Data-analytiikan soveltaminen ammattijalkapalloon

Turun yliopisto
Tietotekniikan laitos
LuK-tutkielma
Tietojenkäsittelytiede
Joulukuu 2024
Sami Eloneva

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

TURUN YLIOPISTO
Tietotekniikan laitos

Sami Eloneva: Data-analytiikan soveltaminen ammattijalkapalloon

LuK-tutkielma, 24 s.
Tietojenkäsittelytiede
Joulukuu 2024

Ammattijalkapallossa erot joukkueiden välillä ovat äärimmäisen pienet. Joukkueet pyrkivät jatkuvasti saavuttamaan etulyöntiaseman vastustajaansa nähden. Data-analytiikka on yksi tapa, miten etulyöntiasemaa pyritään saavuttamaan. Pelaajista ja joukkueen toiminnasta kerätään päivittäin suuria määriä dataa. Tämä tutkielma toteutettiin kirjallisuuskatsauksena. Tutkielman perusteella data-analytiikan sovellusalueet ammattijalkapallossa voidaan jakaa pelaaja ja joukkuetason analyysiin. Keskeisimmät sovellusalueet keskittyivät suorituksien arviointiin sekä taktiikan analysointiin. Koneoppimismenetelmät ovat keskeisessä osassa kerätyn datan analysoinnissa, koska ne mahdollistavat tarkemman ja tehokkaamman analyysin suurista ja monimutkaisista datakokoelmista. Data-analytiikan hyödyt näkyvät erityisesti toiminnan optimoinnissa ja päätöksenteon tukemisessa, mutta sen käyttöön liittyy myös haasteita, kuten datan puolueellisuus ja tulkinnan vaikeudet. Tässä tutkielmassa tutkitaan, mitä sovellusalueita ja hyötyjä data-analytiikalla on ammattijalkapallossa sekä mitä haasteita tähän liittyy. Tutkielma osoitti, että vaikka data-analytiikan merkitys kasvaa, sen täysi potentiaali ammattijalkapallossa on vielä hyödyntämättä.

Asiasanat: data-analytiikka, sovellusalueet, koneoppiminen, ammattijalkapallo, hyödyt, haasteet

Sisällys

1 Johdanto **1**

2 Data-analytiikka urheilussa **4**

2.1 Data-analytiikka ja massadata 4

2.2 Datan kerääminen jalkapallossa 6

2.3 Analyysimenetelmät 8

3 Data-analytiikan sovellukset ammattijalkapallossa **10**

3.1 Pelaajien analyysi 11

3.1.1 Data-analytiikka ja koneoppiminen pelaajien arvioinnissa . . . 11

3.1.2 Suorituksien mittarit 13

3.2 Joukkuetason analyysi 14

3.2.1 Taktiikan analyysi ja optimointi 15

3.2.2 Menetelmiä joukkuetason analyysiin 16

3.3 Haasteet ja rajoitukset 18

4 Yhteenveto **22**

Lähdeluettelo **25**

Kuvat

1.1 Aineistonhakuprosessi 2

Taulukot

3.1 Pelaajien analyysiä koskevat aineistot 10

3.2 Joukkuetason analyysiä koskevat aineistot 11

1 Johdanto

Data-analytiikan rooli on kasvanut huomattavasti ammattiurheilussa ja siitä on tullut myös osa ammattijalkapalloa. Ammattiurheilun korkeiden panoksien ja pienien marginaalien vuoksi joukkueet pyrkivät saamaan pienintäkin kilpailuetua. Joukkueet pyrkivät saavuttamaan kilpailuetua hyödyntämällä pelaajista ja otteluista kerättyä dataa. Analytiikan avulla voidaan saada merkittäviä havaintoja pelaajista, joukkeesta ja taktiikoista. Valmentajat ja pelaajat pystyvät hyödyntämään havaintoja harjoituksien ja suorituksien parantamiseksi. [1] Urheilussa analytiikalla siis pyritään mallintamaan ja analysoimaan pelaajien tai joukkueen suorituskykyä tieteellisten menetelmien avulla, hyödyntäen kerättyä dataa ja analytiikan menetelmiä datan analysoinnissa ja tulkinnassa [2].

Data-analytiikan juuret urheilussa ulottuvat 1960-luvulle asti, jolloin analyysi perustui yksinkertaisiin muistiinpanoihin. Nopean teknologian kehityksen tuloksena data-analytiikan soveltaminen urheiluun on muuttunut paljon vuosien saatossa. [2] Nykyään edistykselliset seurantajärjestelmät ja algoritmit tuottavat jatkuvasti uutta dataa pelaajista ja ottelutapahtumista, mikä on mahdollistanut suurten datamäärien analysoinnin esimerkiksi koneoppimismallien avulla. [3]

Jalkapallon analysointi on erityisen haastavaa muihin joukkuelajeihin verrattuna, sillä suuri kenttä, useat pelaajat, pitkä peliaika ja jatkuvasti muuttuva pelitilanne tekevät siitä monimutkaisen analyysinkohteen. [3] Jalkapallo on maailman suosituin laji, jota pelaavat miljoonat ja seuraavat yli miljardit ihmiset ympäri maailman.

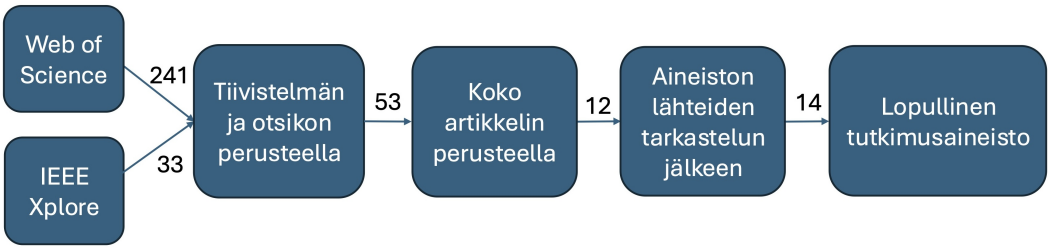
Lajin ympärille on muodostunut teollisuus, joka tuottaa valtavasti tuloja pelaajille, seuroille, valmentajille ja henkilökunnalle. Nykyään jalkapalloseuroja ei nähdä pelkästään urheiluseuroina, vaan enemmänkin yrityksinä. [4] Pelaajien ja joukkueen analysointi sekä jatkuva kehitys on elinehto seuran suorituskyvyn ylläpitämiseksi.

Tämän tutkielman tavoitteena on selvittää data-analytiikan soveltamisen hyötyjä ja sovellusalueita ammattijalkapallossa sekä tarkastella data-analytiikan tuomia haasteita ammattijalkapallon kontekstissa. Tutkielman tutkimuskysymykset ovat:

TK1: Mitkä ovat data-analytiikan sovellusalueet ja hyödyt ammattijalkapallossa?

TK2: Mitkä ovat data-analytiikan haasteet ammattijalkapallossa?

Tutkielma toteutettiin kirjallisuuskatsauksena. Tietokannat joissa haku toteutettiin oli Web Of Science ja IEEE. Nämä tietokannat valikoituivat, sillä ne tuottivat osuvimmat tulokset hakulauseelle. Haussa käytetty hakulause oli (*"data analy*"* OR *"big data"* OR *"sport* analy*"* OR *"football analy*"*) AND (*"professional football"* OR *"professional soccer"* OR *"professional clubs"* OR *"elite soccer"*). Hakutulokset suodatettiin englanninkielisiksi. Aineistonhakuprosessi on esitelty kuvassa 1.1.



Kuva 1.1: Aineistonhakuprosessi

Ensimmäinen haku tuotti yhteensä 274 artikkelia, joista jokaisesta käytiin otsikko ja tiivistelmä läpi. Ensimmäisellä kerralla karsittiin joukosta kaikki artikkelit, jotka eivät vastanneet tutkimuskysymyksiin tai eivät koskeneet ammattijalkapalloa. Ensimmäisen karsinnan jälkeen jokainen artikkeli luettiin läpi, joista valikoitui mukaan 12 artikkelia. Näistä vielä käytiin läpi lähdeluettelot, joista mukaan valikoitui

vielä kaksi soveltuvaa artikkelia. Lopullinen tutkimusaineisto sisältää 14 artikkelia.

Tutkielma koostuu johdannon lisäksi kolmesta luvusta. Luvussa 2 käydään läpi mitä data-analytiikka on ja miten dataa kerätään. Luvussa 3 esitellään kirjallisuuskatsauksessa käsitellyt artikkelit, jotka voitiin jakaa teemoittain pelaajia ja joukkuetta koskeviin. Luvussa 4 on yhteenveto tutkielmasta, jossa vastataan tiiviisti asetettuihin tutkimuskysymyksiin.

2 Data-analytiikka urheilussa

Data-analytiikka on avannut uusia mahdollisuuksia urheilussa, kuten suorituskyvyn parantamisessa. Tämän luvun tarkoituksena on määritellä data-analytiikka, esitellä erilaisia datan keruumenetelmiä jalkapallossa sekä tarkastella analyysimenetelmiä kerätylle datalle.

2.1 Data-analytiikka ja massadata

Data-analytiikassa yhdistetään ja kerätään tietoa useasta eri lähteestä jatkokäyttöä varten. Datasta on tavoitteena saada hyödyllisiä havaintoja esimerkiksi päätöksenteon tai ennustamisen tueksi. [5] Dataa tulisi prosessoida asiallisesti, jotta raakadatasta saadaan poimittua hyödyllistä tietoa. Tiedon prosessointiin kuuluu esimerkiksi tiedonintegrointia, -puhdistusta, -muuntamista, -vähentämistä ja -louhintaa. [6]

Datan prosessointi alkaa keräämisellä ja valikoinnilla. Valikoinnilla tarkoitetaan relevantin tiedon tunnistamista ja erottelua kerätystä datasta tai tietokannoista. Eri tietolähteistä tulevat tiedot on yhdistettävä eli integroitava yhtenäiseksi kokonaisuudeksi. Integroinnin jälkeen data puhdistetaan virheellisistä, epäjohdonmukaisista ja tarpeettomista tiedoista. Tarvittaessa datan määrää voidaan myös vähentää. Seuraavaksi data muunnetaan analyysikelpoiseen muotoon hyödyntäen menetelmiä, jotka yksinkertaistavat dataa ja supistavat sen määrää. [6]

Tiedonlouhinnan tavoitteena on löytää datasta informaatiota, mitä aikasemmin ei ollut saatavilla. Tiedonlouhintaa voidaan suorittaa koneoppimismenetelmillä. Lo-

puksi tulokset arvioidaan ja tulkitaan, mikä on keskeistä onnistuneelle päätöksenteolle ja johtopäätöksille. [6]

Koneoppiminen on keskeinen sovellusalue data-analytiikassa. Koneoppiminen on yksi tekoälyn osa-alue [7]. Koneoppimisessa tavoitteena on parantaa tietokoneohjelman suorituskykyä kouluttamisen avulla. Ohjelmaa opetetaan suorittamaan tiettyä tehtävää harjoitusaineistolla. Koneoppimisen avulla voidaan automatisoida analyyttisten mallien rakentaminen monimutkaisten tehtävien, kuten kohteen tunnistamisen, suorittamiseksi. Koneoppiminen on osoittautunut hyvin soveltuvaksi data-analytiikan menetelmiin, kuten regressioon ja klusterointiin. Koneoppimismallit pystyvät tuottamaan luotettavia sekä toistettavissa olevia tuloksia oppimalla aikaisemmista laskelmista. Mallia opetetaan siis soveltuvalla harjoitusaineistolla, jolloin se oppii tunnistamaan datasta piilotettuja yhteyksiä ja monimutkaisia rakenteita. [8]

Koneoppimisessa käytetyt tekniikat voidaan jakaa kolmeen osaan, ohjattuun oppimiseen (engl. *supervised learning*), ohjaamattomaan oppimiseen (engl. *unsupervised learning*) ja vahvistettuun oppimiseen (engl. *reinforcement learning*). Ohjatussa oppimisessa syötteet yhdistetään valmiiksi tiettyihin tuloksiin. Ohjatussa oppimisessa ohjelmaa koulutetaan siis datalla, missä syötteet ja myös halutut tulokset on ennaltamääritelty. Ohjaamattomassa oppimisessa ohjelma oppii tunnistamaan rakenteita datasta puolestaan ilman ennaltamääriteltyjä vastauksia. Vahvistusoppiminen perustuu virheiden kautta oppimiseen. [7] Vahvistusoppimisessa ohjelma oppii parantamaan suorituskykyään palautteen avulla. Lisäksi edistyneimpiä tekniikoita kuten konvoluutioverkkoja (engl. *Convolutional Neural Network*) voidaan hyödyntää analytiikassa, jotka soveltuvat erityisen hyvin kuvadatan prosessointiin [7].

Datan määrä ja monimutkaisuus kasvaa jatkuvasti [5]. Dataa saadaan nykyään useista eri lähteistä, kuten kannettavista ja puettavista laitteista, sensoreista, internetistä, multimediasovelluksista ja monista muista lähteistä. Massadatalla (engl. *big data*) tarkoitetaan niin suurta datan määrää, että perinteiset tietojärjestelmät eivät

pysty sitä käsittelemään. Nykypäivän data, jota tulisi analysoida, ei ole ainoastaan laajaa vaan koostuu myös eri datan tyypeistä. [6]

Massadatan määrittelemisen apuna voidaan käyttää kolmea V:tä. Kolmella V:llä viitataan massadatan keskeisiin ominaisuuksiin: määrään (engl. *volume*), nopeuteen (engl. *velocity*) ja valikoimaan (engl. *variety*). Tämä tarkoittaa, että dataa on valtavasti, sitä syntyy nopeasti ja sitä tulee monessa eri muodossa sekä useista eri lähteistä. On kuitenkin todettu, että nykyään nämä kolme V:tä eivät enää riitä massadatan kuvailemiseksi. [6] Urheilun kontekstissa massadatan keskeisiin ominaisuuksiin on lisätty arvo (engl. *value*) ja todenmukaisuus (engl. *veracity*) [9].

Massadatalla on ollut ennennäkemättömiä vaikutuksia urheiluteollisuuden kehityksessä. Urheilussa kerätty data on todella monipuolista. Datan lähteet ovat laajoja ja datan tyypit ovat monimutkaisia. [9]

2.2 Datan kerääminen jalkapallossa

Jalkapallossa kerätään jatkuvasti dataa useista eri lähteistä. Datan lähteitä ovat esimerkiksi fyysinen data, ottelupöytäkirjat, pelitapahtumakohtainen data ja seurantadata [10]. Tässä alaluvussa esitetään erilaisia datan keräämisen menetelmiä jalkapallossa.

Yksi merkittävimmistä tavoista kerätä pelaajilta dataa on puettavien teknologioiden avulla. Puettavalla teknologialla viitataan laitteisiin ja sensoreihin, joita voidaan esimerkiksi käyttää terveydentilan seurantaan. Laite kiinnitetään urheilijan kehoon tai vaatteisiin. Laitteet tarjoavat merkityksellistä dataa niin urheilijoille kuin myös valmentajille harjoituksien ja suorituksien parantamiseksi. [11]

Sensorit mittaavat dataa urheilijan kehosta. Sensoreita on useita erilaisia, ja urheilussa merkittävimmät ovat fysiologiset-, ympäristö- ja sijaintisensorit. Fysiologiset sensorit mittaavat kehon biologisia prosesseja, kuten lihasten sähköistä aktiivisuutta (EMG), sydämen toimintaa (ECG) ja aivojen sähköistä aktiivisuutta (EEG).

Ympäristösensorit mittaavat ympäristöä, kuten ilmanlaatua, lämpötilaa ja kosteutta. Ympäristösensoreilla on mahdollista seurata ympäristön vaikutuksia urheilijan suorituksiin sekä terveyteen. [11]

Sijaintisensorit seuraavat urheilijoiden liikkeitä ja sijaintia. Sijaintia on mahdollista seurata Global Navigation Satellite System- (GNSS), Ultra-Wideband- (UWB) ja kamerapohjaisten järjestelmien avulla. GNSS viittaa satelliittipohjaisiin navigaatiojärjestelmiin, joista tunnetuin on Global Positioning System (GPS). GNSS vastaanottaa signaaleja ja analysoi niiden ajoitusta ja sijaintia ja määrittää pelaajan tarkan sijainnin. [11] GPS-järjestelmä on laajalti käytössä ammattijalkapallossa, missä seurat saavat havaintoja pelaajien aktiivisuudesta harjoituksissa sekä otteluissa [12]. Jalkapallossa yleisessä käytössä ovat GNSS-jäljittimet, jotka mittaavat esimerkiksi juostua matkaa, juoksunopeutta ja kiihdytyksiä [10].

Sensorit eivät kuitenkaan ole ainut tapa kerätä dataa jalkapallossa. Konenäön kehittyminen on mahdollistanut pelaajien sekä otteluiden analysoimisen myös videon avulla. Konenäkö on tietojenkäsittelytieteen osa-alue, joka tutkii, kuinka tietokoneet voivat prosessoida ja ymmärtää visuaalista dataa, kuten kuvia ja videoita [13]. Konenäön avulla on mahdollista suorittaa monimutkaisia analyyseja, kuten pelaajien tunnistamista ja luokittelemista esimerkiksi pelinumeron avulla. Tämän avulla on mahdollista seurata kaikkia tapahtumia, missä pelaaja on ollut osallisena. [14]

Jalkapallossa pelitapahtumakohtaisella datalla tarkoitetaan kaikkia tapahtumia, missä pallo on mukana. Perinteisesti pelitapahtumakohtaista tietoa on kerätty katsomalla videoita ja luokittelemalla tapahtumia. Tapahtuman luokittelemisella tarkoitetaan jonkin merkinnän liittämistä tapahtumaan. Merkinnällä voidaan ilmaista esimerkiksi, onko tapahtuma onnistunut vai epäonnistunut. Nykyään voidaan kerätä dataa pelitapahtumista joukkueen stadionilla sijaitsevan kiinteän asennuksen avulla, käyttäen korkearesoluutioisia kameroita. Keräämistä kutsutaan optiseksi keräämiseksi, joka ilmoittaa pelaajien ja pallon sijainnin useita kertoja sekunnissa.

Tämä on kuitenkin todella kallista ja on siksi vain käytössä korkeimmissa sarjoissa. [10]

Perinteisin tapa kerätä dataa on ottelupöytäkirjoista. Ottelypöytäkirjat sisältävät tietoa niin pelaajista kuin joukkueista ottelukohtaisesti. Jalkapallossa nämä tiedot voivat olla kokoonpanot, vaihdot, maalit ja syöttöpisteet. Pöytäkirjaan voidaan myös sisällyttää tarkempaakin dataa, kuten pallonhallintaprosentti tai onnistuneiden syöttöjen määrä. Verrattuna aiemmin luvussa mainittuihin keräämisen menetelmiin ottelupöytäkirjoja on pidetty jo vuosikymmenten ajan. [10]

2.3 Analyysimenetelmät

Tässä alaluvussa tarkastellaan erilaisia kerätyn datan analyysimenetelmiä. Menetelmän valinnassa on tärkeää tarkastella kriittisesti käsiteltävissä olevaa dataa. On tunnistettava, mitä riippuvuuksia aineistoissa on sekä mihin tarkoitukseen dataa halutaan käyttää. [10] Data-analytiikan menetelmillä etsitään hyödyllistä tietoa käytettävissä olevasta datasta, ja menetelmiä on kolmenlaisia: kuvailevia, ennakoivia ja ohjailevia. [15]

Kuvaileva analytiikka koostuu vaihtelevista tekniikoista, jotka edustavat nykyistä ja aiemmin tapahtunutta. Kuvaileva analytiikka on määritelty menneisyyden tarkastelemiseksi päätöksenteon tueksi. Kuvaileva analytiikka pyrkii vastaamaan kysymykseen: mitä on tapahtunut. [16]

Kuvailevassa analyysissä hyödynnetään monia tilastollisia mittareita kuten keskiarvoa, mediaania, moodia ja keskihajontaa. Tulokset esitetään usein graafisesti, kuten kaavioina ja kuvaajina. Klusterointi on eräs edistyneempi kuvailevan analyysin menetelmä. Sen tavoitteena on ryhmitellä samankaltaiset objektit. Tällöin samassa ryhmässä olevat objektit ovat ominaisuuksiltaan samankaltaisia ja vastaavasti eri ryhmissä olevat eroavat toisistaan. [15] Tämän avulla voidaan löytää samankaltaisuuksia datasta. Klusterointi on yksi ohjaamattoman oppimisen menetelmä.

Ennakoivassa analytiikassa pyritään havaitsemaan datasta toistuvia trendejä ja yhteyksiä. Ennakoiva analyysi tarkastelee tapahtumien juurisyitä esimerkiksi tiedonlouhintaa hyödyntäen. Ennakoivalla analyysilla pyritään ennustamaan tulevaisuuden tapahtumia. Ennakoiva analytiikka pyrkii vastaamaan kysymykseen: mitä voisi tapahtua. [16]

Ennakoiva analytiikka hyödyntää tilastollisia malleja tulevaisuuden tapahtumien ennustamiseen. Yleisimpiä ennustavia malleja ovat regressio, päätöspuut, neuroverkot, tukivektorikone (SVM) ja lähimmän naapurin luokitin. Regressio on eräs yleisimmistä ennakoivan analytiikan malleista. Se on laajalti käytössä analytiikassa, ja sen avulla voidaan ennustaa sekä jatkuvia että diskreettejä muuttujia. Regressiomallit ovat joustavia ja tarjoavat selityksen muuttujien vaikutuksesta lopputulokseen. [15] Regressio on yksi ohjatun oppimisen menetelmä.

Ohjaavaa analyysiä käytetään päätöksenteossa antamaan suosituksia mahdollisiin toimenpiteisiin. Ohjaavassa analyysissä hyödynnetään matemaattista ohjelmointia, heuristisia hakuja ja simulaatiomallinnuksia optimaalisten toimintatapojen löytämiseksi. [15] Ohjaava analyysi pyrkii olemaan tukena päätöksenteossa tarjoamalla ennusteiden perusteella suosituksia toimenpiteisiin. Ohjaava analyysi pyrkii vastaamaan kysymykseen: mitä meidän tulisi tehdä. [16]

3 Data-analytiikan sovellukset ammattijalkapallossa

Data-analytiikaa on mahdollista soveltaa monella eri tavalla ammattijalkapallossa. Tässä luvussa tarkastellaan data-analytiikan soveltamista ammattijalkapalloon kirjallisuuden pohjalta. Tavoitteena on selvittää, millaisia hyötyjä ja mahdollisuuksia data-analytiikka tarjoaa sekä millaisiin sovellusalueisiin sen käyttö keskittyy. Lisäksi käsitellään analytiikkaan liittyvä haasteita ja rajoituksia. Kirjallisuudessa havaitut sovellusalueet voidaan jakaa pelaajien sekä joukkuetason analyysiin. Taulukossa 3.1 on esitelty aineistot koskien pelaajien analyysiä ja taulukossa 3.2 on esitelty aineistot koskien joukkuetason analyysiä.

Aineisto	Sovellusalue	Käytetyt menetelmät
[10], [17], [18], [19]	Suorituskyvyn arviointi	Logistinen regressio, koneoppiminen, klusterointi, pelitapahtumakohtainen data, luokittelu, k-lähimmän naapurin menetelmä
[10], [20]	Pelaajien hankinta	Koneoppiminen, ennakoiva analyysi
[21]	Loukkaantumisriskin ennustaminen	XGBoost, sääntöpohjaiset mallit, GPS-seurantadata
[22]	Käyttäytymisen analyysi	Konenäkö, visualisointijärjestelmät

Taulukko 3.1: Pelaajien analyysiä koskevat aineistot

Aineisto	Sovelluskohde	Käytetyt menetelmät
[23], [24], [25], [26], [17]	Taktinen analyysi	Koneoppiminen, Voronoin diagrammi, seurantadata, aika- ja paikkatiedot, visuaalinen esitys
[19]	Hyökkäyspelaaminen	Koneoppminen, seurantadata
[27]	Syöttöjen analyysi	Optinen seuranta, koneoppiminen
[28]	Vastaprässin tunnistaminen	Koneoppiminen, XGBoost, seurantadata
[29], [17]	Ohjaava ja ennustava analytiikka päätöksenteon tueksi	Tekoäly, tilastollinen oppiminen, vahvistusoppiminen, peliteoria, konenäkö

Taulukko 3.2: Joukkuetason analyysiä koskevat aineistot

3.1 Pelaajien analyysi

Pelaajista mitataan ja kerätään jatkuvasti dataa harjoituksissa sekä otteluissa. Datan saatavuus on kasvanut paljon, ja yksi selittävä tekijä on edistykset pelaajien seuraamisessa. Suuri muutos tuli kun Kansainvälinen jalkapalloliitto (FIFA) antoi luvan käyttää langattomia sensoreita pelaajien toimintojen sekä paikan seurantaan. [23]

3.1.1 Data-analytiikka ja koneoppiminen pelaajien arvioinnissa

Koneoppimista voidaan hyödyntää data-analytiikassa ja se soveltuu myös hyvin ammattijalkapallon analysointiin. Koneoppimisen mallit voidaan jakaa yksittäisiin ja kaikkiin pelitilanteisiin keskittyviin malleihin. Yksittäisiä toimintoja, kuten syöttöä tai laukausta mittaavat mallit käyttävät usein odotusarvoa arvioidakseen toimintaa. Näitä malleja opetetaan tilastollisilla luokittelijoilla, kuten logistisella regressiolla. Harjoitusdata voi sisältää esimerkiksi laajan otannan aiemmin suoritetuista laukauk-

sista. Kaikkia tilanteita huomioivat mallit pyrkivät analysoimaan pelaajan kokonaisvaltaista vaikutusta otteluun. Mallit tarkastelevat pelitilan muuttumista pelaajien toimintojen perusteella. Jalkapallossa keskiössä ovat maalintekoon ja voittoon perustuvat lähestymiset. Esimerkiksi malli voisi tarkastella joukkueen mahdollisuutta voittaa ottelu, tietyssä pelin tilassa. [10]

Pelitapahtumakohtaista dataa hyödynnetään paljon pelaajien analysoinnissa. Siinä otetaan huomioon tapahtumia missä yksilö on ollut osallisena [17], [19]. Koneoppiminen on eräs tapa miten pelitapahtumakohtaista dataa voidaan käsitellä ja analysoida. Koneoppimisessa tapahtumakohtaista dataa voidaan hyödyntää pelaajien suorituksien analysoinnissa tarkastelemalla esimerkiksi syöttöjä tai maaleja. Lukuisia menetelmiä voidaan hyödyntää analysoinnissa. Herold ym. [19] kirjallisuuskatsauksessa yleisimmiksi menetelmiksi ilmenivät regressio, klusterointi, luokittelu ja k-lähimmän naapurin menetelmä. Menetelmät keskittyivät ohjaamattomaan ja ohjattuun oppimiseen. [19]

Uusien pelaajien hankinnassa voidaan hyödyntää data-analytiikkaa. Uusien lahjakkaiden pelaajien tunnistaminen on merkittävä osa joukkueen menestystä. Valmentajat arvioivat pelaajien potentiaalia ja hyödyntävät siihen useista eri lähteistä kerättyä dataa. Pelaajien hankinnassa dataa voidaan hyödyntää kuvailevaa, ennakoivaa ja ohjaavaa analytiikkaan. Raporttien ja kaavioiden avulla voidaan esimerkiksi kuvailla pelaajan taitoja. Lisäksi voidaan paikantaa optimaalinen pelipaikka pelaajalle sekä ennustaa pelaajan suorituskykyä tulevalla kaudella. [20] Koneoppimista voidaan soveltaa myös pelaajien hankintaan [10], [19]. Sen avulla voidaan ennustaa, kuinka hyvin pelaaja sopisi osaksi joukkuetta ja arvioimaan pelaajan tulevaa kehitystä. [10]

Data-analytiikan avulla voidaan ennustaa pelaajien kuormitusta sekä loukkaantumisen riskiä. Esimerkiksi huippuseurat kuten Arsenal ja Manchester City käyttävät data-analytiikkaa loukkaantumisien ehkäisemisessä [20]. Pelaajien loukkaantu-

misriskiä on mahdollista eunnustaa tarkastelemalla yli- tai aliharjoittelua [21]. Tutkimuksessa mallit keskittyivät erityisesti kontaktittomien loukkaantumisien ehkäisemiseen. Datana hyödynnettiin GPS-seurantalaitteiden avulla kerättyä dataa pelaajista. Tutkijat loivat kolme erilaista mallia, kaksi sääntöpohjaista mallia ja yhden koneoppimismenetelmän käyttäen XGBoost-algoritmia. XGBoost on algoritmi, joka yhdistää useita heikkoja malleja kuten päätöspuita ja optimoi niitä tarkkuuden parantamiseksi. XGBoostin avulla loukkaantumisia pystyttiin ennustamaan jopa 90 % tarkkuudella. Mallia voidaan hyödyntää etenkin harjoitusohjelman muokkaamisessa esimerkiksi vähentämällä tai lisäämällä kuormitusta pelaajille. [21] Jatkuva kuormituksen lisääntyminen pelaajille on tällä hetkellä ammattijalkapallossa ajankohtainen keskustelun aihe. Tutkimuksessa kehitettyä mallia voitaisiin hyvin hyödyntää joukkueen suorituskyvyn ylläpitämiseksi säätelemällä pelaajien kuormituksia.

Pelaajien käyttäytimistä kentällä tarkastellaan jatkuvasti. Yksi tapa analysoida pelaajien toimintaa on visualisointi, jossa hyödynnetään pelaajista kerättyä dataa. Tutkimuksessa kehitettiin järjestelmä, joka tunnistaa pelaajien tekemät virheelliset juoksut ja tämän avulla suorittaa "mitä-jos" (engl. *what-if*) analyysiä. Järjestelmän toiminta perustuu pelaajista kerättyihin x- ja y-koordinaatteihin ja tutkimuksessa käytettiin sääntöpohjaista lähestymistapaa. Paikkadatan avulla voitiin tunnistaa vapaa tila, johon liike olisi voitu tehdä, sekä havaita tilanteessa olevat pelaajat ja virheelliset liikkeet. Järjestelmää voidaan käyttää esimerkiksi valmennuksen tukena näyttämällä optimaalisimmat juoksulinjat tietyissä tilanteissa. Esimerkiksi toistuvia virheellisiä juoksuja voidaan korjata valmennuksen avulla analysoimalla järjestelmän tuottamaa tietoa ja hyödyntämällä sitä harjoituksien suunnittelussa. [22]

3.1.2 Suorituksien mittarit

Pelaajien suorituskykyä voidaan arvioida useilla eri mittareilla. Tunnetuimmat ammattijalkapallossa olevat mittarit ovat maaliodottama eli xG (engl. *expected goals*)

ja syöttöodottama eli xA (engl. *expected assists*). Maaliodottama määrittelee todennäköisyyden maalin syntymiselle. Sen avulla voidaan kuvata tilannetta tai koko ottelua. Syöttöodottama puolestaan laskee todennäköisyyden maaliinjohtavan syötön mahdollisuudelle. Mittarin tehtävänä on tarjota tietoa mitä aikaisemmin ei ollut tarjolla. Merkittävä etu esimerkiksi xG:llä on, että se kertoo pelin todellisesta kulusta paljon enemmän kuin pelin tulos. Jalkapallossa syntyy vähän maaleja, joten mittareilla kuten xG ja xA on mahdollista analysoida pelin todellista luonnetta paremmin. [10], [17] Satunnaisuudella on vaikutus ottelun lopputulokseen, jonka vuoksi esimerkiksi maaliodottamaa pidetään parempana pelin kulun mittarina. Maaliodottama sisältää vähemmän satunnaisuutta verrattuna syntyneisiin maaleihin. [18] Tästä syystä esimerkiksi joukkueen tehokkuutta pallon takaisin voittamisen jälkeen arvioitiin perustuen maaliodottamaan syntyneiden maalien sijasta [28].

Koneoppimista voidaan hyödyntää myös mittareiden luomisessa [10]. Maaliodottamaa laskettaessa tulee ottaa huomioon monia eri tekijöitä. Brechot ja Flepp [18] tutkimuksessaan laski maaliodottamia pääsarjoista ja otokseen sisältyi yhteensä 170 688 laukausta 7 304 ottelusta. Tutkimuksessa otettiin huomioon yksittäisen laukauksen sijainti, laukauksen tyyppi sekä se kehon osa, jolla laukaus suoritettiin. Laukausten todennäköisyydet lasketaan yhteen summaamalla, mistä muodostuu ottelun lopullinen maaliodottama. Vastaavasti voidaan laskea päästettyjen maalien odottama. Tutkimuksessa nostettiin esille, että maaliodottamaa laskiessa tulisi myös ottaa huomioon puolustajien antamaa painetta, yksittäisen pelaajan maalintekokykyä, maalivahdin taitoa ja pelaajan liikettä tilanteeseen. Näitä ei kuitenkaan laskelmassa otettu huomioon sillä riittävän tarkkaa dataa ei ollut saatavilla tutkimuksen aikana.

3.2 Joukkuetason analyysi

Pelaajien analysoinnin lisäksi on tärkeää analysoida joukkueen kokonaisvaltaista toimintaa. Tässä alaluvussa tarkastellaan joukkuetason analyysiä. Ensin tarkastellaan

taktiikan analyysiä ja optimointia. Tämän jälkeen esitellään kirjallisuudessa käytettyjä menetelmiä joukkuetason analyysiin.

3.2.1 Taktiikan analyysi ja optimointi

Joukkueiden taktiikoiden tunnistaminen on monimutkainen tehtävä ammattijalkapallossa. Taktiikalla tarkoitetaan harkittujen toimintojen toteuttamista tietyn tavoitteen saavuttamiseksi. Jalkapallossa tämä tavoite on selkeä, ottelun voittaminen. Oikean taktiikan valitseminen on merkittävä tekijä lopputuloksen kannalta. Taktiikoiden analyysi vaatii dataa useista eri lähteistä. Taktiikoiden analyysissä tulee ottaa huomioon pelaajien tekniset taidot, pelaajien suorituskyvyt ja joukkueiden muodostelmat. [23]

Taktiikoiden analysoimiseen on useita eri menetelmiä. Erilaisten pelityylien analysointi on tärkeää otteluun valmistautumisessa. Joukkueiden keskipisteen arviointi nostettiin mielenkiintoiseksi tarkastelunkohteeksi. Keskipisteen analysoinnissa tarkastellaan joukkueen geometrista keskipistettä, joka lasketaan kaikkien pelaajien sijaintien perusteella. Sen avulla voidaan arvioida koko joukkueen käyttäytymistä kentällä. Myös joukkueen syöttöjen käyttäytymistä on tutkittu esimerkiksi verkostoanalyysin avulla, jossa pelaajat mallinnetaan solmuiksi ja pelaajien väliset syötöt painotetuiksi pisteiksi. Verkostoanalyysistä voidaan helposti havaita avainpelaajat, sillä heillä on enemmän yhteyksiä muihin pisteisiin. Lisäksi monia koneoppimisen malleja on käytetty taktiikoiden analysoimisessa kuten muodostelmien automaattisessa tunnistamisessa. [23]

Seurantadatalla on merkittävä rooli taktiikoiden analysoinnissa [17], [26]. Hyökkäyspelaaminen on yksi jalkapallon keskeisimmistä osa-alueista. Koneoppimista ja seurantadataa hyödyntämällä voidaan analysoida hyökkäyspelin tehokkuutta ja edistää sitä. Seurantadatan käyttö voidaan jakaa neljään pääalueeseen: yksilötason syöttöanalyysiin, joukkuetason syöttöanalyysiin, aika- ja paikkatiedon hyödyntämiseen

maalipaikkojen luomisessa sekä pallon takaisin voittamisen analysointiin. Seurantadatan avulla voidaan tehdä hyödyllisiä löydöksiä joukkueiden välillä. Voidaan esimerkiksi vertailla vastustajia ja tehdä datan perusteella ratkaisevia päätöksiä taktiikan valitsemisessa. [19]

Massadatan myötä ammattijalkapallon analysoinnissa ollaan alettu hyödyntämään tekoälyn menetelmiä, jotka voidaan jakaa kolmeen osa-alueeseen: tilastolliseen oppimiseen, peliteoriaan ja konenäköön. Tilastollisessa oppimisessa pyritään tunnistamaan datasta ominaisuuksia, jotka kuvaavat esimerkiksi pelaajien ja joukkueiden pelityylejä. Peliteoriassa tutkitaan pelaajien taktista päätöksentekoa erilaisissa tilanteissa. Näitä menetelmiä voidaan erityisesti hyödyntää ohjaavassa ja ennustavassa analytiikassa. Menetelmien yhteisenä tavoitteena on tarjota valmentajille ja pelaajille tukea päätöksentekoon. Vaikka jokainen näistä ovat yksinään merkittäviä, niin todellinen hyöty saadaan menetelmiä yhdistämällä. Tilastollisen oppimisen ja peliteorian yhdistämisellä voidaan analysoida esimerkiksi joukkueen muodostelmia ja tämä auttaa pelaajien sekä muodostelman valitsemisessa. Konenäön ja tilastollisen oppimisen yhistämisellä voidaan puolestaan erottaa videosta hyvinkin vaikeaa informaatiota ja hyödyntää sitä ennustavien mallien luomisessa. [29]

Valmennukselle päätöksenteon arviointi on elintärkeä tehtävä. Koneoppimismenetelmien kuten vahvistusoppimisen avulla voidaan tarjota reaaliaikaista päätöksenteon tukea valmennukselle. Nykyisissä vahvistusoppimisen malleissa pelikenttä ja pelaajien sijainnit esitetään kaksiulotteisina matriiseina, mikä mahdollistaa niiden käsittelyn konvoluutioneuroverkkojen avulla havaintojen saamiseksi. Mallin avulla voidaan saada tarkka kuva tilanteesta kentällä. [17]

3.2.2 Menetelmiä joukkuetason analyysiin

Otteluanalyytikot keskittyvät erityisesti joukkeen kykyyn voittaa pallo takaisin sekä siihen, mitä tapahtuu pallonriiston jälkeen. Pallon takaisin voittaminen perustuu

vahvasti vastustajan analysointiin ja oman taktiikan suunnitteluun. Vastaprässin (engl. *counterpressing*) luomiseen on olemassa lukuisia erilaisia lähestymistapoja. Bauer ja Anzer [28] tekemässä tutkimuksessa luotiin koneoppimismalli, joka pystyi analysoimaan joukkueen kykyä voittaa pallo takaisin. Tutkimuksessa hyödynnettiin XGBoost-algoritmia. Harjoitusdatan muodostamisessa hyödynnettiin manuaalista luokittelua sekä asiantuntijoiden määrittelemiä ominaispiirteitä vastaprässille. Tutkimuksessa hyödynnettiin seurantalaitteiden avulla kerättyä dataa, kuten optisten seurantajärjestelmien tuottamaa dataa. Yhteensä tarkasteltavia otteluita oli 4 118.

Bauer ja Anzer [28] totesivat, että malli pystyi tunnistamaan vastaprässin tilanteita riittävällä tarkkuudella. Havaittiin, että maalin todennäköisyys kasvaa merkittävästi, mikäli vastaprässi onnistuu. Sen sijaan, jos vastaprässi epäonnistuu, maalin päästämisen todennäköisyys kasvaa. Malli myös pystyi tunnistamaan optimaalisen paikan riistää pallo, mikä oli sivurajan lähettyvillä vastustajan kenttäpuoliskolla. Mallilla on moniakin hyödyllisiä käyttökohteita. Mallin avulla voidaan havaita samankaltaisuuksia joukkueissa sekä analysoida miten pallo voidaan voittaa takaisin mahdollisimman nopeasti ja tehokkaasti. Mallin avulla saatiin informaatiota, mitä ei ollut mahdollista saada perinteisillä menetelmillä. Otteluiden analysointiin, kuten vastustajien taktiikoiden ja tottumusten tunnistamiseen, käytetään yleensä runsaasti resursseja ja aikaa. Merkittävä hyöty koneoppimisen soveltamisesta ammattijalkapalloon on ajansäästäminen ja suuremman otoksen tarkastelu. [28]

Syöttö voidaan luokitella jalkapallon yhdeksi tärkeimmäksi elementiksi. Syöttö on yksi ottelun aikana yleisimmin esiintyvistä tapahtumista. Tämä tekee siitä merkittävän analyysin kohteen. [17] Syötön tehokkuutta voidaan tutkia data-analytiikan avulla. Goes ym. [27] esittelivät tutkimuksessaan menetelmän, jolla tehokkuutta voidaan analysoida. Tutkimus keskittyi vastustajan muodostelmaan ja pelaajien liikkeisiin. Datana käytettiin seurantadataa, jonka saatiin optisen keräämisen avulla. Mitä enemmän syöttö liikuttaa vastustajan pelaajia sekä muodostelmaa sitä tehokkaam-

pana malli pitää syöttöä. Mallin avulla voidaan tehdä hyvinkin yksityiskohtaista analyysiä joukkueen syöttöjen tehokkuudesta, mikä vaikuttaa paljon taktiikan valitsemisessa. Syöttöjä arvioimalla voidaan optimoida joukkueen taktiikkaa. Mallin avulla voidaan tunnistaa syöttöjä, jotka saavat vastustajan liikkumaan eniten, mikä puolestaan voi johtaa vastustajan väsymiseen. [27] Mallia voidaan hyvin soveltaa taktiikan optimoimisessa syöttöjen tehokkuuden lisäämiseksi.

Aika- ja paikkatietoja hyödynnetään taktiikoiden analysoinnissa, erityisesti pelitapahtumakohtaisessa datassa. Näiden tietojen avulla voidaan luoda visuaalisia otteluanalyysejä, kuten lämpökarttoja, jotka kuvaavat pelaajien tai joukkueen liikkumista kentällä. Lämpökartta on visuaalinen esitys koko kentästä, missä pelaajien toimintaa voidaan havainnollistaa väreillä. Lämpökartta tarjoaa kattavan kuvan pelin tapahtumista ja voi paljastaa valmentajalle, kuinka taktiikat muuttuvat ottelun aikana. Lämpökartasta on mahdollista havaita vastustajajoukkueen taktiikoiden muutoksia. Lisäksi pelaajien liikeratoja on mahdollista myös mallintaa visuaalisesti. Tämän avulla voidaan tutkia pelaajien tekemiä juoksuja ja optimoida niitä. [24]

Toinen esimerkki aika- ja paikkatietojen soveltamisesta jalkapallossa on Higgins ym. [25] tekemä tutkimus, jossa tarkasteltiin kentän hallinnan arviointia. Tutkimuksessa hyödynnettiin Voronoin diagrammia kentän jakamiseen eri alueisiin. Kentän hallinnan arviointi keskittyy alueiden tunnistamiseen, joilla yksittäinen pelaaja voi kontrolloida palloa. Mallista on mahdollista havaita alueita, missä syntyy eniten maaleja tai luodaan eniten maalipaikkoja. Tutkijat havaitsivat, että maalien määrän ja kentän hallinan välillä oli vain heikko positiivinen korrelaatio.

3.3 Haasteet ja rajoitukset

Data-analytiikan soveltamiseen ammattijalkapalloon liittyy monia haasteita. Kirjallisuudessa nousi vahvasti esille kuinka jalkapallo on yksi haastavimmista ammattiurheilun analyysin kohteista. Tämän alaluvun tarvoitteena on esitellä kirjallisuudessa

mainittuja haasteita ja rajoitteita.

Ammattijalkapallossa tapahtumiin vaikuttavia tekijöitä on paljon. Massadatan myötä analyysin tekemiseen on tullut mukaan lukuisia haasteita. Taustatekijöiden suuri määrä ja monimutkaisuus muodostaa yhden keskeisen haasteen analyysissä. Ottelun aikana muuttuvia tekijöitä on paljon ja otteluun osallistuu samanaikaisesti 22 pelaajaa, mikä tekee ottelun analysoimisesta erityisen haastavaa [29]. Perinteisimmät mittarit, kuten keskiarvo eivät ota taustatekijöitä huomioon analyysissä [23].

Jokaisen vaikuttavan tekijän ottaminen huomioon analyysissa on vaikeaa [17]. Esimerkiksi aikaisemmin esitelty malli maaliodottaman laskemiseen ei ottanut huomioon laskelmissa jokaista maaliin vaikuttavaa tekijää. Tämä voi aiheuttaa mallin ali- tai ylisuorittamista, joka johtaa epäluotettaviin tuloksiin. [18] Myös loukkaantumisriskien ennustaminen oli haastavaa, sillä kaikkia ihmisten biologisia eroavaisuuksia on vaikea ottaa huomioon. [21]

Analyysiä tehdessä on hankala arvioida tapahtumien juurisyitä. Tämä aiheuttaa haasteita tapahtumien luokittelussa. Selitys siitä miksi jokin tapahtuma epäonnistui voi usein jäädä epäselväksi. Tilanne, jossa syöttö menee vastustajalle, voitaisiin luokitella vain huonoksi syötöksi. Tämän taustalla voi kuitenkin olla toisen pelaajan huono juoksu, joka ei enään kertoisi syötön laadusta vaan toisen pelaajan huonosta juoksusta. [10]

Pelitapahtumakohtainen data on subjektiivista, jonka keräämistä tavallisesti suorittavat ihmiset, jotka ovat alttiita virheille. Tämä voi johtaa virheellisen datan tulkintaan. Epäonnistunut keskitys, joka johtaakin maaliin saatetaan luokitella laukaisuksi. [10] Virheellisesti luokiteltu data voi aiheuttaa ongelmia koneoppimismallien opettamisessa, sillä se voi johtaa mallin epätarkkuuteen ja virheellisiin tuloksiin.

Otoskokoon liittyy haasteita ja rajoitteita. Joukkeet pelaavat vain rajoitetun määrän otteluita yhdessä kaudessa. Jokainen pelaaja ei pelaa jokaista ottelua, joten

dataa kyseisen pelaajan pelaamisesta on vain rajoitettu määrä. Datan vanheneminen myös liittyy otoskoon haasteisiin. Jalkapallossa taktiikat ja pelaajat vaihtuvat jatkuvasti. Dataa, joka on kerätty aikaisemmilta kausilta ei usein enään pidetä olennaisena tietona. [10]

Otoksen keräämiseen liittyy myös haasteita. Esimerkiksi pelaajan ja maalivahdin henkilökohtaisten taitojen määrällistä arviointia, eli niiden muuttamista luvuiksi, on vaikea toteuttaa tarkasti. On myös havaittu, että useasti otoksissa on riittämätön määrä laukauksia erottaakseen yksilön taidon ja satunnaisuuden. [18] Nopeat tilanteet, missä esiintyy ennalta arvaamattomia liikkeitä aiheuttavat haasteita analyysiä suorittavalle järjestelmälle tai ammattilaiselle. Ongelmana on liikettä seuraavien kameroiden antaman datan epätarkkuus. [19] Kameroiden datan keräämiseen liittyy ongelmia luotettavuudessa. Kamerat keskittyvät kentällä kiinnostavimpaan kohteeseen, minkä vuoksi useita pelaajia jää kameran kuvan ulkopuolelle. Toinen haaste on se kuinka pelaajat usein peittävät toisiaan, mikä tekee pelaajien seuraamisesta ja tunnistamisesta haastavaa. [29]

Ammattijalkapallossa kerätyn datan analysoinnissa on myös riskinsä datan puolueellisuudelle. Puolueellisuutta aiheuttaa esimerkiksi hyvät pelaajat tai joukkueet, jotka suorittavat enemmän tapahtumia verrattuna muihin pelaajiin. Joten myös kerätyssä datassa on enemmän informaatiota paremmista pelaajista. [10] Syöttöjen tehokkuuden analyysissä nostettiin esille kuinka data oli kerätty pääasiassa yhdestä joukkueesta, joka voi aiheuttaa puolueellisuutta datassa ja tuloksissa [27].

Kirjallisuudessa korostuu haasteet datan tulkinnassa, erityisesti siinä, kuinka havainnot voidaan esittää valmennukselle ymmärrettävässä muodossa. Edistyneempien koneoppimismenetelmien, kuten neuroverkkojen monimutkaisuus tekee tulkinnasta haastavaa. Ongelmana on tunnistaa, mitkä tekijät vaikuttavat lopputuloksiin, jolloin hyödyllisen ja konkreettisen tiedon tarjoaminen valmennukselle vaikeutuu. [19] Havaintojen esittämistä vaikeuttaa myös yhteistyön puute tietojenkäsittelijöiden ja

valmennuksen välillä [19], [23], [20].

Dataa kerätään useista eri lähteistä, mikä aiheuttaa haasteita tietojen integroinnissa. Lisäksi massadata on vaikeuttanut datan varastointia ja saatavuutta [23]. Esimerkiksi aiemmin esitellyssä mallissa yksi maaliodottaman laskemiseen liittyvä ongelma oli datan saatavuus. Laskelmassa ei voitu ottamaan huomioon kaikkia mahdollisia taustatekijöitä sillä ei ollut saatavilla riittävän tarkkaa dataa. [18] Useista eri datalähteistä koottu tieto luo erityisiä haasteita myös jalkapallossa käytettäville tietojärjestelmille [26].

4 Yhteenveto

Data-analytiikka on noussut merkittäväksi osaksi urheilua ja viime vuosien aikana myös ammattijalkapalloa. Data-analytiikka tarjoaa merkittäviä hyötyjä joukkueille ja pelaajille. Sen tavoitteena on parantaa joukkeen ja pelaajien suorituskykyä. Data-analytiikan hyödyntämiseen liittyy kuitenkin myös haasteita, jotka voivat aiheuttaa ongelmia soveltamisessa. Ammattijalkapallossa data-analytiikan soveltamiseen on useita eri mahdollisuuksia ja menetelmiä. Tässä tutkielmassa tavoitteena oli kirjallisuuskatsauksen perusteella saada kattava käsitys data-analytiikan sovellusalueista ammattijalkapallossa, sen tuomista hyödyistä ja haasteista.

Tutkimuskysymys 1 kysyi: *Mitkä ovat data-analytiikan sovellusalueet ja hyödyt ammattijalkapallossa?* Data-analytiikkaa voidaan hyödyntää ammattijalkapallossa sekä pelaajien että joukkuetason analyysissä. Pelaajien osalta keskeisiä sovellusalueita ovat, suorituskyvyn analyysi, pelaajien hankinta, loukkaantumisriskin ennustaminen ja käyttäytymisen analyysi. Joukkuetason sovellusalueita ovat taktinen analyysi, syöttöjen ja hyökkäyspelaamisen analyysi sekä vastaprässin tunnistaminen. Molemmilla tasoilla koneoppiminen oli vahvasti esillä. Pelaajien analyysiin verrattuna joukkuetason analyysissä hyödynnettiin enemmän seurantadataa eli tietoa pelaajien ja pallon paikasta ottelun tai harjoituksien aikana. Data-analytiikan keskeisin hyöty on uuden tiedon saaminen. Data-analytiikan avulla saadaan informaatiota, jota ei aikaisemmin ollut saatavilla, mikä tukee harjoituksien optimointia, päätöksentekoa ja toiminnan kehittämistä. Erityisesti valmennus hyötyy edistyneistä analytii-

kan menetelmistä, jotka säästävät aikaa ja mahdollistavat syvällisemmän pelaaja- ja joukkuetason analyysin.

Tutkimuskysymys 2 kysyi: *Mitkä ovat data-analytiikan haasteet ammattijalkapallossa?* Ammattijalkapallo luokiteltiin yhdeksi vaikeimmaksi analysoitavaksi kohteeksi muihin joukkuelajeihin verrattuna. Data-analytiikalle erityisesti asettaa haasteita useat muuttuvat tekijät, nopeat tilanteet sekä kentän iso koko. Merkittäviä haasteita nousi esille otoksen keräämisessä ja luokittelussa, datan puolueellisuudessa ja kaikkien vaikuttavien tekijöiden huomioon ottamisessa. Lisäksi datan tulkinnan ongelmat nousi keskeiseksi haasteeksi. Tietojenkäsittelijöiden ja valmennustiimin välillä on vielä suuri kommunikaation puute, joka on esteenä data-analytiikan suurimman potentiaalin saavuttamisessa. Haasteena oli myös yksinkertaisen ja ymmärrettävän datan esittäminen valmennustiimille.

Kirjallisuuskatsaus rajattiin koskemaan ainoastaan aineistoja, jotka käsittelivät nimenomaan ammattijalkapalloa. Tarkoituksena ei ollut tarkastella data-analytiikan kaikkia mahdollisia sovellusalueita, vaan tarjota kattava kuva sen mahdollisuuksista olemassa olevan tutkimuksen perusteella. Rajauksen vuoksi aineistoja, jotka keskittyivät yksinomaan jalkapallon yleiseen analyysiin, ei otettu mukaan. Tämä tarkoittaa, että joitakin mahdollisia sovellusalueita on saattanut jäädä tarkastelun ulkopuolelle. Erikoistilanteita käsitteleviä tutkimuksia ei löytynyt, vaikka se on merkittävä analytiikan alue, mikä saattaa johtua rajauksesta. Jatkotutkimuksia ajatellen tutkielmaa voisi laajentaa käsittelemään jalkapalloa kokonaisuutena ja sisällyttää myös harrastus ja puoliammattilaisen jalkapallon. Data-analytiikan haasteisiin ammattijalkapallon analysoinnissa tulisi myös kiinnittää enemmän huomiota, erityisesti tutkien mahdollisia ratkaisuja esiteltyihin ongelmiin.

Tämä tutkielma osoitti, että data-analytiikan soveltamisella ammattijalkapalloon on olemassa merkittävä potentiaali joukkueille sekä pelaajille, erityisesti suorituksien optimoinnissa. Samalla tuotiin esille data-analytiikan käyttöön liittyviä

haasteita, kuten datan tulkintaan ja otokseen liittyvät ongelmat. Tulosten perusteella data-analytiikan mahdollisuuksia ei ole vielä hyödynnetty täyteen potentiaaliin asti. Data-analytiikan merkitys alalla kasvaa jatkuvasti, mutta haasteisiin on kehitettävä ratkaisuja, jotta voidaan saavuttaa data-analytiikan todellinen potentiaali ammattijalkapallon kontekstissa.

Lähdeluettelo

[1] O. Olaniyan, B. Dehe, D. Bamford ja S. Ward, ”Enhancing On-Pitch Learning Capabilities with Data Analytics and Technologies in Elite Sports”, *European Sport Management Quarterly*, vol. 24, nro 4, s. 1195–1214, 2023. DOI: 10.1080/16184742.2023.2270565.

[2] E. Morgulev, H. Azar O ja R. Lidor, ”Sports analytics and the big-data era”, *International journal of data science and analytics*, vol. 5, nro 4, s. 213–222, 2018. DOI: 10.1007/s41060-017-0093-7.

[3] G. Liu, Y. Luo, O. Schulte ja T. Kharrat, ”Deep Soccer Analytics: Learning an Action-Value Function for Evaluating Soccer Players”, *Data Mining and Knowledge Discovery*, vol. 34, nro 5, s. 1531–1559, 2020. DOI: 10.1007/s10618-020-00705-9.

[4] O. Müller, A. Simons ja M. Weinmann, ”Beyond Crowd Judgments: Data-driven Estimation of Market Value in Association Football”, *European Journal of Operational Research*, vol. 263, nro 2, s. 611–624, 2017. DOI: 10.1016/j.ejor.2017.05.005.

[5] D. Chong ja H. Shi, ”Big Data Analytics: A Literature Review”, *Journal of Management Analytics*, vol. 2, nro 3, s. 175–201, 2015. DOI: 10.1080/23270012.2015.1082449.

[6] C.-W. Tsai, C.-F. Lai, H.-C. Chao ja A. V. Vasilakos, ”Big Data Analytics: A Survey”, *Journal of Big Data*, vol. 2, nro 21, 2015. DOI: 10.1186/s40537-015-0030-3.

[7] S. Mittal ja O. P. Sangwan, ”Big Data Analytics Using Machine Learning Techniques”, teoksessa *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2019, s. 203–207. DOI: 10.1109/CONFLUENCE.2019.8776614.

[8] C. Janiesch, P. Zschech ja K. Heinrich, ”Machine Learning and Deep Learning”, *Electronic Markets*, vol. 31, s. 685–695, 2021. DOI: 10.1007/s12525-021-00475-2.

[9] B. Zhongbo ja B. Xiaomei, ”Sports Big Data: Management, Analysis, Applications, and Challenges”, *Complexity*, vol. 2021, nro 6676297, s. 11, 2021. DOI: 10.1155/2021/6676297.

[10] J. Davis, L. Bransen, L. Devos et al., ”Methodology and evaluation in sports analytics: Challenges, approaches, and lessons learned”, *Machine Learning*, vol. 113, s. 6977–7010, 2024. DOI: 10.1007/s10994-024-06585-0.

[11] A. Ç. Seçkin, B. Ateş ja M. Seçkin, ”Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities”, *Applied Sciences*, vol. 13, nro 18, 2023. DOI: 10.3390/app131810399.

[12] V. De Silva, M. Caine, J. Skinner et al., ”Player Tracking Data Analytics as a Tool for Physical Performance Management in Football: A Case Study from Chelsea Football Club Academy”, *Sports*, vol. 6, nro 4, s. 130, 2018. DOI: 10.3390/sports6040130.

[13] Y. Matsuzaka ja R. Yashiro, ”AI-Based Computer Vision Techniques and Expert Systems”, *AI*, vol. 4, nro 1, s. 289–302, 2023. DOI: 10.3390/ai4010013.

[14] B. T. Naik, M. F. Hashmi ja N. D. Bokde, ”A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions”, *Applied Sciences*, vol. 12, nro 9, 2022. DOI: 10.3390/app12094429.

[15] L. Duan ja Y. Xiong, ”Big Data Analytics and Business Analytics”, *Journal of Management Analytics*, vol. 2, nro 1, s. 1–21, 2015. DOI: 10.1080/23270012.2015.1020891.

[16] M. K. Saggi ja S. Jain, ”A Survey towards an Integration of Big Data Analytics to Big Insights for Value-Creation”, *Information Processing & Management*, vol. 54, nro 5, s. 758–790, 2018. DOI: 10.1016/j.ipm.2018.01.010.

[17] Z. Pu, Y. Pan, S. Wang et al., ”Orientation and Decision-Making for Soccer Based on Sports Analytics and AI: A Systematic Review”, *IEEE/CAA Journal of Automatica Sinica*, vol. 11, nro 1, s. 37–57, 2024. DOI: 10.1109/JAS.2023.123807.

[18] M. Brechot ja R. Flepp, ”Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals”, *Journal of Sports Economics*, vol. 21, nro 4, s. 335–362, 2020. DOI: 10.1177/1527002519897962.

[19] M. Herold, F. Goes, S. Nopp, P. Bauer, C. Thompson ja T. Meyer, ”Machine learning in men’s professional football: Current applications and future directions for improving attacking play”, *International Journal of Sports Science & Coaching*, vol. 14, nro 6, s. 798–817, 2019. DOI: 10.1177/1747954119879350.

[20] R. de la Torre, L. O. Calvet, D. Lopez-Lopez, A. A. Juan ja S. Hatami, ”Business Analytics in Sport Talent Acquisition: Methods, Experiences, and Open Research Opportunities”, *International Journal of Business Analytics (IJBAN)*, vol. 9, nro 1, s. 1–20, 2022. DOI: 10.4018/IJBAN.290406.

[21] T. Piłka, B. Grzelak, A. Sadurska, T. Górecki ja K. Dyczkowski, ”Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors”, *Sensors*, vol. 23, nro 3, 2023. DOI: 10.3390/s23031227.

[22] M. Stein, D. Seebacher, R. Marcelino et al., ”Where to Go: Computational and Visual What-If Analyses in Soccer”, *Journal of Sports Sciences*, vol. 37, nro 24, s. 2774–2782, 2019. DOI: 10.1080/02640414.2019.1652541.

[23] R. Rein ja D. Memmert, ”Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science”, *SpringerPlus*, vol. 5, nro 1410, 2016. DOI: 10.1186/s40064-016-3108-2.

[24] V. Machado, R. Leite, F. Moura, S. Cunha, F. Sadlo ja J. L. Comba, ”Visual Soccer Match Analysis Using Spatiotemporal Positions of Players”, *Computers & Graphics*, vol. 68, s. 84–95, 2017. DOI: 10.1016/j.cag.2017.08.006.

[25] L. Higgins, T. Galla, B. Prestidge ja T. Wyatt, ”Measuring the Pitch Control of Professional Football Players Using Spatiotemporal Tracking Data”, *Journal of Physics: Complexity*, vol. 4, nro 2, 2023. DOI: 10.1088/2632-072X/acb67d.

[26] F. Goes, L. Meerhoff, M. Bueno et al., ”Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review”, *European Journal of Sport Science*, vol. 21, nro 4, s. 481–496, 2021. DOI: 10.1080/17461391.2020.1747552.

[27] F. R. Goes, M. Kempe, L. A. Meerhoff ja K. A. P. M. Lemmink, ”Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches”, *Big Data*, vol. 7, nro 1, s. 57–70, 2019. DOI: 10.1089/big.2018.0067.

[28] P. Bauer ja G. Anzer, ”Data-Driven Detection of Counterpressing in Professional Football: A Supervised Machine Learning Task Based on Synchronized Positional and Event Data with Expert-Based Feature Extraction”, *Da-*

ta Mining and Knowledge Discovery, vol. 35, nro 5, s. 2009–2049, 2021. DOI: 10.1007/s10618-021-00763-7.

[29] K. Tuyls, S. Omidshafiei, P. Muller et al., ”Game Plan: What AI Can Do for Football, and What Football Can Do for AI”, *Journal of Artificial Intelligence Research*, vol. 71, s. 41–88, 2021. DOI: 10.1613/jair.1.12505.