilla monessa eri muodossa, kuten videotallenteina, numeroina tai kirjallisina raporteina (Azar et al. 2018). Urheilusta kerättyjä tilastoja voidaan pitää hyvin jäseneltynä datana, sillä ne lajitellaan usein tapahtuman kirjaamishetkellä, kuten esimerkiksi maalit tai rangaistukset. Videotallenteet ja muista multimediatyyppisistä lähteistä tuotettu data voidaan puolestaan luokitella jäsentelemättömäksi dataksi, koska sitä on haastava käsitellä perinteisin työkaluin. (Schumaker et al. 2010.) Luvussa 4 käydään läpi, miten kerättyä

dataa voidaan jalostaa ja analysoida niin, että sitä pystytään hyödyntämään tehokkaasti pyrkiessä tuottamaan joukkueelle kilpailukykyä paremman suorituskyvyn muodossa.

Maailmassa eri lähteistä tulevan tiedon määrä kasvaa jatkuvasti ja tämä pätee myös urheilussa. Urheilutapahtumasta voidaan kerätä hyvin monimuotoista dataa, kuten esimerkiksi pelin aikana syntyvät perinteiset suorituskykytilastot, jotka kertovat joukkueiden ja pelaajien tekemistä ratkaisuista kentällä (Alamar 2013). Varsinkin ammattiurheilussa jokainen pelikentällä tapahtuva suoritus näkyy jollain tavalla tilastoissa. Yksittäiset tilastot, kuten esimerkiksi pelivälineen hallinta-aika ei välttämättä yksinään kerro kovinkaan paljoa päätöksentekijöille, mutta niitä yhdistelemällä voidaan saavuttaa hyvinkin laadukasta aineistoa suorituskyvyn tarkkailuun. Alamar (2013) mainitsee myös, että nykyisin erittäin suosittu datan lähde on otteluista saatava videomateriaali, jonka avulla pystytään tunnistamaan pelaajien liikkeitä ja ottelun kokonaisuus hyvin tarkasti.

Videotallenteiden ja perinteisten ottelutilastojen lisäksi joukkueurheilussa hyödynnetään nykyisin IoT (Internet of Things) laitteita. IoT, eli suomeksi esineiden internet tarkoittaa dataa kerääviä laitteita, jotka toimivat yhdessä internettiin yhdistettynä muodostaen ikään kuin laitteiden verkoston (Hovi 2018; Sedkaoui 2018). Urheilussa usein käytettyjä IoT -laitteita ovat esimerkiksi sykettä, stressitasoja, veren laktaattipitoisuuksia tai muita kehon toimintaa mittaavia puettavia laitteita, kuten sykemittarit ja muut anturit (Miller 2015). Puettavia laitteita ei kuitenkaan ole sallittu vielä vuonna 2015 minkään urheilun pääsarjassa varsinaisten oikeiden pelien aikana (Barlow 2015). Puettavien IoT laitteiden lisäksi teknologiaa voidaan hyödyntää myös urheilukentälle asennettavissa sensoreissa tai pelivälineeseen lisätyssä sirussa. Gowda et al. (2018) esittelevät tekstissään iBall nimistä krikettipalloa, johon on asennettu siru, joka tuottaa jatkuvasti dataa pallon liikkeistä, kuten nopeudesta ja pyörimisestä. Samankaltainen ratkaisu on toteutettu myös jääkiekkoon Suomessa ja se on käytössä muutamissa jäähalleissa, joissa pelataan Suomen korkeinta jääkiekkosarjaa, eli Liigaa. Ratkaisussa kiekkoon on asennettu bluetooth teknologialla varustettu siru, mikä on yhteydessä jäähalliin asennettuihin sensoreihin ja tuottaa näin reaaliaikaista dataa analysoitavaksi. (Wisehockey Oy 2020.) IoT laitteiden tulo urheiluun tuo joukkueille valtavasti uusia mahdollisuuksia kehittää suorituskykyä entistä tehokkaammin. Data-analytiikan hyödyntämisen avulla harjoituksia voidaan optimoida pelitilanteiden mukaiseksi ja urheilijoiden suorituskykyä harjoittelussa voidaan tarkkailla niin, että siitä saatava hyöty olisi mahdollisimman optimaalinen.

Joukkueurheilutapahtumista voidaan siis kerätä sekä määrällistä, että laadullista dataa, joka voi olla laadultaan hyvin jäsenneiltyä tai jäsennelemätöntä. Dataa voidaan kerätä

ottelusta syntyvistä tilastoista ja raporteista, otteluiden videotallenteista tai puettavista, pelivälineeseen kiinnitetyistä ja urheiluareenoihin kiinnitetyistä IoT laitteista.

3.3 Datan kerääminen

Kuten aikaisemmin on mainittu, kerätty data voi olla jäsenneltyä tai jäsen telemätöntä. Jäsenneltyä helposti analysoitavaa dataa joukkueurheilusta saadaan esimerkiksi ottelun aikana kerätyistä tilastoista. Itse urheiluorganisaatioiden lisäksi ottelutilastoja keräävät myös monet ulkoiset toimijat, jotka luovat tilastoista analyyskejä ja jakavat dataa urheilu-seuroille ja -sarjoille (Schumaker et al. 2010). Perinteisten ottelutilastojen kerääminen on hyvin yksinkertaista ja kerättävät tilastot vaihtelevat paljon lajeittain. Tilastollista otteludataa kirjataan yleensä taulukkomuotoon numeerisina arvoina, jolloin niitä on helppo käsitellä.

Jäsen telemätöntä videodataa kerätään laajalti ympäri maailmaa parhaissa joukkueurheilusarjoissa. Videoista prosessoitavan datan määrä on valtava ja sitä tulee tarjota myös reaaliaikaisena päätöksentekijöille. Tällaista videodataa tuottavia järjestelmiä on olemassa paljon, mutta niiden ominaisuudet eivät eroa kovinkaan paljoa toisistaan. Esimerkiksi NBA koripalloliigassa on käytössä SportVU-järjestelmä, joka pystyy tallettamaan 25 datapistettä sekunnissa kentän jokaisesta pelaajasta ja pelivälineestä (Sellitto 2015). Toisena esimerkkinä voidaan nostaa jalkapallossa käytössä oleva TRACAB-järjestelmä, joka tuottaa kolmen tyyppistä dataa, välittömästi saatavilla olevaa reaaliaikaista dataa, 15 sekunnin viiveellä tulevaa koneellisesti prosessoitua hieman virheettömämpää dataa, sekä muutaman tunnin päästä kuvaamisesta tuotettua tarkkaa ja prosessoitua dataa (Linke et al. 2020).

Kuten aiemmin mainittiin, puettavia älykkäitä laitteita ei toistaiseksi ole täysin sallittu tunnetuimmissa joukkueurheilusarjoissa otteluiden aikana, mutta niistä saadaan valtavaa hyötyä harjoittelussa. Puettavien mittauslaitteiden kehitys on ollut erittäin nopeaa lähivuosina ja niiden suosio on jatkuvassa kasvussa urheilussa (Martin 2016). Dataa voidaan kerätä lähes jokaisesta urheilijan kehon osasta, kuten olkapäästä, kyynärpäästä, ranteesta, polvesta, sekä nilkasta ja näin ollen eri urheilulajien lajispesifejä liiketoja voidaan hahmottaa hyvinkin tarkasti (Ai et al. 2018). Ylivoimaisesti yleisin puettava laite on ranteessa pidettävä mittari, mutta monissa urheilulajeissa mittaamista on järkevämpää hyödyntää myös muissa kehon osissa, kuten esimerkiksi amerikkalaisessa jalkapallossa ja jääkiekossa olkapääsuojien alla (Aroganam et al. 2019). Vaikka tässä tutkimuksessa ei keskitytä varsinaisesti dataa keräävien laitteiden teknisiin ominaisuuksiin,

analysointia varten on hyvä tietää millä menetelmillä laitteet dataa keräävät. Puettavat laitteet voivat kerätä dataa esimerkiksi kiihtyvyyssantureiden, gyroskooppien, magnetometrien, GPS:än, sykemittareiden ja painesensoreiden avulla, joiden ohjelmistot on ohjelmoitu yleensä C, C++, Aduino, Java, Swift tai Python ohjelmointikielillä. Laitteet tuottavat esikäsiteltyä dataa älylaitteille, jotka ovat usein yhteydessä esimerkiksi pilvipalveluihin. (Aroganam et al. 2019.)

Ei puettavat, yleensä pelivälineisiin asennettavat älykkäät laitteet toimivat datan keräämisen kannalta samalla tavalla kuin useimmat puettavatkin laitteet. Laitteet pystyvät sensoreiden avulla havaitsemaan objektin nopeuden, pyörimisen, törmäykset ja lentoradan, sekä lähettämään dataa (Ebling 2016). Aiemmin mainittujen älykkäiden krikettipallo ja jääkiekko esimerkkien tekniikkaa voidaan hyödyntää esimerkiksi jalkapallossa ja golfpallossa, sekä erilaisissa peleissä käytettävissä mailoissa (Ebling 2016; Mendes Jr et al. 2016).

Videoista saatavan multimediadatan, sensoreiden, sekä älykkäiden puettavien- ja pelivälineisiin asennettujen laitteiden yhdessä tuottamaa dataa voidaan näyttää reaaliajassa suoraan älykkäiden laitteiden, kuten älypuhelimien, tablettien, kellojen ja tietokoneiden avulla. Data voidaan myös ladata suoraan pilvipalveluihin myöhempää käyttöä varten. (Barbosa 2018.)

3.4 Datan keräämisen haasteet

Kuten ylempänä on käyty läpi, erityyppisen datan lähteitä on joukkueurheilussa valtavasti. Datalähteiden määrä myös kasvaa jatkuvasti lisää, sekä myös näistä eri lähteistä saatavan datan määrä kasvaa (Alamar 2013). Osittain datan määrän takia joukkueurheilun päätöksentekijät, eli valmentajat, pelaajat, analyttikot ja joukkueen johto käyttää entistä vähemmän aikaa puhtaasti filosofisille argumenteille ja enemmän aikaa datan etsimiseen, sekä analysointiin (Barlow 2015). Yhdeksi haasteeksi datan keräämisessä joukkueurheilussa voidaan pitää päätöksentekijöiden analyysille jäävää aikaa.

Kun dataa on valtavasti, se liikkuu ja muuttuu nopeasti, sekä esiintyy vaihtelevassa muodossa, voidaan puhua massadatalle (engl. big data) ominaisista piirteistä (Hovi 2018; Sedkaoui 2018). Myös Miller (2015) mainitsee, että monien lähteiden ja laitteiden yhdistelemisestä aiheutuva datavirta aiheuttaa ammattuurheilujoukkueille ongelmia pysyvä teknologian kehityksen mukana. Näin ollen toisena datan keräämisen haasteena joukkueurheilussa voidaankin pitää päätöksenteolle olennaisen datan keräämistä ja louhimista valtavista datamassoista.

Vaikka kilpaurheilu pyörii nykyisin hyvin pitkälti datan ja analytiikan ympärillä, se on kuitenkin edelleen hyvin inhimillinen ala, jossa ihmiset tekevät lopullisia päätöksiä ja välillä päätökset on tehtävä hyvinkin nopeasti (Azar et al. 2018). Baerg (2017) kertoo tekstissään, että osalla urheilijoista ei ole tarpeeksi tietoa, miten dataa voidaan tehokkaasti hyödyntää harjoittelun ja varsinkin tunnetuimmilla urheilijoilla saattaa olla ympärillä jopa liikaakin ihmisiä analysoimassa ja ohjeistamassa mitä saa tehdä ja mitä ei. Tämä tuo esille ongelman siitä, että kaikilla saman lajin urheilijoilla ei ole tasavertaisia mahdollisuuksia hyödyntää urheilutapahtumista saatavaa dataa. Lisäksi koska urheiluorganisaatiot saavat pääosin suurimmat tuottonsa fanien ja median kautta, tällöin myös urheilijoiden elämäntavalliset tiedot ovat usein näille hyvin avointa (Miller 2015). Osittain tämän takia urheilijat ovat huolissaan omasta yksityisyyden suojastaan, sillä ihmiset saattavat median kautta saada tietoa urheilijoiden elämäntavoista, harjoittelusta, rituaaleista, ja sijainnista (Baerg 2017).

Joukkueurheiludataa käsittelevän luvun perusteella monet joukkueurheiluorganisaatiot, sekä yksittäiset urheilijat saattavat joutua kärsimään tietoähkystä (engl. information overload) suorituskkyä kehittävässä harjoittelussaan. Yhtenä ratkaisuna ongelmaan on hyvin pitkälti automatisoidut data-analytiikkajärjestelmät, jotka tuottavat valmiiksi jäseneltyä informaatiota päätöksentekijöille reaaliajassa (Alamar 2013; Azar et al. 2018; Link 2018; Miller 2015). Näitä järjestelmiä käsitellään luvuissa 4 ja 5.

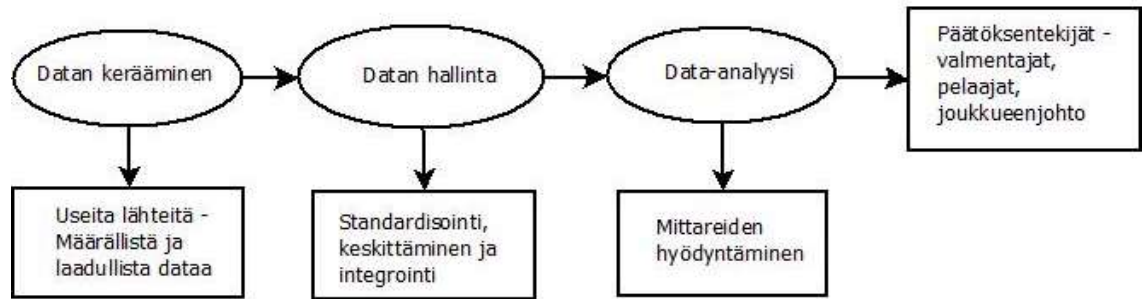
4. DATAN JALOSTAMINEN JOUKKUEURHEILUSSA

Luvussa 3 käytiin läpi, miksi joukkueurheilusta halutaan kerätä dataa, millaista dataa kerätään ja miten dataa kerätään. Luvussa 4 tarkastellaan tarkemmin, miten joukkueurheilusta kerättyä dataa voidaan jalostaa sellaiseen muotoon, että sitä voidaan hyödyntää tehokkaasti päätöksenteossa.

Alaluvussa 4.1 tarkastellaan yleisellä tasolla data-analytiikan prosessia joukkueurheilussa. Lopuissa alaluvuissa käydään läpi datan jalostamiselle olennaisia vaiheita kuten datan esikäsittelyä, tiedon louhintaa ja päätöksentekovaihetta helpottavaa datan visualisointia. Datan jalostamisen tarkoituksena on muuntaa ja prosessoida dataa helpommin ymmärrettävään muotoon, eli informaatioksi.

4.1 Data-analytiikka joukkueurheilussa

Analytiikan ja data-analytiikan termeillä viitataan alun perin erityyppisten datamallien hakemiseen, tulkitsemiseen ja viestimiseen (Davenport 2006). PohjoisAmerikassa urheilussa käytettävästä data-analytiikasta voidaan käyttää termiä urheiluanalytiikka. Alamar (2013) kuvailee urheiluanalytiikkaa jäsenneilyn historiallisen datan hallintana, jonka avulla voidaan luoda ennustavia- ja ohjailevia analyttisiä malleja. Urheiluanalytiikka on myös tietojärjestelmien ja urheiludatan hyödyntämistä päätöksenteossa, kun halutaan saavuttaa joukkueelle kilpailuetua urheilutapahtumassa pelikentällä. (Alamar 2013.) Azar et al. (2018) mainitsee tekstissään, että luotettavan ja systemaattisen analysoinnin avulla päätöksentekijät, eli valmentajat, urheilijat ja joukkueen johto voivat vahvistaa päätöksentekoprosessiaan. Kuvassa 2 on esiteltyä Azar et al. (2018) hahmottelema urheilun data-analytiikan viitekehys.



Kuva 2: Urheilussa käytettävän data-analytiikan viitekehys. (Azar et al. 2018)

Kuvassa 2 on esiteltyä Azar et al. (2018) muodostama yksinkertaistettu urheiluanalytiikan viitekehys, jota voidaan hyödyntää prosessin selkeyttämisessä. Viitekehys ei poikkea juurikaan perinteisimmistä liiketoiminnan analytiikan- ja tiedonhallinnan malleista, jotka lähtevät liikkeelle datan keräämisestä, sekä päättyvät päätöksentekoon ja tiedon hyödyntämiseen. Vertailukohteena voidaan pitää laajalti tunnettua Choon tiedonhallinnan prosessimallia, joka pitää sisällään vaiheet tiedontarpeiden tunnistamiselle, tiedon hankinnalle, tiedon jakelulle, tiedon käytölle ja toiminnan mukauttamiselle (Choo 2002). Urheiluanalytiikan viitekehyksessä datan kerääminen pitää sisällään urheilutapahtumissa tapahtuvan datan keräämisen eri vaiheet, joita käsiteltiin tarkemmin luvussa 3. Datan- tai tiedonhallinnalla puolestaan mahdollistetaan tiedon parempi saatavuus, sekä näin ollen pystytään säästämään päätöksentekijöiden aikaa (Alamar 2013). Hovi (2018) kuvailee tiedon hallintaa kattoterminä, joka pitää sisällään esimerkiksi käsitteiden mallintamista, tietovarastointia, tiedonhallintaan liittyviä turvallisuusasioita, tietoarkkitehtuuria, talletusratkaisuja ja tietojen laatuasioita. Urheiluanalytiikan osalta toimiva tiedonhallinta mahdollistaa siis urheilutapahtumista saadun datan tehokkaan uudelleenkäytön esimerkiksi joukkueen harjoituksissa. Malli voitaisiin kuvitella myös lähtevät päätöksentekijöistä, jotka asettavat tiedon tarpeet datan keräämiselle.

Miller (2015) avaa tekstissään, että urheiludataan liittyviä ongelmia ratkoessa parhaita työkaluja ohjelmointiin on R ja Python, joista R soveltuu erityisesti tilastollisiin ratkaisuihin ja Python tarjoaa hyödyllisiä kirjastoja esimerkiksi koneoppimiseen ja muihin tiedon louhinnan ratkaisuihin. Myös Harjamäki (2018) mainitsee TTY Porin selvityksessä, että yleisesti data-analytiikan sovelluksissa kaksi käytetyintä ohjelmointikieltä ovat juuri R ja Python. Kuten tekstissä on aiemmin mainittu, urheilussa käytettävän datan kattava analysointi tarvitsee pitkälle automatisoituja analyttisiä järjestelmiä, jotka hyödyntävät juuri tiedon louhintaa ja koneoppimista.

Ofoghi et al. (2013) listaa tekstissään neljä urheiludatan analyysissä olennaista tavoitetta, jotka ovat listattuna alla:

- Löytää suorituskymällejä, jotka kuvaavat sitä, miten urheilija tai joukkue voi lisätä mahdollisuuksiaan päättää suoritus tai ottelu jossain tietyssä asemassa tai tilanteessa. Esimerkiksi voittaa ottelu kahden maalin erolla.
- Ennustaa urheilijan tai joukkueen suorituksia pohjautuen aiemmin tiedettyihin harjoittelu- tai ottelusuorituksiin. Esimerkiksi mitä pelaajan tulee tehdä rangaistuslaukauksessa, -potkussa tai -heitossa, jotta onnistuminen olisi mahdollisimman todennäköistä.
- Mahdollistaa reaaliaikainen päätöksenteko siitä, mitä toimia/reaktioita tai taktiikoita tarvitaan sen hetkisen tai tulevan tapahtuman aikana. Esimerkiksi häviöllä olevan joukkueen viimeisten peliminuuttien ratkaisut.
- Tiettyjen urheilutapahtumien tai tilanteiden päävaatimusten tunnistaminen ja oikeiden pelaajien valitseminen juuri tähän tapahtumaan tai tilanteeseen. Esimerkiksi jääkiekossa yli- ja alivoimakentällisten valinta.

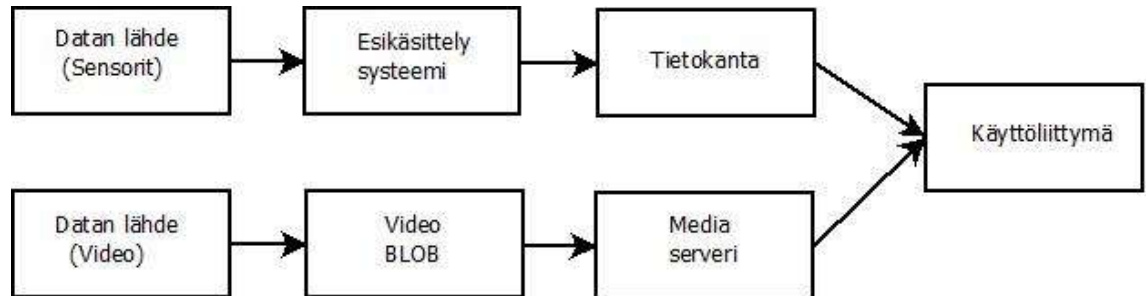
Jotta yllä oleviin tavoitteisiin voidaan päästä, kerättyä dataa pitää voida jalostaa päätöksenteon kannalta edullisempaan muotoon. Seuraavissa alaluvuissa on esiteltyinä kolme datan jalostamisen kannalta olennaista työvaihetta, jotka ovat datan esikäsittely, tiedon louhinta ja datan visualisointi.

4.2 Datan esikäsittely

Ennen kuin eri lähteistä kerättyä dataa voidaan alkaa analysoida, tulee sitä siistiä helpommin käsiteltävään muotoon. Urheilusta saatavan datan esikäsittelyssä tulee ottaa huomioon varsinkin datan valmistelu ja -lajittelu analyysiä varten (Ofoghi et al. 2013). Yleisellä tasolla datan siistiminen vaatii tiedon tarpeiden perusteella joitakin seuraavista kuudesta toimenpiteestä, vääränlaisen tai ei tarpeellisen datan siistiminen, muodon muokkaaminen tarvittavaksi, useista lähteistä kerätyn datan yhdistäminen, datan normalisointi yhtäläiseen muotoon, puuttuvien arvojen korvaaminen ja poikkeavan tai häiriintyneen (engl. noisy) datan tunnistaminen (García et al. 2015).

Otteluista kerätyn tilastollisen datan esikäsittely on kohtalaisen yksinkertaista, esimerkiksi tarkastelemalla, että kokonaisluvut ovat oikeassa muodossa, tai teksti, kuten pelaajien ja joukkueiden nimet ovat yhtäläisellä formaatilla kirjattuja. Videoista tai sensorien avulla kerätyn datan esikäsittely on huomattavasti haastavampaa ja usein tehokasta

vasta kuin prosessi on lähes täysin automatisoitua. Yleisesti huippujoukkueurheiluo-
nisaatiot ovatkin ulkoistaneet nämä prosessit. (Schumaker et al. 2010.) Kretzer et al.
(2015) ehdottavat modernille video- ja sensoridataa käsittelevälle järjestelmälle neljävai-
heista arkkitehtuuria, ehdotuksen tietovirtakaavio näkyy kuvassa 4.



Kuva 3: Ehdotettu sensoreista ja videolähteistä saadun datan käsittelyjärjestelmästä tietovirtakaaviona. (Kretzer et al. 2015)

Kaavion ensimmäisissä vaiheissa kerätään data pelaajiin ja/tai pelivälineisiin kiinnite-
tyistä sensoreista tai videolta. Toisessa vaiheessa esikäsittelykomponentti vastaanottaa
kaikki sensori- tai videotiedot, suodattaa ne ja tallentaa tietokantaan tai mediaserverille.
BLOB (engl. Binary Large Object) tarkoittaa suuria binäärisiä datasettejä, eli tässä ta-
pauksessa videodataa. Käsitelty data saadaan käyttöön käyttöliittymän avulla. (Kretzer
et al. 2015.)

4.3 Tiedon louhinta

Eri lähteistä kerättyä esikäsiteltyä dataa voidaan lähteä jalostamaan hyödyllisempään
muotoon erilaisilla tiedon louhinnan menetelmillä, jotka mahdollistavat erilaisten toistu-
vien mallien ja yhtäläisyyksien löytämisen datamassoista. Tiedon louhintaa voidaankin
kuvailla eräänlaisena massadata-analytiikkana tai sen olennaisena työkaluna. Tiedon
louhinta käsitettä käytetään varsinkin silloin, kun tavoitteena on muuntaa suuria tai jat-
kuvasti virtaavia datamassoja päätöksentekoa varten edullisempaan muotoon. (Azzalini
et al. 2012.) Tästä päätellen modernissa joukkueurheilussa tehokas data-analytiikan
hyödyntäminen vaatii tiedon louhinnan ratkaisuja. Luvun 1 kuvan 1 mukaisesti tiedon
louhinta keskittyy enimmäkseen tiedon jalostamiseen datasta informaatioksi (Ofoghi et
al. 2013). Olennainen osa tiedon louhintaa on koneoppiminen, mikä mahdollistaa suu-
rien datamassojen käsittelyn tehokkaasti, tiedon louhinnassa käytetty koneoppiminen voi
olla ohjattua (engl. Supervised learning), ohjaamatonta (engl. Unsupervised learning) tai
vahvistettua (engl. Reinforcement learning). Ohjatussa oppimisessa kone opetetaan op-

pimaan suhteita kohdemuuttujan ja muiden muuttujien välillä. Ohjaamattomassa oppimisessa kone oppii sille syötetystä datasta ilman ohjaamista tai tavoitteen asettamista ja vahvistetussa oppimisessa kone oppii pohjautuen sen ympäristöstä saatavasta palautteesta. (Dangeti 2017.)

Urheilussa käytetään paljon erilaisia tiedon louhinnan työkaluja erilaisten yhtäläisyyksien selvittämiseksi ja datan jalostamiseksi. Yleisimpiä metodeja datan käsittelyyn joukkueurheilussa on klusterointi (engl. Clustering), luokittelu (engl. Classification), suhteiden mallintaminen (engl. Relationship modeling) ja assosiaatiosäännöt (engl. Association rule mining) (Ofoghi et al. 2013). Metodit ovat esiteltynä lyhyesti alla hyödyntäen kuvitteellisia esimerkkejä.

Klusteroinnissa kone pystyy käsittelemään ryhmittelemätöntä dataa ja jakamaan datan automaattisesti ryppäisiin niiden eri ominaisuuksien mukaan. Klusterointi hyödyntää ohjaamatonta oppimista, joten se ei vaadi erikseen koneen opettamista. (Delen & Olson 2008.) Esimerkiksi joukkueen fyysisen harjoittelun suunnittelussa voidaan hyödyntää klusterointia lajitellessa urheilijat eri ryhmiin kuntoluokituksen mukaisesti.

Luokittelu on puolestaan ohjattua oppimista, jossa koneelle syötetään niin sanottu harjoittelu data, jonka pohjalta kone osaa automaattisesti ennustaa ja luokitella datan sen ominaisuuksien mukaan (Delen & Olson 2008). Luokittelussa tulokset tai luokat pyritään esittämään helppolukuisessa puukavion muodossa, joka auttaa päätöksentekijöitä keskittymään helposti tiettyyn pelaajaryhmään (Zhou 2012). Esimerkiksi kuinka suuri osa eurooppalaisessa jalkapallossa tietyltä sektorilta potkaistu potku menee maaliin (Brefeld et al. 2019).

Suhteiden mallinnuksessa tavoitteena on yleensä löytää toiminto tai malli, joka kuvaa parhaiten ennustettujen ja riippuvien ominaisuuksien välistä suhdetta. Joukkueurheilussa paljon käytettyjä suhteiden mallinnustapoja on neuroverkot tai syväoppiminen, jotka koostuvat useista itsesäätelevistä prosessointimenetelmistä, jotka toimivat yhdessä tiiviisti toisiinsa kytketyissä verkostoissa. Perinteisempänä suhteiden mallinnustyökaluna voidaan pitää lineaarista regressiota. (Ofoghi et al. 2013.) Esimerkiksi millä todennäköisyydellä päädytään haluttuun lopputulokseen, jos tehdään jokin asia oikein mutta ei pystytä tunnistamaan kaikkia ulkoisia vaikuttavia tekijöitä.

Assosiaatiosääntöjen avulla urheilussa voidaan liittää esimerkiksi olosuhteiden, liikkeiden, päätösten, tilanteiden tai tapahtumien vaikutuksen johonkin tiettyyn tilanteeseen tai tulokseen päätyemisessä (Ofoghi et al. 2013). Assosiaatiosääntöjen kanssa puhutaan usein ostoskorianalyysistä, mikä kuvaa asiakkaan todennäköisyyttä ostaa tuote Y, kun

hän on ostanut jo tuotteen X (García et al. 2015). Samalla tavalla, jos joukkue voittaa jonkun tietyn tilanteen pelikentällä, assosiaatiosääntöjen avulla voidaan tarkastella millä todennäköisyydellä se johtaa johonkin tiettyyn toiseen tilanteeseen.

4.4 Datan visualisointi

Kun datasta on löydetty toistuvia yhtäläisyyksiä, jotka johtavat johonkin haluttuun tai ei haluttuun lopputulokseen, sen visualisointi on hyvin olennaista päätöksentekijöille ongelman hahmottamiseksi. Visualisointi mahdollistaa tehokkaan datan ymmärtämisen ja siitä oppimisen (Miller 2015). Näin ollen tiedon tasoissa visualisoinnin avulla voidaan siirtyä informaation tasolta tietämyksen tasolle ja niin edelleen tietämyksen tasolta harjoittelun avulla osaamiseksi, joka puolestaan näkyy suorituskyvyn kehittymisenä. Tarkkaa datan visualisointia voidaan hyödyntää joukkueen taktisessa harjoittelussa tai yksittäisen urheilijan suorituskyyä tukevassa harjoittelussa, joko jälkeenpäin näytettynä tai reaaliajassa (Basole & Saupe 2016). Videotallenteilta saatu informaatio on suoraan visuaalisessa muodossa, mutta varsinkin tilastoista ja sensoreiden avulla kerätyn datan visualisointi voi tuoda suorituskyvyn kehittämisen kannalta valtavasti lisäarvoa urheilujoukkueelle.

Joukkueurheilussa visualisoinnin avulla voidaan tunnistaa ja osoittaa esimerkiksi alueita, josta todennäköisesti syntyy maali, yleisiä tai onnistuneita syöttöketjuja tai kuvioita, sekä perinteisiä kaavioita ja kuvaajia (Schumaker et al. 2010). Visuaalista dataa urheilutapahtumista tuottaa joukkueiden analytikkojen ja yhteistyökumppaneiden lisäksi yleensä myös urheiluliigat. Esimerkiksi PohjoisAmerikan suurimman urheilusarjat (MLB, NBA, NHL) tuottavat yhteistyökumppaneineen paljon erilaisia visuaalisia raportteja, joita voidaan hyödyntää sarjan, joukkueen ja yksilöiden suorituskyvyn kehittämisessä (Schumaker et al. 2010).



Kuva 4: Lämpökartta jalkapallon maalia kohti menevien potkujen todennäköisyysjakaumasta. (Bialkowski et al. 2015)

Kuvassa 4 on esiteltyä esimerkki visualisoinnin hyödyntämisestä maalinteon tarkastelemisessa jalkapallossa. Kuvassa vasemmanpuoleinen kartta näyttää kaikkien potkujen todennäköisyysjakauman, eli miltä alueilta pallo todennäköisesti potkaistaan maalia kohti. Oikeanpuoleinen kartta puolestaan esittää maaliin menneiden potkujen todennäköisyysjakaumaa. Esimerkin mukainen maalinteon lämpökartta on tehty hyödyntäen pelaajien ja pallon sijaintidataa ja tilastoja. (Bialkowski et al. 2015.) Lämpökartta on yksi hyvä esimerkki urheiludatan visualisoinnin hyödyistä, mutta sen toimintaperiaatteen syvällisempi tarkastelu ei ole tässä tutkimuksessa tarpeellista.

5. SUORITUSKYKYÄ INFORMAATIOSTA

Tässä luvussa tarkastellaan miten jalostetusta datasta, eli informaatiosta voidaan jalostaa edelleen tietoa ja osaamista urheilujoukkueelle. Luvussa käsitellään joukkueen suorituskykyyn vaikuttavia tekijöitä yksilöiden, joukkueen ja valmennuksen näkökulmasta. Lisäksi käydään läpi, miten datasta saatu informaatio hyödynnettyä tehokkaasti joukkueen harjoittelussa ja urheiluotteluissa.

Kuten johdantoluvussa esitellystä kuvasta 1 voidaan huomata, siirtyessä datasta ja informaatiosta ylöspäin tiedon tasoissa, tieto muuttuu eksplisiittisestä hiljaiseksi ja myös ohjelmoitavuus vähenee samalla. Tästä syystä urheilujoukkueen suorituskyvyn kehittämiseen tähtäävä data-analyysiprosessi alkaa tässä vaiheessa olemaan hyvin inhimillisten ratkaisujen varassa. Varsinkin kilpaurheilujoukkueet hyödyntävät kuitenkin päätöksentekoprosessin tukena harjoitteluista ja otteluista kerättyä dataa harjoittelun suunnittelussa, loukkaantumisten ehkäisemisessä ja personalisoitujen palautumisstrategioiden suunnittelussa (Basole & Saupe 2016).

5.1 Joukkueen suorituskykyyn vaikuttavat tekijät

Urheilijan suorituskyky urheiluottelussa riippuu hyvin pitkälti harjoittelusta ja fyysisestä kuntotasosta, sekä oikeanlaisesta kehon ja mielen valmistautumisesta otteluun (Martin 2016). Urheilun suorituskykydatan ja -mallien avulla voidaan opastaa yksittäisen urheilijan päivittäisiä valintoja, harjoittelua, ottelussa tehtäviä ratkaisuja ja pelipäivän päätöksiä tukemaan suorituskykyä (Miller 2015). Urheilujoukkue puolestaan koostuu useista yksilöllisiä päätöksiä tekevästä urheilijoista ja valmentajista, joilla on yhteinen tavoite suoriutua urheiluottelussa mahdollisimman hyvin. Joukkueen suorituskykyyn vaikuttaa useita tekijöitä, joita voidaan luokitella esimerkiksi pitkän- ja lyhyen aikavälin tekijöihin (Ofoghi et al. 2013).

Joukkueen suorituskykyä mitattaessa olennaisena osana on alaluvussa 3.1 mainitut yksilön suorituskykyä mittaavat KPI-mittarit, joiden avulla pystytään tarkastelemaan mitkä asiat vaikuttavat suoritukseen ja lopputulokseen merkittävästi. Urheilijan lajitaitoa mittaavat KPI:t ovat hyvin lajispesifejä, kuten jalkapallon tai jääkiekon lajitaidot, mutta useimmat mitattavat fyysiset ominaisuudet ovat huipputason joukkueurheilijoilla kohtalaisen yhtäläisiä (Kniffin et al. 2017). Tällaisia fyysisiä ominaisuuksia ovat esimerkiksi nopeus, voima ja kestävyys (Hughes et al. 2012; Kniffin et al. 2017; Kretzer et al. 2015).

Toinen varsinkin pitkällä aikavälillä joukkueen suorituskykyyn vaikuttava tekijä on yksilöiden loukkaantumiset ja niistä palautuminen (Miller 2015). Mikäli joukkue joutuu pelaamaan osan kauden otteluistaan ilman tärkeitä pelaajia tai he pelaavat puolikuntoisena, vaikutukset joukkueen suorituskykyyn ovat väistämättömiä. Harjoittelussa tai ottelussa yleisesti sattuneita loukkaantumisia ja niihin johtaneita tilanteita voidaan tutkia videotallenteiden ja erilaisten sensoreiden avulla (Bussey & Bartlett 2013).

Joukkueurheilussa yksittäisten pelaajien lisäksi valmentajien vaikutus joukkueen suorituskykyyn on valtava. Valmentajien vastuulla on tehdä päätöksiä ottelukokoonpanoista, pelaajien roolituksesta ja sijoittumisesta, harjoittelun suunnittelusta, pelistrategiasta, sekä pelin aikana taktiikasta (Miller 2015). Näiden perusteella joukkueen suorituskykyyn voidaan sanoa koostuvan hyvin pitkälti joukkueen yksilöiden ja valmentajien välisestä yhteistyöstä. Suorituskykyyn pystytään vaikuttamaan harjoittelussa ja pelitapahtumissa tehdyillä lyhyen tai pitkän aikavälin päätöksillä, joita pystytään tukemaan jalostetun ja analysoidun urheiludatan avulla.

5.2 Päätöksenteko harjoittelussa

Perustietona voidaan pitää, että ihminen pystyy kehittämään ja ylläpitämään omaa suorituskykyään harjoittelun avulla, joka sisältää paljon laadukkaita toistoja. Tämä pätee myös urheilussa, jossa harjoittelun avulla voidaan kehittää erilaisia lajitekniisiä taitoja tai joukkueena erilaisten pelikuvioiden toimintaa. Joukkueurheilussa ja yksilöurheilussa fyysisten suoritusten analysointi on kohtalaisen samankaltaista, mutta joukkueurheilussa analysoitavia on useita eri roolitusten omaavia urheilijoita samalla kentällä, jolloin analyysin kohde voi olla esimerkiksi hyökkäävä vastaa puolustaja tai maalivahti, tai ryhmä pelaajia (Ofoghi et al. 2013). Koska joukkueurheilun analysointiin liittyy paljon inhimillisiä ja sattumanvaraisia muuttujia, myös joukkueen suorituskykyyn vaikuttaa useiden asioiden summa. Alla on esiteltynä muutamia olennaisia tapoja vaikuttaa joukkueen suorituskykyyn harjoittelun kautta.

Huipputason urheilujoukkueissa harjoittelumäärät ovat valtavia ja yksittäiset urheilijat kärsivät paljon suorituskykyä alentavasta ylikuormitustilasta, joka ilmentyy usein palautumisen heikkenemisenä, stressitasojen nousuna ja fyysisinä loukkaantumisina (Drust et al. 2017). Ongelmaan voidaan vaikuttaa löytämällä urheilijoille sopiva harjoittelurytmitys, jolloin palautuminen harjoituksesta on optimaalinen ennen seuraavaa harjoitusta. Optimaalisen harjoittelurytmituksen löytämisessä voidaan hyödyntää data-analytiikan

työkaluja, kuten ihmisen elintoimintoja mittaavia puettavia sensoreita ja tilastollisia menetelmiä (Link 2018). Näin ollen yksi suorituskykyä edistävä tapa hyödyntää dataa joukkueen harjoittelussa on tarkastella pelaajien kuormitusta ja jaksottaa harjoittelua.

Joukkueharjoittelun rytmityksen lisäksi myös harjoitteiden sisältö on hyvin oleellista joukkueen suorituskyvyn edistämisen kannalta. Analysoimalla tilastoista ja videotallenteilta kerättyä dataa, voidaan tunnistaa oman joukkueen ja vastustajien vahvuuksia ja heikkouksia, sekä luoda niistä hyödyllisiä KPI-mittaristoja (Link 2018). Näitä mittaristoja voidaan hyödyntää harjoittelun suunnittelussa ja myöhemmin siirtää otteluihin. McGarry et al. (2013) esittävät kirjassaan neljä tapaa lisätä suorituskykyä harjoittelun kautta, demonstraatiot, ohjeistukset, palaute ja harjoitteluorganisaatiot. Demonstraatioharjoittelussa pyritään näyttämään joukkueelle tai urheilijalle demonstraatioita jostain tietyistä suorituksesta ja opettamaan yksityiskohtaisesti jotain tiettyä taitoa tai kuviota. Ohjeistamisessa puolestaan annetaan enemmän mahdollisuutta soveltamiselle harjoituksessa, jolloin harjoitus kuvaa enemmän pelitilannetta ja näin ollen helpottaa pelaamista paineen alla. Palautteen antamisella tarkoitetaan nimensä mukaisesti suorituksen jälkeistä palautetta, mikä mahdollistaa virheiden tehokkaan korjaamisen. Harjoitteluorganisaatioilla tarkoitetaan harjoittelua, jossa ei juurikaan anneta ohjeistusta alkuun, esimerkki tällaisesta harjoituksesta voi olla harjoitustilanteessa pelattavat pelit ja kilpailut. (McGarry et al. 2013.)

5.3 Päätöksenteko pelitilanteissa

Päätöksenteko pelitilanteissa perustuu pääosin lyhyeen tai reaaliaikaiseen päätöksentekoon. Urheilussa joukkueet voivat hyödyntää dataa ja analytiikkaa kokoonpanojen ja joukkueen pelinsisäisen strategian määrittelemisessä maksimoidakseen suorituskykynsä ottelussa (Miller 2015). Huipputasen urheilujoukkueilla on yleensä käytössään useisiin eri tehtäviin erikoistuneita valmentajia päävalmentajan lisäksi, jotka vastaavat esimerkiksi erikoistilanteista ja vastustajien analysoinnista. (Erhardt et al. 2014). Tämä mahdollistaa nopeiden päätösten tekemisen eri pelitilanteissa.

Reaaliaikaisessa päätöksenteossa automatisoidut ja nopeat datankäsittelyjärjestelmät ovat erittäin oleellisessa osassa. Ofoghi et al. (2013) mainitsee tekstissään, että reaaliaikaisessa päätöksenteossa käytetyimpiä työkaluja on suhteiden mallinnus- tai assosiaatiosääntöalgoritmeja hyödyntävät järjestelmät. Joukkueurheilussa varsinkin pelaajien osalta pelitilannepäätökset tulevat usein kuitenkin harjoittelun kautta automaationa.

5.4 Modernit analytiikkajärjestelmät joukkueurheilussa

Kuten tekstissä on useaan kertaan mainittu, datan tehokkaan hyödyntämisen joukkueurheilussa mahdollistaa pitkälle automatisoidut järjestelmät, jotka vastaavat niin datan keräämisestä, jalostamisesta ja analysoinnista. Järjestelmissä hyödynnetään erilaisia data-analytiikan keinoja, kuten koneoppimista, tiedon louhintaa ja tekoälyä (engl. artificial intelligence). Alla on esiteltynä muutamia tunnettuja analytiikkajärjestelmien tarjoajia, joita käytetään ammattijoukkueurheilussa päätöksenteon tukena.

Stats Perform tarjoaa analytiikkapalveluita päätöksenteon tueksi jalkapallo, koripallo, rugby, kriketti, Amerikkalaisen jalkapallon, ja baseball joukkueille. Yritys tekee yhteistyötä 500 joukkueen kanssa ympäri maailmaa tarjoten palveluita esimerkiksi suorituskyvyn analysointiin, pelaajien tarkkailuun ja pitkän aikavälin strategiselle suunnittelulle. (Stats Perform Group 2020.)

ChyronHegon tarjoaa videolta ja puettavista laitteista kerättyä visuaalista dataa urheilujoukkueille. Aiemmin tekstissä mainittu jalkapallossa ja baseballissa käytettävä TRACAB paikannusjärjestelmä on ChyronHegonin tuottama palvelu, jonka avulla urheilujoukkueet pystyvät tarkkailemaan urheilijan kuntotasoja, ehkäisemään loukkaantumisia, tekemään uusia taktisia päätöksiä ja parantaa mahdollisuuksiaan onnistua. (ChyronHego Corporation 2020.)

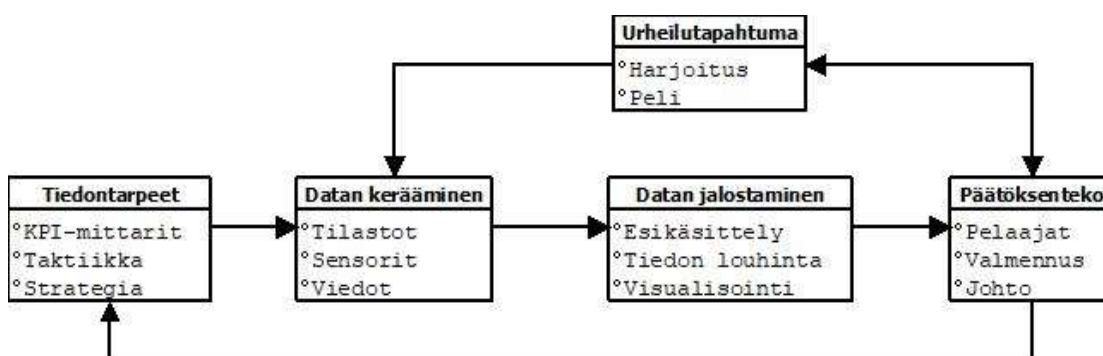
Elias Sports Bureau puolestaan toimii yhteistyössä suurten ammattuurheilusarjojen kanssa kuten MLB, NFL, NBA, NHL, MLS ja WNBA. Yhtiö tarjoaa asiakkaiden tarpeiden mukaan kustomoituja palveluja liittyen esimerkiksi grafiikkaan tai tilastointiin. (Elias Sports Bureau Inc 2020.)

6. JOHTOPÄÄTÖKSET

Tutkimuksen viimeisessä luvussa vedetään tiivistetysti yhteen tutkimuksen tulokset, arvioidaan tutkimuksen tuloksia ja pohditaan mahdollisia jatkotutkimusmahdollisuuksia.

6.1 Yhteenveto

Tässä kandidaatintyön tutkimuksessa tutkittiin data-analytiikan hyödyntämistä joukkueurheilussa, pääpainona toimi joukkueen suorituskyvyn kehittäminen. Tutkimuksen rakenne noudattaa kuvan 1 mukaisesti tiedon jalostamisen eri vaiheiden etenemisjärjestystä alkaen tiedontarpeiden määrittelystä ja datan keräämisestä, sekä päättyen jalostetun tiedon hyödyntämiseen päätöksenteossa. Toisin sanoen tutkimuksessa liikutaan tiedon tasojen pyramidin alimmalta tasolta ylöspäin hyödyntäen urheiludataa. Toisaalta tutkimuksen perusteella prosessin voidaan todeta olevan jatkuva, sillä kilpailukyvyn ylläpitämiseksi suorituskykyä pystyttävä ylläpitämään tai kehittämään jatkuvasti. Kuvassa 5 on esiteltyä pääpiirteittäin yhteenvetokuvaajana tutkimuksen aihealue. Kuvasta voidaan huomata, että prosessi lähtee liikkeelle tiedontarpeiden määrittelystä, mutta tiedontarpeet määrittelee kuitenkin päätöksentekijät. Tiedontarpeiden määrittelystä siirrytään urheilutapahtumista saatavan datan keräämiseen ja jalostamiseen. Kuvassa nuolet osoittavat prosessin vaiheiden kulkusuunnan. Kaksisuuntainen nuoli urheilutapahtuman ja päätöksenteon välillä tarkoittaa sitä, että urheilutapahtumaa voidaan hyödyntää päätöksenteon perusteena tiedontarpeiden asettamisessa.



Kuva 5: Tulosten yhteenveto.

Kuten kuvasta 5 voidaan huomata, suorituskyvyn kannalta olennaisimpia päätöksentekijöitä joukkueurheilussa ovat yksittäiset pelaajat, valmentajat ja joukkueen johto. Päätöksentekijät hyödyntävät urheilusta kerättävää dataa otteluanalyseissä, pelistrategian-

ja harjoittelun suunnittelussa, loukkaantumisten ehkäisemisessä, joukkueen ja pelaajien suorituskyvyn hallinnassa, kehittämisessä ja ennustamisessa, sekä joukkueen rakentamisessa. Dataa pystytään keräämään urheiluotteluissa urheilijoiden suorituksista ja ratkaisuksista, valmentajien ja johdon päätöksistä, sekä muista ottelun sisällä tapahtuvista tapahtumista. Lisäksi dataa kerätään joukkueen harjoituksista ja urheilijoiden arkisista toimista, kuten syömisestä, nukkumisesta ja kehon toiminnoista. Datan kerääminen tapahtuu urheilussa pääosin tilastojen, videoiden ja erilaisten sensorien avulla.

Tutkimuksessa havaittiin myös, että dataa on saatavilla hyvin paljon, se liikkuu nopeasti, sitä saatavilla monessa eri muodossa ja sen käsittely perinteisin tilastollisin menetelmin on haastavaa, joten se täyttää massadatalle tyypilliset tunnusmerkit. Näin ollen kerättyä dataa tulee jalostaa, jotta siitä saataisiin irti mahdollisimman suuri hyöty suorituskyvyn kehittämisen kannalta. Joukkueurheiludatan jalostaminen pitää sisällään eri vaiheita, joihin voi lukeutua datan esikäsittelyä, tiedon louhintaa ja -visualisointia. Datan esikäsittelyyn lukeutuu datan siistimistä, muokkaamista, yhdistelyä, normalisointia, puuttuvien arvojen korvaamista ja poikkeavan tai häiriintyneen datan tunnistamista. Urheilussa käytettäviä tyypillisiä tiedon louhinnan menetelmiä ovat esimerkiksi klusterointi, luokittelu, suhteiden mallintaminen ja assosiaatiosäännöt.

Datan kerääminen ja jalostaminen onnistuu erilaisten automatisoitujen yhdessä toimivien ohjelmoitavien järjestelmien avulla, mutta niistä saatavan informaation siirtäminen suorituskyvyksi urheiluottelussa vaatii inhimillisiä päätöksiä ja tekoja päätöksentekijöiltä. Päätöksiä tehdään sekä pitkällä että lyhyellä aikavälillä. Pitkän aikavälin päätöksiä ovat esimerkiksi joukkueen rakenteelliset päätökset ja harjoitteluun vaikuttaminen. Pitkän aikavälin päätöksentekoa voidaan tukea esimerkiksi tilastollisen datan avulla. Lyhyen aikavälin päätöksiä voidaan tehdä myös harjoittelussa esimerkiksi pelin taktiikkaa harjoitellessa. Otteluissa lyhyen aikavälin päätökset tehdään reaaliajassa tuotetun videomateriaalin avulla tai harjoittelun kautta hankitun osaamisen avulla.

6.2 Tulosten arviointi ja jatkotutkimusmahdollisuudet

Tutkimuksen tulokset avaavat näkökulmia siihen, miten data-analytiikkaa voidaan hyödyntää joukkueurheilun päätöksenteossa. Aihealue osoittautui kuitenkin ennakoitua laajemmaksi ja varsinaisia yksiselitteisiä tapoja urheiludatan analysoinnille ei paljastunut. Tarkempia ja yksityiskohtaisempia tuloksia olisi voitu saavuttaa rajaamalla tutkimus johonkin tiettyyn urheilulajiin tai menetelmään, kuten esimerkiksi tilastollisen data-ana-

lyysin hyödyntämiseen jääkiekossa. Tutkimuksessa tavoitteena oli kuitenkin saada yleiskuva joukkueurheilussa käytettävästä analytiikasta ja siinä käytettävistä työkaluista. Tässä tavoitteessa tutkimus on kohtalaisen onnistunut.

Haasteelliseksi tutkimuksessa osoittautui käytettävien menetelmien ja työkalujen nopea teknologinen kehitys, sillä jopa 10 vuotta vanhat menetelmät alkavat olla alalla erittäin vanhentuneita. Esimerkiksi IoT -laitteiden hyödyntäminen ja kehitys urheilussa on ollut pinnalla alalla vasta muutamia vuosia ja tämän takia tietoa laitteiden toiminnasta tai käyttötarkoituksista on saatavilla hyvin vähän. Alun perin tutkimuksessa oli tavoitteena löytää vastauksia myös urheilussa käytettävän data-analytiikan tilasta Suomalaisessa urheilussa, mutta aineistoa tähän ei juurikaan ollut saatavilla Wisehockey -analytiikkajärjestelmän lisäksi.

Data-analytiikan työkalut ja menetelmät kehittyvät jatkuvasti myös urheilussa, joten aihealue sisältää myös runsaasti jatkotutkimusmahdollisuuksia varsinkin tiedonhallinnan, analytiikan, datatieteen ja muiden tietoteknisten aihealueiden osalta. Puettavien laitteiden jatkuvan kehityksen myötä myös ihmisen suorituskykyyn ja elintoimintoihin liittyviä ongelmia voidaan lähestyä urheilussa käytettävän analytiikan kautta. Lisäksi koska aihealue ei ole Suomessa tieteellisesti kovin laajalti tutkittua ja ainakaan aineistoa ei juurikaan ole saatavilla, tutkimuksen tekeminen Suomessa käytettävän urheiluanalytiikan tilasta olisi tarpeellista.

LÄHTEET

- Ai, L., Cao, Y., Mi, Q., Sun, K., Qiao, M., Wang, Q., Yuan, X., Zang, S., Zhang, J. & Zuo, Y., 2018. Human motion monitoring in sports using wearable graphene-coated fiber sensors. *Sensors & Actuators: A.Physical*, 274, pp. 132-140.
- Alamar, B.C., 2013. *Sports Analytics. A guide for coaches, managers, and other decision makers*. New York: Columbia University Press.
- Aroganam, G., Manivannan, N. & Harrison, D., 2019. Review on Wearable Technology Sensors Used in Consumer Sport Applications. *Sensors (Basel, Switzerland); Sensors (Basel)*, 19(9), pp. 1983.
- Azar, O., Morgulev, E. & Lidor, R., 2018. Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), pp. 213-222.
- Azzalini, A., Scarpa, B. & Walton, G., 2012. *Data analysis and data mining an introduction*. Oxford: Oxford University Press.
- Baerg, A., 2017. Big Data, Sport, & the Digital Divide: Theorizing How Athletes Might Respond to Big Data Monitoring. *Journal of Sport & Social Issues*, 41(1), pp. 3-20.
- Barbosa, T., 2018. Smart sport equipment: reshaping the sports landscape. *Motricidade*, 14(2-3), pp. 1-2.
- Bar-Eli, M., Plessner, H. & Raab, M., 2011. *Judgement, decision making and success in sport*. Malden, Mass: Wiley.
- Barlow, J., 2015. *Data Analytics in Sports*. O'Reilly Media, Inc.
- Basole, R.C. & Saupe, D., 2016. Sports Data Visualization [Guest editors' introduction]. *IEEE Computer Graphics and Applications*, 36(5), pp. 24-26.
- Bialkowski, A., Carr, P., Lucey, P., Matthews, I. & Monfort, M., 2015. "Quality vs Quantity": Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data, *SPORTS ANALYTICS CONFERENCE 2015*.
- Brefeld, U., Davis, J., Van Haaren, J. & Zimmermann, A., 2019. *Machine Learning and Data Mining for Sports Analytics 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings*. 1 edn. Cham: Springer International Publishing.

- Bussey, M. & Bartlett, R., 2013. *Sports Biomechanics: Reducing Injury Risk and Improving Sports Performance*. Florence: Taylor and Francis.
- Carvalho, A., Ferreira, J., Ratten, V. & Miragaia, D., 2019. Interactions between financial efficiency and sports performance. *Journal of entrepreneurship and public policy*, 9(1), pp. 84-102.
- Choo, C.W., 2002. *Information Management for the Intelligent Organization, Art of Scanning the Environment*. USA: Information Today.
- Chyronhego Corporation, 2020. *Sports Tracking: Revealing the DNA of Sport*. Available: <https://chyronhego.com/products/sports-tracking/> [Jul 31, 2020].
- Dangeti, P., 2017. *Statistics for machine learning: build supervised, unsupervised, and reinforcement learning models using both Python and R*. Birmingham, England: Packt Publishing.
- Davenport, T.H., 2006. *Competing on Analytics*. Harvard Business Review.
- Delen, D. & Olson, D.L., 2008. *Advanced Data Mining Techniques*. 1 edn. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Drust, B., Nedergaard, N.J., Robinson, M.A. & Vanrenterghem, J., 2017. Training Load Monitoring in Team Sports: A Novel Framework Separating Physiological and Biomechanical Load-Adaptation Pathways. *Sports medicine (Auckland); Sports Med*, 47(11), pp. 2135-2142.
- Ebling, M.R., 2016. IoT: From Sports to Fashion and Everything In-Between. *IEEE Pervasive Computing*, 15(4), pp. 2-4.
- Elias Sports Bureau Inc, 2020. *About Elias*. Available: <http://www.esb.com/about/> [Jul 31, 2020].
- Erhardt, N., Martin-Rios, C. & Harkins, J., 2014. Knowledge flow from the top: the importance of teamwork structure in team sports. *European Sport Management Quarterly*, 14(4), pp. 375-396.
- García, S., Luengo, J. & Herrera, F., 2015. *Data Preprocessing in Data Mining*. 1 edn. Cham: Springer International Publishing.
- Gowda, M., Dhekne, A., Shen, S., Choudhury, R., Yang, S.X., Yang, L., Golwalkar, S. & Essanian, A., 2018. IoT Platform for Sports Analytics. *GetMobile: Mobile Computing and Communications*, 21(4), pp. 8-14.
- Gratton, C., 2012. *The global economics of sport*. London: Routledge.

Harjamäki, J., 2018. Selvitys data-analytiikan nykytilasta ja data-analytiikan hyödyntämisestä Satakunnassa. TTY Pori: Analyttinen-hanke.

Hovi, J., 2018. Data-alan termien selitykset ja kuvaukset. Available: <https://www.ari-hovi.com/3274-2/> [Jul 17, 2020].

Hughes, M., Caudrelier, T., James, N., Donnelly, I., Kirkbride, A. & Duschesne, C., 2012. Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position. *Journal of human sport and exercise*, 7(2), pp. 402-412.

Kniffin, K.M., Howley, T. & Bardreau, C., 2017. Putting Muscle into Sports Analytics: Strength, Conditioning, and Ice Hockey Performance. *Journal of Strength and Conditioning Research; J Strength Cond Res*, 31(12), pp. 3253-3259.

Kretzer, M., Mädche, A. & Shah, F.A., 2015. Designing an Analytics Platform for Professional Sports Teams, Thirty Sixth Conference on Information Systems 2015.

Laihonen, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J., Pekkola, S., Virtanen, P., Vuori, V., & Yliniemi, T., 2013. Tietojohdaminen. Tampereen teknillinen yliopisto, Tietojohdamisen tutkimuskeskus Novi.

Link D. Sports Analytics. *German Journal of Exercise and Sport Research*. 2018a;48(1):13-25.

Linke, D., Link, D. & Lames, M., 2020. Football-specific validity of TRACAB's optical video tracking systems. *PLoS one; PLoS One*, 15(3), pp. e0230179.

Martin, L., 2016. *Sports Performance Measurement and Analytics: The Science of Assessing Performance, Predicting Future Outcomes, Interpreting Statistical Models, and Evaluating the Market Value of Athletes*. PH Professional Business.

McGarry, T., O'donoghue, P. & Sampaio, A., 2013. *Routledge handbook of sports performance analysis*. London: Routledge.

Mendes Jr, J., Vieira, M., Pires, M. & Stevan Jr, S., 2016. Sensor Fusion and Smart Sensor in Sports and Biomedical Applications. *Sensors (Basel, Switzerland); Sensors (Basel)*, 16(10), pp. 1569.

Miller, T., 2015. *Sports Analytics and Data Science: Winning the Game with Methods and Models*. PH Professional Business.

Ofoghi, B., Zeleznikow, J., Macmahon, C. & Raab, M., 2013. Data Mining in Elite Sports: A Review and a Framework. *Measurement in Physical Education and Exercise Science*, 17(3), pp. 171-186.

Researchandmarkets, 2019. Sports - \$614 Billion Global Market Opportunities & Strategies to 2022 - ResearchAndMarkets.com. Available: <https://www.business-wire.com/news/home/20190514005472/en/Sports---614-Billion-Global-Market-Opportunities> [Jul 2, 2020].

Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), pp. 163-180.

Schumaker, R.P., Solieman, O.K., Chen, H., Sharda, R. & Voss, S., 2010. *Sports Data Mining*. 1 edn. Boston, MA: Springer US.

Sedkaoui, S., 2018. *Data analytics and big data*. London: ISTE.

Sellitto, C., 2015. *Big Data can give Athletes the Winning Edge*. Rockaway: Advantage Business Media.

Stats Perform Group, 2020. *Sports Data - Sports AI, Technology, Data Feeds*. Available: <https://www.statsperform.com/> [Jul 31, 2020].

Wisehockey Oy, 2020. *Real-Time Sport Analytics Platform: Overview of the Wisehockey system*. Available: https://wisehockey.com/wp-content/uploads/2020/02/wh_overview_en.pdf.

Zhou, Q., 2012. Research on the Guidance Methods of Sports Training Based on Data Mining Technology. *Applied mechanics and materials*, 155–156, pp. 590-595.