TURUN YLIOPISTO
Kauppakorkeakoulu

DATAN JA DATAMALLIEN HYÖDYNTÄMINEN ALIHINNOITELTUJEN HUIPPUJALKAPALLOILIJOIDEN LÖYTÄMISESSÄ JA HANKKIMISESSA

Tietojärjestelmätieteen kandidaatintutkielma

Laatija:
Ilmari Ollila

Ohjaaja:
Yliopisto-opettaja Kaisa Kukkonen

5.5.2025
Turku

Kanditutkielma

Oppiaine: Tietojärjestelmätiede
Tekijä: Ilmari Ollila
Otsikko: Datan ja datamallien hyödyntäminen alihinnoiteltujen huippujalkapalloilijoiden löytämisessä ja hankkimisessa
Ohjaaja: Yliopisto-opettaja Kaisa Kukkonen
Sivumäärä: 38 sivua
Päivämäärä: 5.5.2025

Nykypäivän jalkapallossa yksi maali saattaa merkitä miljoonien eurojen tuloja, tai tulojen menetystä. Seurat maksimoivat urheilullista hyötyä talouden kustannuksella ja monet seurat ovat ajautuneet taloudellisiin vaikeuksiin huonon pelaajanhankintastrategian takia.

Pelaajat ovat oleellinen osa seurojen lisäarvon tuottoa. Henkilöstöresurssina jalkapallopelaajat ovat nousseet miljoonien eurojen arvoisiksi. Pelaajamarkkinat ovat äärimmäisen kilpailulliset. Urheilujohtajien tavoitteena on vuosia ollut urheilullisen menestyksen saavuttaminen mahdollisimman pienellä taloudellisella panostuksella. Data-analytiikka antaa seuroille mahdollisuuden löytää pelaajamarkkinoilta kyvykkäitä pelaajia, jotka eivät ole tunnettuja tai kalliita. Data-analytiikalla onkin pelaajanhankinnassa paljon historiaa muissa urheilulajeissa. Jalkapallossa data-analytiikka on jäänyt esimerkiksi baseballia ja koripalloa jälkeen sen kaoottisen ja dynaamisen luonteen vuoksi. Jalkapallon kautta voidaan tarkastella teknologian kehityksen mahdollistamaa data-analytiikalla saatavaa lisähyötyä. Koneoppiminen ja data-analytiikan kehittyminen ovat antaneet mahdollisuuden jalkapallolle päästä muiden urheilulajien tasolle data-analytiikan hyödyntämisessä. Kehittyneet data- ja hinnoittelumallit mahdollistavat seurojen tehokkuuden lisäämisen pelaajamarkkinoilla.

Avainsanat: data-analytiikka, jalkapallo, urheiluanalytiikka, koneoppiminen

Sisällysluettelo

1 Johdanto 7

2 Seuran urheilulliseen menestykseen vaikuttavat pelaajamittarit 9

2.1 Pelaajat menestyksen mahdollistajina 9

2.2 Pelaajien hankkiminen dataa hyödyntäen 10

2.3 Pelaajiin sovellettavat datamallit 11

2.3.1 Koneoppimisen hyödyntäminen mallien rakentamisessa 12

2.3.2 Maalintekoyritysten mallintaminen 12

2.3.3 Syöttöjen ja tilankäytön mallintaminen 13

2.3.4 Yleisen suorituskyvyn datamallit 16

3 Alihinnoiteltujen pelaajien löytäminen 19

3.1 Pelaajamarkkinoiden toiminta 19

3.2 Pelaajien hinnoittelu 21

3.2.1 Pelaajan markkina-arvon indikaattorit 21

3.2.2 Joukkopohjainen pelaaja-arviointi 23

3.2.3 Datapohjaiset hinnoittelumallit 24

3.3 Pelaajien alihinnoittelun hyödyntäminen 26

4 Yhteenveto ja johtopäätökset 28

Lähteet 34

Kuvioluettelo

Kuva 1 Voronoin diagrammi 13

Kuva 2 Jalkapalloon sovellettu Voronoin diagrammi 14

Kuva 3 PlayeRank viitekehyksen toiminta. Mukaellen Pappalardo ym. (2019) 16

Kuva 4 Pelaajan markkina-arvon muodostuminen 21

Kuva 5 Pelaajan markkina-arvon indikaattorit 23

Taulukkoluettelo

Taulukko 1. Esimerkkejä pelaajien datamalleista ja niiden tarkoituksista 18

Taulukko 1 Viitekehys datavetoisen pelaajahankintaprosessin vaiheista esimerkeillä 31

Error! Bookmark not defined.

Taulukko 3 Kehittyneiden suorituskykymittareiden yhdistäminen pelaajahankintaprosessiin 33

1 JOHDANTO

Työväen harrastuksena alkaneesta jalkapallosta on muodostunut miljardibisnes, jossa joukkueet kilpailevat pokaalien ja saavutusten lisäksi miljoonien arvoisista tv-sopimuksista sekä geopoliittisista tavoitteista. Toisin kuin useimmissa muissa yrityksissä, taloudellinen menestys ei ole jalkapalloseuroille aina kaikista tärkeintä (Késenne, 2006). Seurat ovat valmiita maksamaan miljoonia pelaajista, jotka mahdollistavat urheilullisen menestyksen. Yleisin yhdistävä tekijä menestyneillä seuroilla on suuri budjetti, josta suurin osa kuluu pelaajien palkkoihin ja mahdollisiin siirtosummiin. Pienemmilläkin budjeteilla toimivilla seuroilla on kuitenkin mahdollisuus pärjätä toimimalla pelaajamarkkinoilla muita tehokkaammin. Haasteena ovat kuitenkin oikeanlaisten pelaajien tunnistaminen ja hankinta markkinoilla, joissa kilpailu on todella suurta. Monet seurat ovat ajautuneet taloudellisiin vaikeuksiin ja esimerkiksi historiallinen ranskalaisseura Bordeaux hakeutui hiljattain konkurssiin. Lukuisat konkurssit ja likvidaatiot kertovat seurojen urheilullisen menestyksen janoamisen lisäksi kehnosta taloudellisesta osaamisesta ja pelaajahankintastrategiasta.

Seurat ovat vuosikymmeniä koettaneet saada etua pelaajamarkkinoilla kykyjenetsintäosastoihin panostamalla. Pelaajia on aina arvioitu tilastojen perusteella. Teknologian kehityksen myötä pelaajista saatavilla olevien eri tilastojen ja datan määrä ja laatu on kasvanut räjähdysmäisesti 2020-luvulle tultaessa. Pelaajien tuottamasta datasta voidaan luoda malleja sekä pelaajia arvioida yksinkertaisten tilastojen kuten maalien sijaan kehittyneillä metriikoilla (Pappalardo ym. 2019). Pelaajien hankinta perustuu nykyään dataosastojen tuottamiin raportteihin. Tutkielmassa tarkastelen datalähtöisen pelaajahankintaprosessin tuomaa urheilullista lisäarvoa.

Jalkapallo on datan hyödyntämisen suhteen monia muita urheilulajeja jäljessä. Osittain tämä johtuu jalkapallon kaoottisuudesta verrattuna diskreetimpiin lajeihin, kuten baseballiin ja amerikkalaiseen jalkapalloon (Brooks ym. 2016). Amerikkalaisiin urheilulajeihin verrattuna jalkapallossa haasteena on myös eri pääsarjojen määrä ympäri maailmaa. Sarjojen laatuerot vaikeuttavat pelaajien arviointia datan avulla. Eri sarjoista on myös eri määrä dataa saatavilla. Data-analytiikasta on paljon tutkimusta, mutta tutkimus usein jää eri mallien rakentamisen tasolle, eikä niiden hyödyntämistä seurojen kannalta tarkastella perusteellisesti. Data-analytiikkaa koskeva tieteellinen tutkimus käsittelee malleja usein ainoastaan teoreettisella tasolla. On havaittavissa selkeä tutkimusaukko koskien datamallien hyödyntämistä käytännössä sekä itse pelaajahankintaprosessin parantamista. Tässä tutkielmassa tarkastelen datamallien ja mittareiden lisäksi hinnoittelumalleja, jotta koko pelaajahankintaprosessista voidaan muodostaa eheä ja käytännöllinen kokonaiskuva, jota voidaan hyödyntää myös urheilujohdon näkökulmasta.

Jalkapallon kautta voimme tarkastella data-analytiikan kehittymistä ja erilaisten kyvykkyyksien muodostumista teknologian kehityksen avulla. Tässä tutkielmassa tulen tarkastelemaan datan käyttöä pelaajien arvioinnissa erityisesti seurojen kannalta. Tutkielman tutkimusongelma on:

- Miten jalkapalloseurat voivat parantaa urheilullista menestystä löytämällä alihinnoiteltuja pelaajia datan avulla?

Vastaan tutkimusongelmaan seuraavien tutkimuskysymysten avulla:

- Minkälaista dataa on saatavilla pelaajista?
- Miten joukkueet voivat löytää alihinnoiteltuja pelaajia datan avulla?

Tutkielmani on tehty kirjallisuuskatsauksena ja siinä käsitellään tietojärjestelmätieteiden, tietojenkäsittelytieteiden, urheilutieteiden ja kauppatieteiden artikkeleita. Tutkielman luvussa 2 selvitän, mitä urheilullinen menestyminen tarkoittaa ja miten pelaajat vaikuttavat siihen. Tarkastelen historiaa data-analytiikan soveltamisesta urheiluun ja miten data-analytiikan kehitys mahdollistaa uudet kyvykkyydet seurojen pelaajahankintaan. Luvun 2 lopussa käsittelen, miten data-analytiikan kehitystä on hyödynnetty pelaajien arviointiin suunniteltujen datamallien rakentamisessa sekä tarkastelen muutamia koneoppimista hyödyntäviä malleja. Luvussa 3 tutkin pelaajamarkkinoiden toimintaa, historiaa ja teoriaa. Tarkastelen erilaisia hinnoittelumalleja ja mekanismeja, joiden avulla joukkueet voivat saada lisähyötyä löytämällä alihinnoiteltuja pelaajia. Luvussa 4 vastaan tutkimuskysymyksiin sisältölukujen 2 ja 3 perusteella. Yhteenvedossa käsittelen myös löydöksiä, lukujen nivoutumista yhteen tutkimusongelman kannalta ja mahdollisia tulevaisuuden tutkimuskohteita.

Tutkielmani ei siis koske muita datan käyttötarkoituksia, kuten vammojen estoa tai kunnonhallintaa tai pelitaktiikkaa ja otteluvalmisteluja koskevaa dataa. Pelaajien hankinnassa dataa voidaan hyödyntää myös taloudellisen hyödyn saavuttamiseksi, ja taloudellinen ja urheilullinen menestys ovatkin kulkeneet käsikädessä. Tämän tutkielman kontekstissa tarkastelen pelaajienhankintaprosessia urheilullisen lisähyödyn saavuttamiseksi. Erilaiset mittarit soveltuvat paremmin pelaajien suorituskyvyn mittaamiseen ja useiden seurojen tavoite onkin urheilullinen menestys taloudellisen voiton sijaan. Huippujalkapalloksi luokittelen isojen eurooppalaisten jalkapallomaiden pääsarjat. Datamalleissa ja hinnoittelumalleissa on käytetty dataa esimerkiksi Englannista, Saksasta, Italiasta ja Espanjasta. Tämän tutkielman kontekstissa, ja käytännön kannalta sarjat ovat vertailukelpoisia keskenään.

2 SEURAN URHEILULLISEEN MENESTYKSEEN VAIKUTTAVAT PELAAJAMITTARIT

Tämän tutkielman kontekstissa urheilullinen menestys tarkoittaa maan pääsarjassa menestymistä. Huippujalkapallossa menestymistä mitataan ottelusta saatavien pisteiden perusteella. Joukkueet saavat voitosta kolme pistettä ja tasapelistä yhden pisteen. Häviöstä ei saa pisteitä. Joukkueet sijoittuvat keräämiensä pisteiden perusteella sarjataulukossa. Eniten pisteitä kauden aikana kerännyt joukkue julistetaan liigan voittajaksi. Sarjataulukon viimeiset joukkueet putoavat alempaan sarjaan. Sarjataulukon kärkipään joukkueet pääsevät vuosittain kilpailemaan muita Euroopan huippujoukkueita vastaan UEFA:n järjestämissä turnauksissa (Oberstone, 2009). Myös näistä turnauksista saatavat sponsori- ja sopimustulot ovat merkittäviä, varsinkin pienempien joukkueiden kohdalla. (Mustafi ym. 2025)

Joukkueiden kesken jaettavat tv-sopimusrahat ovat merkittävä tulolähde, ja joillekin joukkueille jopa elinehto. Joukkueiden tulolähteiden pääpainon vaihtuminen palkintorahoista ja lipputuloista sopimusrahoihin tarkoittaa, että jokainen voitettu piste voi olla elintärkeä kaupallisesti, vaikka mestaruus olisikin saavuttamattomissa. (Hall ym. 2002)

Toisin kuin taloudellisia voittoja maksimoivissa pohjoisamerikkalaisissa urheiluseuroissa, eurooppalaisissa jalkapalloseuroissa tavoitteeksi on usein mielletty tai asetettu urheilullisen menestyksen saavuttaminen. Monet seurat tekevät tappiota vuodesta toiseen ja omavaraisuus on yleisesti heikolla tasolla. Taloudellinen tavoite on usein toiminnan mahdollistaminen ja kannattavuuden ylläpito eikä tulojen maksimointi. Pohjois-Amerikassa sarjat ovat suljettuja ja seurat toimivat yhteisesti sovitun palkkakaton alla. Eurooppalaisessa jalkapallossa joukkueet toimivat laajalti ilman ulkopuolista vaikutusta kattojärjestöiltä. (Sæbø & Hvattum, 2019)

2.1 Pelaajat menestyksen mahdollistajina

Késennen (2006) mukaan aiemmin on kuvattu yksinkertaistetusti, että joukkueet koettavat maksimoida omaa voittoprosenttiaan otteluissa peluuttamalla parhaita mahdollisia pelaajia, joita he ovat hankkineet tai kehittäneet. Voittomaksimointi (engl. Win maximization) on seurojen toimintamalli, jossa joukkueen voittoprosentin maksimointi on seuran tärkein tavoite. Késenne (2006) on kuitenkin kehittänyt seurojen tavoitteiden jaottelua. Pelkkä voittomaksimointi ei voi olla joukkueen operatiivinen tavoite, koska seurat eivät voi hallita voittoprosenttiaan. Taloudelliset voitot tai panostukset mahdollistavat joukkueen toiminnan. Voittomaksimointi ei tarkoita nykypäivän käsityksen mukaan taloudellisten voittojen huomiotta jättämistä, vaan että niitä ei maksimoida. Monet voittomaksimisoivat

seurat jahtaavat samoja pelaajia avoimilla markkinoilla, mikä johtaa sekä palkkojen, että siirtosummien nousuun. Monet seurat ovat ajautuneet taloudelliseen ahdinkoon maksettuaan pelaajista ylihintaa.

Jos seurojen tuottama päätuote mielletään suoritukseksi pelin aikana, voidaan resurssiperusteista näkökulmaa hyödyntää joukkueiden suorituskyvyn ymmärtämiseksi. Barneyn (1991) resurssiperusteisessa näkökulmassa (engl. Resource Based View) yritykset pyrkivät luomaan kestävää kilpailullista etua luomalla strategioita, jotka hyödyntävät mahdollisimman tehokkaasti yrityksen resursseja. Henkilöstöresurssit ovat osana lisäarvonluontia ja urheiluseuroissa nämä henkilöstöresurssit ovat pääosassa. McNamaran ym. (2013) mukaan monissa ympäristöissä, kuten elokuva-alalla, joukkueurheilussa, konsultoinnissa ja akatemiassa, henkilöstöresurssit ovat fyysisiä resursseja tärkeämpiä. Näissä ympäristöissä kilpailuetu saavutetaan yrityksissä, joissa on eniten henkilöstön kyvykkyyttä. Kilpailu näistä kyvykkyyksistä on näillä aloilla kiivasta. Seurojen pitää siis olla kilpailijoitaan tehokkaampia resurssien (pelaajien) valikoinnissa ja niiden kyvykkyyksien tunnistamisessa ja ennustamisessa äärimmäisen kilpailullisilla globaaleilla markkinoilla (Mustafi ym. 2025). Urheiluseurat pyrkivät luomaan kestävää kilpailuetua myös sisäisten resurssien kehittämisellä ja näiden molempien resurssityyppien yhteen sovittelulla ja hallinnoinnilla. (Lechner & Gudmundsson, 2012)

Seurojen menestys korreloi merkittävästi niiden budjettien kanssa. Rahakkaimmilla seuroilla on mahdollisuus kilpailla markkinoilla kalliimmista resursseista eli pelaajista. Pelaajat, joiden odotetaan tuovan merkittävää lisäarvoa joukkueellensa, ovat hinnoiteltu korkealle. Pienemmillä budjeteilla toimivien seurojen on siis urheilullisen menestyksen takaamiseksi löydettävä alihinnoiteltuja pelaajia pelaajamarkkinoilta (Lechner & Gudmundsson, 2012). Eurooppalaisessa jalkapallossa pelaajamarkkinat ovat avoimet ja pelaajista voidaan tehdä tarjouksia rahallisen korvauksen muodossa. Pelaajista kilpaillaan tarjotun palkan lisäksi tarjoamalla suurempia siirtosummia toisille joukkueille. Pelaajien vapaa liikkuvuus, palkkakaton puuttuminen ja lyhyet sopimukset tarkoittavat, että seurat voivat vapaasti hyödyntää taloudellista voimaansa ja ”ostaa menestyksensä”. Pelaajat esiintyvät viikoittain suurien yleisöiden edessä, joten markkinat toimivat melko tehokkaasti ja pelaajien ominaisuudet ovat laajalti yleisön tiedossa. (Hall ym. 2002)

2.2 Pelaajien hankkiminen dataa hyödyntäen

Seurojen menestykseen vaikuttaa kuitenkin lukuisat asiat, ja oikein koottu joukkue on enemmän kuin osiensa summa. Mestaruuksia tai sarjapisteitä ei jaeta paperilla vahvemmalle tai kalliimmalle joukkueelle (Lechner & Gudmundsson, 2012). Urheilullisen menestyksen saavuttaminen pienellä budjetilla on luonnollisesti ollut monien seurojen omistajien haave jo vuosikymmeniä. Dataa on hyödynnetty pelaajien arviointiin jo pitkään esimerkiksi baseballissa. Baseball soveltuu luonteensa vuoksi

erittäin hyvin data-analytiikan hyödyntämiseen. Data-analytiikan hyödyntämisen mahdollisuuden pelaajamarkkinoilla saavutettavaan lisähyötyyn toi yleiseen tietoisuuteen baseball-joukkueesta Oakland Athleticsista kertova kirja, Michael Lewiksen *Moneyball: The Art of Winning an Unfair Game*. Kirja kuvailee miten Oakland Athletics pystyi saavuttamaan huomattavaa menestystä maailman parhaassa baseball-sarjassa, MLB:ssä budjetilla, joka oli monta kertaa pienempi kuin useimpien kilpailijoiden. Oaklandin menestys 1980-luvulla perustui Bill Jamesin kehittämään Sabermetrics-malliin, joka pyrki löytämään mittareita pelaajista, mitkä linkittyivät voittamiseen. Oakland luotti omiin mittareihinsa pelaajien arvioinnissa tavanomaisten ja yleisten pelaajien arviointiin käytettyjen tilastojen sijaan, mikä paljasti markkinoilla olevan monia alihinnoiteltuja pelaajia. Data-analytiikan suosio kasvoi Oaklandin menestyksen jälkeen. Moneyball-ilmiöstä on muodostunut kaikenkattava fraasi urheilu- ja analytiikkapiireissä. Useimmiten sillä kuvataan tieteellistä ja analyyttistä tapaa koota kilpailullinen joukkue, mutta siihen kuuluu myös aspekti altavastaajan asemasta, ja sekä pelillisen, että taloudellisen hyödyn löytämisestä pienillä resursseilla. Sitä voidaan kuvailla organisaation filosofiana koskien urheiluanalytiikan alakategoriaa, joka liittyy joukkueiden ja pelaajien suorituskyvyn analyysiin sekä pelaajien rekrytointiin ja kykyjenetsimiseen. (Y. Li, 2021)

Baseball on ollut edelläkävijänä urheiluanalytiikassa, koska se on helposti diskretisoitava laji. Baseballottelu on helposti jaettavissa yksittäisiin ja standardoituihin tapahtumiin, kuten lyöntivuoroihin, joilla on selkeä lopputulema, kuten lyönnin osuma tai juoksu. Jalkapallo-ottelua ei voida jakaa selkeisiin tapahtumiin tai vuoroihin ja sen dynaaminen luonne on ollut esteenä selkeän ja hyödynnettävän datan keräämiseen. Vaikka jalkapallossa on hyödynnetty lähinnä karkeaa ja yksinkertaista dataa, kuten maaleja ja syöttöjä, on sillä kuitenkin historiansa urheiluanalytiikassa (Brooks ym. 2016). Reep ja Benjamin (1968) tutkivat jo vuosikymmeniä sitten syöttöketjujen pituuden merkitystä ottelumenestykseen pelkän käsin syötetyn datan perusteella.

2.3 Pelaajiin sovellettavat datamallit

2020-luvulle tultaessa teknologian kehitys auttaa vastaamaan data-analytiikan soveltamisen haasteisiin jalkapallossa. Sensoriteknologian kehitys, esineiden internet (engl. Internet of Things, IoT) ja erityisesti modernit seurantalaitteet ovat mahdollistaneet uudentyyppisen datan keräämisen. Dataa on enemmän, se on laadukkaampaa ja luotettavampaa. Pelaajista ja pelitapahtumista kerättävä data on kytketty aikaan, paikkaan ja pelaajaan. Täten se auttaa ymmärtämään ja kvantifioimaan monimutkaisia sekä dynaamisia pelitapahtumia. (Brooks ym. 2016; Decroos ym. 2019; Rein & Memmert, 2016)

Ottelutapahtumista kerätty data voidaan jakaa tapahtuma- ja sijaintidataan. Tapahtumadataa kerätään vieläkin usein manuaalisesti. Tapahtumadataa ovat yksinkertaiset tapahtumat ottelun aikana kuten esimerkiksi syötöt, laukaukset, rikkeet ja torjunnat. Sijaintidata kerätään puolestaan automaattisesti esimerkiksi erilaisilla optisilla seurantajärjestelmillä. (Anzer & Bauer, 2021; Herold ym. 2019)

2.3.1 Koneoppimisen hyödyntäminen mallien rakentamisessa

Koneoppimista voidaan hyödyntää monessa urheiluanalytiikan osa-alueessa massiivisen datamäärän käsittelyyn. Koneoppimismalleja voidaan luoda esimerkiksi ennustamaan pelien tuloksia huomioiden muuttujat kuten kotietu, aiemmat kohtaamiset, maajoukkuepelaajien lukumäärä tai maantieteellinen etäisyys joukkueiden välillä. Näitä ennustavia malleja hyödynnetään esimerkiksi vedonlyönnissä tai taktisissa päätöksissä (Tax & Yme Joustra, 2015). Koneoppimista voidaan hyödyntää myös pelaajien loukkaantumisriskin arviontiin (Rommers ym. 2024). Koneoppiminen tarjoaa keinon automatisoida pelitapahtumien tutkimisen huomioiden tapahtumien kontekstuaaliset muuttujat ja siten sen luomat luokittelumallit kykenevät tuottamaan monipuolista dataa (Herold ym. 2019).

Pelaajien kykyjen kvantifioiminen on keskeinen haaste pelaajahankintoja tehdessä. Suorituskyvyn mittarit ovat datan pohjalta luotuja mittareita, joiden tehtävänä on helpottaa laajojen ja tarkkojen pelaajaprofiileiden luomista. Jalkapallossa erilaisten suorituskyvyn avainmittareiden määrittäminen on ollut haastavaa johtuen lajin standardisoitujen pelitilanteiden vähyydestä ja pelipaikkojen moninaisuudesta (Schlenger ym. 2023). Perinteisen datan varaan rakennetut mittarit perustuvat yleiseen dataan, kuten maalimääriin tai keskimääräiseen syötön pituuteen. Nämä mittarit eivät kuitenkaan ota huomioon kontekstia, missä suoritus on tapahtunut (Davis ym. 2024). Koneoppimisen hyödyntäminen pelaajahankintojen kontekstissa perustuu erilaisten mallien rakentamiseen aiemmista otteluista kerätyn datan perusteella. Nämä mallit ovat oppineet historiallisesti datasta, miten tiettyjen toimintojen konteksti vaikuttaa niiden tehokkuuteen ja merkittävyyteen. Mallien avulla voidaan luoda uusia, yksityiskohtaisempia ja tarkempia mittareita pelaajista, joilla on selkeä yhteys suorituskykyyn ottelussa. (Gudmundsson & Horton, 2018)

2.3.2 Maalintekoyritysten mallintaminen

Ehkä tunnetuin koneoppimisella koulutettu malli on maaliodottama (engl. Expected Goals, xG), jota voidaan hyödyntää maalintekopeleissä kuten jääkiekossa tai jalkapallossa. xG on tilastollinen malli,

joka käyttää logistista regressiota kuvaamaan maalintekoyritysten laatua antamalla jokaiselle laukaukselle odotetun maaliarvon nollan ja yhden väliltä (Davis ym. 2024). Malli soveltuu jalkapalloon erinomaisesti, koska vähämaalisena lajina pelkästään tehtyjen maalien perusteella luotu malli olisi vääristynyt datan vähyyden vuoksi. Maalinteon tärkeyden vuoksi laukauksien määrä on usein yhdistettävissä voittamiseen, mutta pelkästään onnistuneiden laukausten arviointi olisi liian karkea abstraktio. xG-malli auttaa erottelemaan vaaralliset laukaukset käyttäen algoritmissään lukuisia eri tekijöitä, kuten laukauksen sijaintia, laukauksen suorittavaa kehonosaa ja puolustajien antamaa painetta laukojalle. (Anzer & Bauer, 2021)

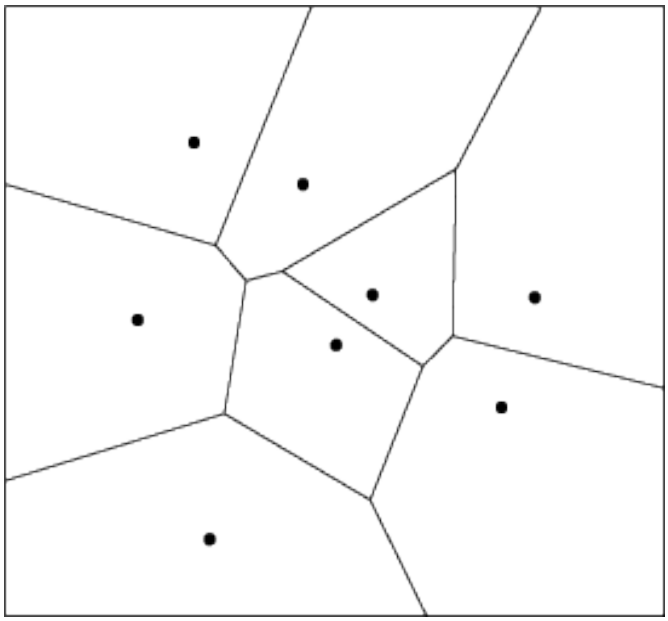
Yksi mallin vahvuuksista on sen ymmärrettävyys. Yksinkertaisuutensa ja selkeytensä vuoksi xG otetaan usein esille esimerkiksi TV-lähetyksissä, blogeissa ja sosiaalisen median päivityksissä. Vähämaalisen luonteensa vuoksi jalkapallo on usein lopputulokseltaan melko satunnainen ja siksi xG on usein tarkempi mittari joukkueen suorituksesta ottelun aikana kuin maalit (Umami ym. 2021). Joukkueet voivat hyödyntää mallia taktisia päätöksiä tehdessä vertailemalla eri joukkueiden laukaisumääriä ja sijainteja suhteessa maalinteon todennäköisyyteen (Anzer & Bauer, 2021). xG-mallilla voidaan kuvata ottelun maalintekotilanteiden määrä ja laatua sekä tapahtuneiden maalien ”onnekkuutta” vertailemalla kahden eri joukkueen odotettua maalikertymää tietyn ottelun aikana (Herold ym. 2019). Anzerin ja Bauerin mukaan (2021) pelaajien arvioinnissa ja vertailussa xG malli auttaa erottamaan hyökkääjien laukaustyylit ja -tavat. Esimerkiksi kahdella maailmanluokan laitahyökkääjällä Jadon Sancholla ja Serge Gnabrylla on täysin erilaiset laukauksien määrät ja laadut. Mallilla voidaan myös etsiä pelaajia, jotka ovat kliinisiä maalintekijöitä, he usein ylittävät maalimäärällään luomansa maaliodottaman.

2.3.3 Syöttöjen ja tilankäytön mallintaminen

Maalintekemiseen vaaditaan useimmiten laukauksen lisäksi syöttö, joka johtaa maaliin. Syöttöjä tapahtuu otteluissa useita satoja, joten niistä syntyy enemmän dataa kuin laukauksista tai maaleista. Syöttöjen merkitystä on mitattu pitkään yksinkertaisilla mittareilla. Syöttöjen määrä, syöttöjen onnistumisprosentti ja syötöstä seuraava laukaus ovat olleet perinteisinä mittareina pelaajien ja joukkueiden syöttöpelin arviointiin. Näiden mittareiden heikkous on sama kuin laukauksia tai maaleja koskevissa yksinkertaisissa mittareissa; ne eivät ota huomioon kontekstia tai sijaintitietoa (Goes ym. 2019). Datan keräämisen kehittyminen mahdollistaa monipuolisempien mallien rakentamisen myös syöttöjen tehokkuuden ja merkityksen määrittämiseen. Uudenlaisen, paikkaan ja aikaan linkittyvän, datan merkitys korostuu syöttöjen arvioinnissa. (Kang ym. 2006)

Brooksin ym. (2016) rakentama malli arvioi syöttöjen tehokkuutta vertailemalla kuinka usein tietyn joukkueen pallonhallintajakso päättyy laukaukseen. Jalkapallossa oivaltava ja peliä edistävä syöttö ei välttämättä johda välittömästi maaliin tai edes laukaukseen. Brooksin ym. (2016) mallissa hyödynnetään koneoppimista määrittämään kuinka paljon keskimäärin syöttö pallonhallintajaksossa johtaa laukauksen todennäköisyyden kasvamiseen. Mallin avulla jokaiselle syötölle voidaan antaa arvoksi sen suhteellinen tärkeys laukauksen luomiseen jossain vaiheessa pallonhallintajaksoa (Pass Shot Value, PSV). Pelaajille voidaan laskea keskiarvo heidän kaikista syötöistään (Average Pass Shot Value, APSV) ja sen perusteella tarkastella heidän pelinrakennustaitojaan. APSV suosii hyökkääjiä, mutta sen avulla voidaan löytää myös keskikenttäpelaajista hyökkäysorientoituneimmat. (Brooks ym. 2016)

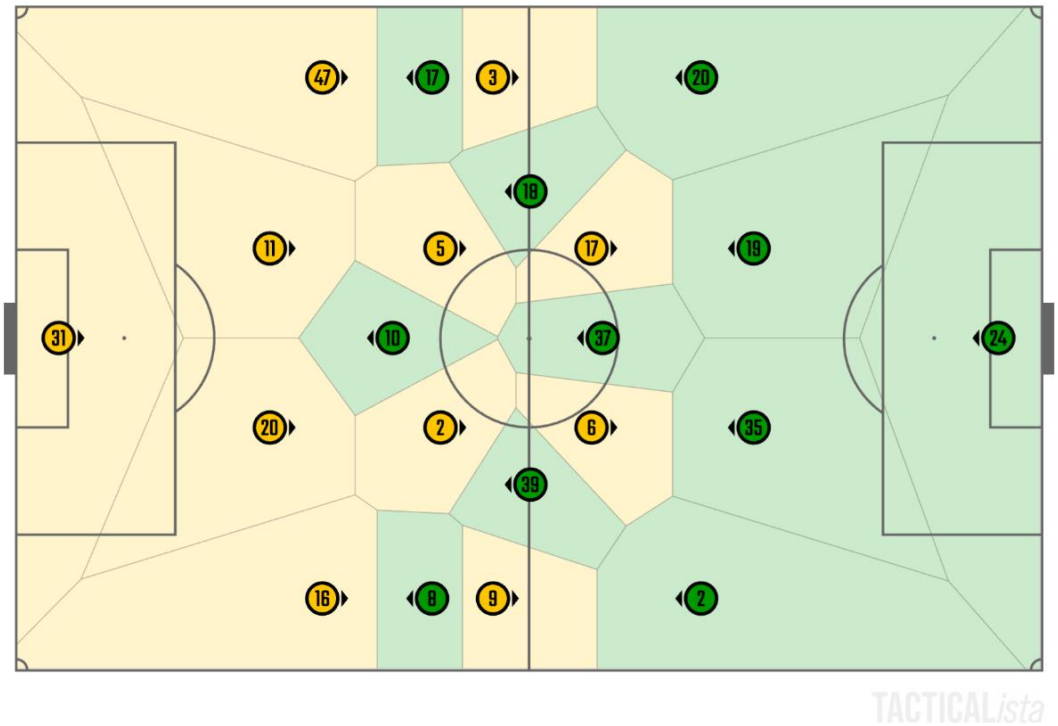
Yksi tapa ymmärtää jalkapalloa on pelitilojen kautta. Joukkueet pyrkivät luomaan ja hyödyntämään tilaa ja minimoimaan vastustajan tilan pelialueella (Herold ym 2019). Takin ja Hasegawan (2000) kehittämässä mallissa he määrittelevät tietyn pelaajan hallitsevan alueen (engl. dominant region), eli osan pelialueesta, jota tietty pelaaja on lähinnä tai minne hänen odotetaan ehtivän ennen muita pelaajia. Malli perustuu Voronoin diagrammiin, joka jakaa matemaattisen tason osiin siinä olevien pisteiden etäisyyksien perusteella. Jokainen piste tasoalueessa hallitsee tiettyä osa-aluetta. Osa-alue määräytyy siten, että jokainen osa osa-alueesta on lähimpänä sitä dominoivaa pistettä, kuin mitään muuta pistettä. (Aurenhammer, 1991)



Kuva 1 Voronoin diagrammi

Takin ja Hasegawan (2000) malli soveltaa Voronoin diagrammia, mutta tekee mallista jalkapalloon soveltuvamman ottamalla huomioon osa-alueiden jaossa niitä dominoivien pisteiden (tässä ta-

pauksessa pelaajien) nopeudet ja suunnat. Mallissa osa-alueet eivät määräydy siis puhtaasti etäisyyksien, vaan sen mukaan, millainen mahdollisuus pelaajalla on ehtiä tietylle alueelle ennen toista pelaajaa. Mallia voidaan hyödyntää pelaajien yhteistyön, syöttelyn ja paineen antamisen kvantifioimiseen. Syöttöjen merkitystä voidaan määrittää niiden luoman uuden tilan avulla. Pelaajien suorituksiin painoarvoa antaa tilan vähyys. (Taki & Hasegawa, 2000)



Kuva 2 Jalkapalloon sovellettu Voronoin diagrammi, Tacticalista.com generoima

Fernandez ym. (2018) laajensivat tilan käyttämisen analyysiä mallillaan. Malli luo Takin ja Hasegawan (2000) mallin kaltaiset alueet mahdollisista alueista, jossa pelaaja voi ottaa pallon haltuunsa. Alueet eivät kuitenkaan ole diskreettejä kuten Voronoin diagrammissa tai vastaavissa malleissa, vaan ne kuvastavat pelaajan vaikutusta tietyn tilan hallintaan. Pelaajat voivat luoda tilaa pelitovereilleen liikkeellään, vaikka heillä ei olisi palloa. Malli määrittelee kaksi eri mittaria. Pelaajan tilanhallinta (engl. Space Occupation Gain, SOG) ja pelaajan tilan luominen (engl. Space Generation Gain, SGG). Pelaajien liikkeellään ja tilanluomisella tehtyä lisäarvoa joukkueellensa painottavat eri pelikentän alueiden painoarvot.

Gudmundssonin (2012) kehittämä malli tarjoaa työkalut syöttöjen merkityksen ja pelaajan syöttelytaitojen määrittämiseen. Yksi mallin työkaluista soveltaa Takin ja Hasegawan mallin hallitsevaa aluetta määrittämällä syötön vastaanottajalle alueen missä hän voi vastaanottaa syötön (engl. passable region). Eri pelaajien syötettävät alueet kertovat, kuinka helppo heille on syöttää pallo. Näiden alueiden perusteella voidaan mallintaa syöttöjen vaikeuden lisäksi pelaajien päätöksentekoa eri syöttö-

vaihtoehtojen välillä, kykyä toteuttaa vaikeita syöttöjä ja pallottomien pelaajien kykyä tarjota mahdollisuus syötölle sekä vastaanottaa syöttö onnistuneesti. Toinen mallin työkalu määrittelee eri syöttelysekvenssien toistuvuuden, ja eri pelaajien muodostamat syöttökuviot. Tämän työkalun avulla voidaan mallintaa, miten joukkue rakentaa hyökkäyksiään ja mitkä pelaajat osallistuvat syöttösekvensseihin eniten ja millä pelialueilla.

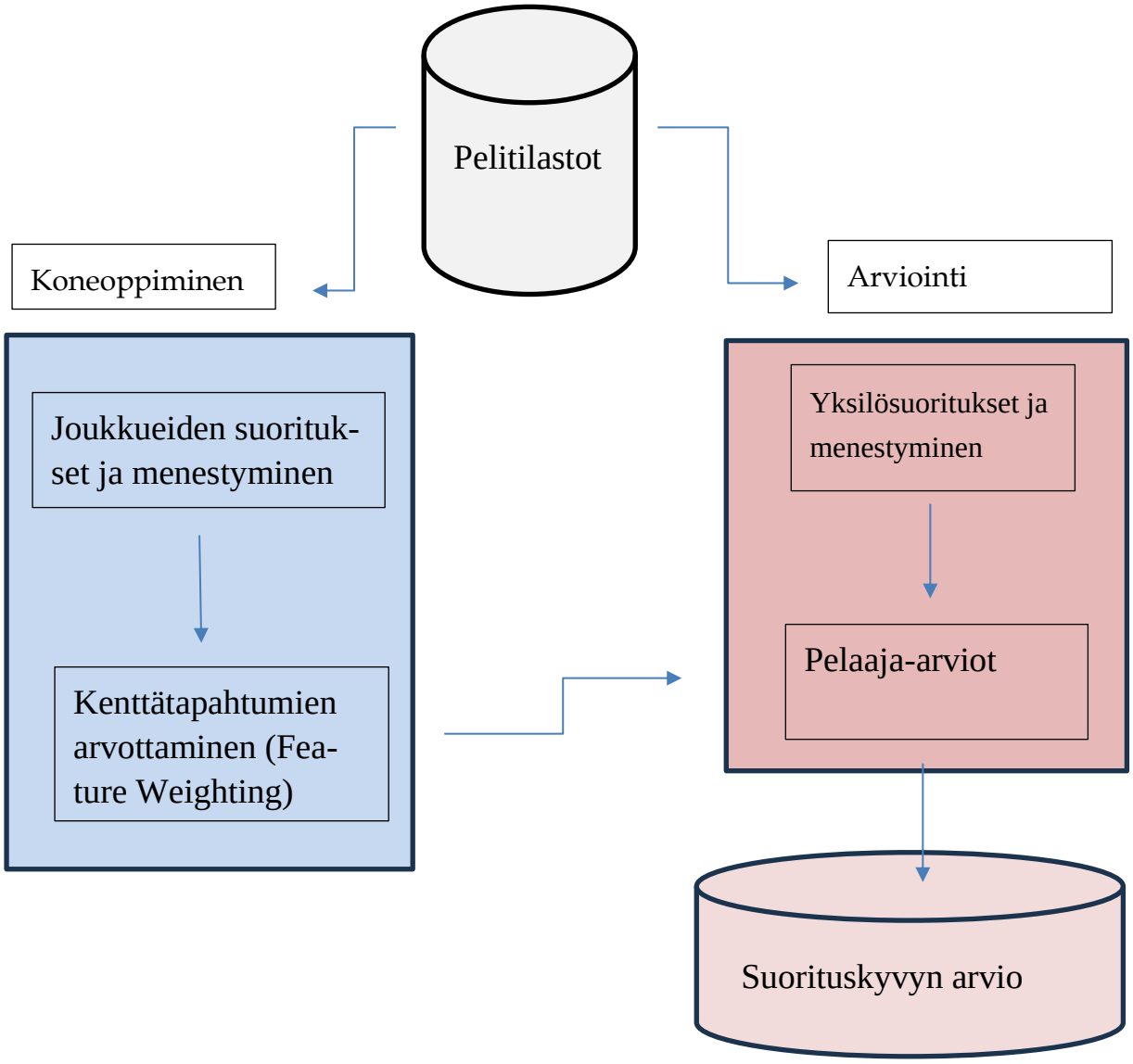
Jalkapallojoukkuetta ja sen pelaajia voidaan pitää verkostona (Duch ym. 2010). Wasserman ja Faust (1994) määrittelevät sosiaalisen verkostoanalyysin (engl. Social Network Analysis, SNA) menetelminä, joilla tutkitaan sosiaalisia verkostoja keskittyen yksilöiden välisiin suhteisiin heidän ominaisuuksiensa sijaan. Duch ym. (2010) kehittivät mallin tavoitteenaan selvittää kuinka paljon yksittäiset pelaajat voivat vaikuttaa monimutkaisen verkoston (jalkapallojoukkueen) toimintaan. Mallissa pallon liike tai ”virtaus” joukkueen pelaajien välillä on kuvattu verkostona. Malli määrittelee pelaajien vaikutuksen ottelun kulkuun laskemalla, kuinka usein pelaaja on osana virtauksia, jotka johtavat laukauksiin. Duch ym. (2010) mallintavat pelaajan virtauksen keskeisyyttä (engl. Flow Centrality, FC) mittarina pelaajan suorituksesta ottelun aikana. Mallia voidaan soveltaa myös joukkueen suoritukseen ja sitä käytettiin vuoden 2008 EM-kisojen joukkueiden ja pelaajien arviointiin. Mallilla esille nousivat voittajajoukkue Espanjan sekä keskikenttälegenda Xavin suoritukset. (Duch ym. 2010)

2.3.4 Yleisen suorituskyvyn datamallit

Decroosin ym. (2019) kehittämä viitekehys pyrkii objektiiviseen pelaajan suorituksen kvantifioimiseen huomioimalla kaikki tapahtumat (syötöt, kuljetukset, keskitykset, ja laukaukset) ja tilannekontekstin, jonka alaisina toiminnot suoritettiin. Valuing Actions by Estimating Probabilities (VAEP) on viitekehys, joka antaa kaikille pelaajan suorittamille tapahtumille arvon perustuen maalintekemisen todennäköisyyden muutokseen (esimerkiksi. +0.05 tai -0.12). Viitekehys muuttaa pelin aikana tehtyjen toimintojen arvot pelaajakohtaiseksi arvioinniksi. Arviointeja vertaamalla perinteisiin tilastoihin ja markkina-arvoihin perustuviin pelaaja arviointeihin voidaan markkinoilta löytää alihinnoiteltuja pelaajia. VAEP-viitekehyksellä voidaan myös jaotella pelaajien pelityylejä painottamalla tiettyjä tapahtumia, kuten kuljettamista. Mittarit, joiden avulla voidaan erotella pelityylejä ovat arvokkaita seuroille, koska pelityylien sovittelu pelitaktiikkaan on oleellinen osa urheilullisen menestyksen rakentamista. Silmämääräinen pelityylien erottelu pelaajakohtaisesti on ollut työlästä kykyjenetsijöille. (Decroos ym. 2019)

Harvat mallit ovat kehitelty yleisen suorituskyvyn mittaamiseen. Monet mittarit arvioivat tiettyjä aspekteja suorituksesta ottelun aikana, kuten maaliodottamaa tai syöttötarkkuutta (Pappalardo ym.

2019). Duchin ym. Flow Centrality mallia voidaan pitää melko yleisenä mittarina, mutta koska se perustuu syöttöjen keskeisyyteen, se soveltuu erityisesti keskikenttäpelaajien ja hyökkääjien arviointiin. Brooks ym. (2016) huomioivat artikkelissaan, että heidän Pass Shot Value -mittari suosii hyökkääjiä. Näiden ongelmakohtien varjossa Pappalardo ym. (2019) kehittivät yleistä suoritusta arvioivan viitekehyksen PlayeRank. Viitekehys koulutetaan arvostamaan tiettyjä kenttätapahtumia suhteessa ottelun lopputulokseen. Tietyt kenttätapahtumat saavat painotetut arvot riippuen myös pelaajan pelipaikasta, jonka PlayeRank määrittelee automaattisesti sijaintitiedon perusteella. PlayeRankin tavoitteena on mallintaa jokaisen pelipaikan pelaajan merkityksen voiton kannalta antamalla pelaajille arvosanan jokaisesta ottelusta perustuen arvotettuihin tapahtumiin. Viitekehyksen avulla voidaan myös tarkastella myös pelaajien monipuolisuutta pelipaikallisesti ja heidän tasaisuuttaan suorituksissa kauden aikana. (Pappalardo ym. 2019)



Kuva 3 PlayeRank viitekehyksen toiminta. Mukaellen Pappalardo ym. (2019)

Taulukko 2 Esimerkkejä pelaajien datamalleista ja niiden tarkoituksista

Tarkoitus	Mallit	Lähteet
Maalintekoyritysten mallintaminen	Expected Goals	(Anzer & Bauer, 2021)
Syöttöjen ja tilankäytön mallintaminen	Pass Shot Value, Dominant Region, Flow Centrality	(Brooks ym. 2016; Duch ym. 2010; Fernandez ym. 2018; Taki & Hasegawa, 2000)
Yleisen suorituskyvyn mallit	PlayeRank, Valuing Actions by Estimating Probabilities	(Decroos ym. 2019; Pappalardo ym. 2019)

3 ALIHINNOITELTUJEN PELAAJIEN LÖYTÄMINEN

Carmichaelin ja Thomasin (1993) mukaan pelaajamarkkinoilla on kaksi päätehtävää: (1) Mahdollistaa joukkueiden pelisuorituksien parantuminen uusien pelaajien avulla ja (2) mahdollistaa pelaajien liikkuvuus joukkueista toisiin parempien mahdollisuuksien (peliajan ja roolin), paremman palkan tai yleisen työtyytyväisyyden perässä. Pantusonin ja Hvattumin (2020) mukaan joukkueen näkökulmasta onnistunut toiminta pelaajamarkkinoilla mahdollistaa joukkueen pitkäaikaisten taloudellisten- sekä lyhytaikaisten urheilullisten tavoitteiden saavuttamisen. Onnistuneeseen toimintaan sisältyy pelaajien ostaminen ja myyminen. Koska jalkapallojoukkueet pohjaavat enemmän henkilöstöresursseihin kuin pääomaan, joukkueiden markkina-arvo riippuu pitkälti sen pelaajien markkina-arvoista (Franceschi ym. 2024). Joukkueen markkina-arvon kasvuun on yhdistetty maltillinen aktiivisuus markkinoilla, oikea-aikainen vanhenevien pelaajien myyminen ja nuorempiin lupaaviin pelaajiin sijoittaminen. Joukkueiden strateginen ongelma on mahdollisimman arvokkaiden pelaajien hankkiminen sille allokoidulla pelaajabudjetilla. Kuten useimmat avoimet markkinat myös jalkapallon pelaajamarkkinat ovat vaikeasti ennustettavissa, ja pelaajien arvoon ja sen muutokseen vaikuttavat monet seikat. (Pantuso, 2017)

3.1 Pelaajamarkkinoiden toiminta

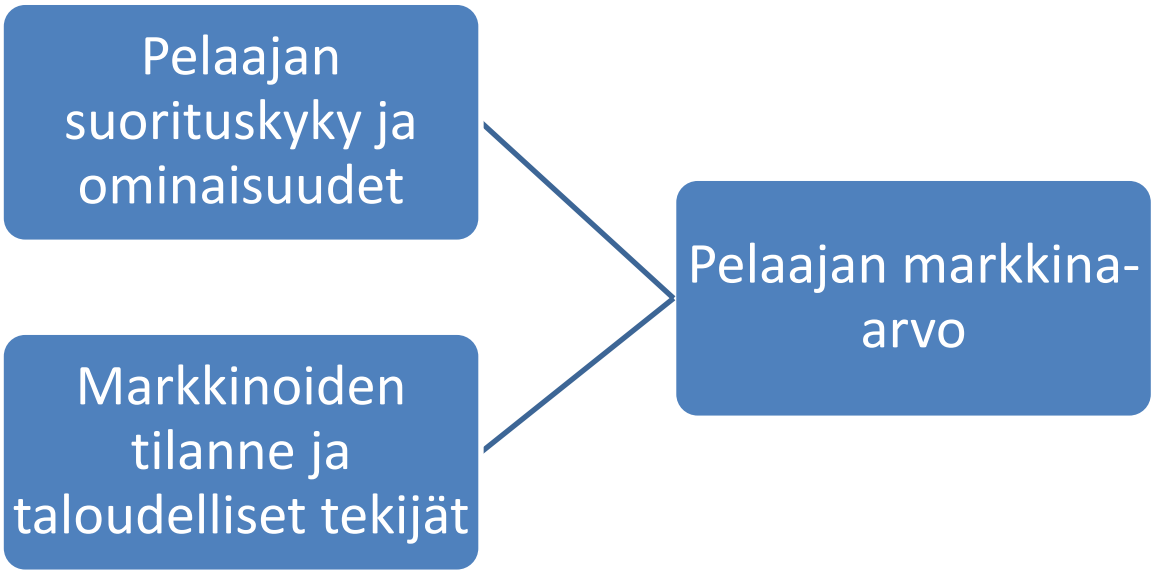
Pelaajat ovat tietyllä tavalla seurojen rahallista omaisuutta. Heihin on kiinnitetty suuri arvo, joka näkyy seurojen taloudellisissa katsauksissa. He ovat myös iso osa seuran arvonluontia, tuotannontekijöinä osana prosessia, joka koittaa maksimoida seuran yleistä hyötyä (engl. utility maximization) tai urheilullista hyötyä (engl. win maximization) (Késenne, 2006). Barneyn (1991) resurssiperusteisesta näkökulmasta pelaajat ovat henkilöstöresurssi, joiden tuottamaa lisähyötyä koitetaan maksimoida yksilöllisillä strategioilla.

Vuoden 1990 Euroopan yhteisöjen tuomioistuimen päätöksen mukaan pelaajat ovat kuitenkin enemminkin työntekijöitä kuin seurojen omaisuutta. Tämä päätös tunnetaan sen alullepanijan, pelaajan Jean Marc Bosmanin nimellä. Tapauksessa pelaaja Jean Marc Bosman halusi vaihtaa seuraa sopimuksensa päätyttyä. Halukas uusi joukkue ja vanha joukkue eivät kuitenkaan päässeet yhteisymmärrykseen siirtosummasta. Bosman voitti tapauksen oikeudessa ja täten sai vaihtaa seuraa ilman siirtosummaa. Bosman-tapaus vaikutti siihen, miten teoreettisesti pelaajia tarkastellaan osana seuroja. Seurojen näkökulmasta Bosman-tapaus tarkoitti sitä, että pelaajat saavat siirtyä heidän sopimuksensa päätyttyä toiseen seuraan ilman siirtokorvausta. Tämä luo seuroille painetta myydä pelaajat, jotka eivät ole kiinnostuneita sopimuksen uusimisesta (Antonioni, 2000; Frick, 2007). Franceschin ym.

(2024) mukaan Bosman-tapaus oli kulmakivi pelaajamarkkinoiden deregulaatiossa, mikä johti markkinoiden aktiivisuuden nousuun. Esimerkiksi EU:n sisäisillä seuroilla ei enää ole kiintiöitä koskien pelaajien kansallisuutta. Tapauksen myötä pelaajien voidaan katsoa olevan poistettavia aineettomia hyödykkeitä seurojen kirjanpidon kannalta.

Bosman-tapauksen myötä pelaajat, jotka ovat ilman sopimusta ovat vapaita agentteja (engl. free agents). Vapaat agentit voivat siirtyä seuraan, joka solmii uuden sopimuksen heidän kanssaan. Jos pelaajalla on sopimus toiseen seuraan, on ostavan ja myyvän seuran ensin sovittava siirtokorvauksesta (engl. transfer fee), minkä jälkeen seura saa aloittaa sopimusneuvottelut pelaajan kanssa. Pelaajia voidaan myös lainata toiselle joukkueelle, tällöin lainaava seura maksaa pelaajan palkan tai osan siitä, mahdollisen lainakorvauksen lisäksi (Pantuso, 2017). Tämän tutkielman kontekstissa tutkin erityisesti pelaajien ostamista, eli seuran vaihtamista siirtokorvausta vastaan.

Eurooppalaisessa jalkapallossa pelaajat vaihtavat seuraa useimmiten rahallista korvausta vastaan. Pelaajien markkina-arvo (engl. market value) kuvastaa arviota siirtokorvauksen suuruudesta, jonka ostava seura odotetaan maksavan tietystä pelaajasta. Ellei seura ole määrittänyt selkeää markkina-arvoa pelaajalle (hintapyyntö) tai pelaaja ole juuri siirtynyt tiettyyn hintaan markkinoilla, markkina-arvot ovat perustuneet toimittajien tai joukkueiden omiin arvioihin. Nykypäivänä pelaajien markkina-arvot perustuvat yksittäisten arvioiden hintojen lisäksi yleisöpohjaisiin (engl. crowd based estimations) tai datalähtöisiin (engl. data-driven estimations) arvioihin (Franceschi ym. 2024; Müller ym. 2017). Itse toteutuneisiin siirtokorvauksiin vaikuttaa pelaajien markkina-arvon lisäksi myös joukkueiden neuvotteluasemat. Pelaajasiirtoneuvottelua voidaan Carmichaelin ja Thomasin (1993) mukaan pitää strategisena neuvotteluna, jossa molemmat osapuolet pyrkivät Nashin neuvotteluperiaatteen mukaan maksimoimaan omaa hyötyään. Ostaja ja myyjä ovat usein eriarvoisessa asemassa. Myyjä voi olla tilanteessa, jossa sen kannattaa luopua pelaajasta hänen ikänsä, kunnon (tai terveyden), suorituskyvyn puutteen tai taktisen sopimattomuuden vuoksi. Myös pelaaja itse voi vaatia siirtoa. Pelaajan myyntiin voi johtaa myös taloudelliset paineet. Ostajan neuvotteluvoimaan voi vaikuttaa akuutit ja yllättävät pelaajapuutteet (omien pelaajien myynnit ja loukkaantumiset) tai urheilullisista tavoitteista lipsuminen. Myös siirtojen ajankohdalla on väliä. Eurooppalaisessa jalkapallossa siirrot ja pelaajien rekisteröinnit tapahtuvat sovittuna aikoina eli siirtoikkunan ollessa auki.



Kuva 4 Pelaajan markkina-arvon muodostuminen

3.2 Pelaajien hinnoittelu

Pelaajien hinnoittelulla, eli arvon määrittämisellä, on monia tarkoituksia. Seuran pelaajat ovat arvotettavissa vuosittaisiin tilinpäätöksiin tai muihin taloudellisiin lausuntoihin. Pelaajien arvot ovat otettava huomioon kaikessa toiminnassa pelaajamarkkinoilla. Pelaajien ostaminen on aina sijoitus, jolla on riskinsä. Pelaajien arviointi antaa myös signaaleja muille seuroille ja itse pelaajan edustajille siitä, kuinka arvokkaana häntä pidetään, tämä toimii perustana siirtokorvauksesta neuvottelun lisäksi pelaajan ja seuran välisiin sopimusneuvotteluihin. (Franceschi ym. 2024)

3.2.1 Pelaajan markkina-arvon indikaattorit

Pelaajien markkina-arvon indikaattorit ovat tekijöitä, joiden perusteella voidaan arvioida tai mallintaa pelaajien mahdollisia markkina-arvoja. Pelaajien markkina-arvon indikaattorit voidaan jakaa kolmeen kategoriaan.

- Pelaajien ominaisuudet
- Pelaajien suorituskyky
- Pelaajien kuuluisuus

Pelaajan ominaisuudet ovat fyysisiä ja demografisia faktoja pelaajasta, kuten ikä, pituus, pelipaikka, jalkaisuus tai kansallisuus. Pelaajan ikä kuvastaa hänen kokemustaan ja potentiaalia. Markkina-arvo yleensä nousee noin 25-vuotiaaksi asti, jolloin se lähtee laskuun. Pelaajan pituus korreloi hyvän pallon puskemistaidon kanssa, mikä on tärkeä taito puolustajille sekä hyökkääjille (Müller ym. 2017). Herm ym. (2014) havaitsivat, että molempijalkaisuudella (engl. ambipedal, eli henkilö kykenee käyttämään kumpaakin jalkaa yhtä hyvin) on positiivinen vaikutus markkina-arvoihin. Vasenjalkaiset ja molempijalkaiset ovat oikeajalkaisia harvinaisempia, ja heillä on tietyillä pelipaikoilla helpompi pelata syöttö- ja kuljetuskulmien avautumisen vuoksi. Pelaajien pelipaikka vaikuttaa pelaajien monipuolisuuteen ja näkyvyyteen yleisön kannalta. Keskikenttäpelaajat voivat usein toimia eri pelipaikoilla, joten he ovat joustavampia ja täten keskimäärin kalliimpia kuin esimerkiksi maalivahdit. Hyökkääjät puolestaan koetaan jännitystä lisääviksi pelaajiksi ja he ovat usein otsikoissa muiden pelipaikkojen edustajia enemmän. (He ym. 2015.; Müller ym. 2017)

Vaikka pelaajien suorituskyky on vaikea määritellä yksiselitteisesti, on olemassa suorituskykyyn liittyviä indikaattoreita, jotka selkeästi linkittyvät pelaajien markkina-arvoon. Euroopan välisillä kentillä esiintymisellä ja maajoukkueessa pelaamisella on positiivinen vaikutus pelaajien markkina-arvoihin (Carmichael & Thomas, 1993; Garcia-del-Barrio & Pujol, 2007). Peliajan merkityksestä markkina-arvoon on tehty eriäviä johtopäätöksiä riippuen datamallista, mitä on käytetty. C. Li ym. (2022) eivät löytäneet yhteyttä peliajasta (mitattuna minuutteina kentällä) ja markkina-arvosta, Müller ym. (2017) puolestaan muodostavat korrelaation. Useilla yksinkertaisilla tilastoilla on positiivinen yhteys markkina-arvoihin. Yleisiä yksinkertaisia tilastoja, joita on käytetty pelaajien arvon arvioinnissa, ovat pelaajan tekemät maalit, syötöt ja voitetut pääpallot. (Carmichael & Thomas, 1993; Frick, 2007)

Pelaajien suorituskyvyn lisäksi heidän kuuluisuutensa ja status ajavat markkina-arvoa ylöspäin. Seurat saattavat maksaa yli muuten odotetun markkina-arvon, koska kuuluisat pelaajat voivat statuksellaan lisätä lipunmyyntiä ja fanipaitojen menekkiä (Behravan & Razavi, 2021). David Beckhamin siirrossa Real Madridiin siirtoneuvotteluissa käytiin keskustelua pelaajan imagon tuomista lisäansioista (Tunaru ym. 2005). Internetin ja sosiaalisen median yleistyminen on antanut uusia tapoja mitata kuuluisutta. Lehmannin ja Schulzen (2008) mukaan pelaajan mainintojen määrä nettilehdessä on yhteydessä pelaajan arvoon. Erilaiset hakukonetulosten ja klikkausten määrä ovat nykyään helposti mitattavissa, ja niiden sekä pelaajan markkina-arvon välillä on selkeä yhteys. (Garcia-del-Barrio & Pujol, 2007; Herm ym. 2014)

Nämä eri indikaattorit käyvät ilmi kaikissa pelaajan markkina-arvon arvioinneissa, mutta niitä voidaan erityisesti hyödyntää markkina-arvoa datan perusteella arvioivissa malleissa.

Kuva 5 Pelaajan markkina-arvon indikaattorit

3.2.2 Joukkopohjainen pelaaja-arviointi

Joukkopohjainen (engl. crowd-driven) pelaaja-arviointi luottaa ”yleisön viisauteen”. Transfermarkt on yksi internetin käytetyimpiä sivustoja pelaajamarkkinoiden seuraamiseen. Sivusto tarjoaa yleisiä tilastoja pelaajista sekä uutisia ja huhuja pelaajasiirroista. Sivuston uniikki ominaisuus on sen tapa arvioida pelaajien markkina-arvoa. Sivusto luottaa joukkopohjaiseen (engl. crowd-based) pelaaja-arviointiin antamalla jokaiselle rekisteröityneelle käyttäjälle mahdollisuuden arvioida pelaajan markkina-arvo. Sivuston periaatteena on, että joukon muodostama keskiarvo on lähellä pelaajan ”todellista” arvoa. Sivuston antamat markkina-arvot ovat historiallisesti olleet hyvin lähellä toteutuneita siirtosummia sekä niitä käytetään usein mediassa ja tutkimuksissa. (Franck & Nüesch, 2012; He ym. 2015; Müller ym. 2017)

Joukkoistamisella (engl. crowdsourcing) on yleisesti ollut ongelmana ihmisten puolueellisuus, tarkoitukselliset manipulaatiot sekä tiedon ja kokemuksen puute (Müller ym. 2017). Hermin ym. (2014) mukaan Transfermarkt on kuitenkin huomioinut nämä yleiset epäkohdat. Sivuston markkina-arviot eivät ole täysin demokraattisia, vaan arviot ovat painotettu hierarkkisesti. Tietyt ”tuomarit” painottavat arviot ja tekevät lopulliset päätökset markkina-arvoista. Joukkopohjaisella pelaaja-arvioinnilla on kuitenkin heikkoutensa. Käyttäjien ja tuomarien puolueellisuus tai tiedon puute saattaa vääristää markkina-arvoja. Markkina-arvot eivät myöskään ole välttämättä ajankohtaisia. Markkina-arvot päi-

vitetään vuoden tai puolenvuoden välein, mikä tarkoittaa, että yksittäiset ottelut tai kuukauden tapahtumat eivät välity markkina-arvoihin. Joukkopohjaiset arvioinnit ovat tarkempia tunnetuimpien pelaajien kohdalla. Tunnettujen seurojen pelaajista on monia arvioita, mistä muodostaa lopullinen markkina-arvo. Tuntemattomissa joukkueissa tai sarjoissa pelaavista pelaajista ei välttämättä ole tarpeeksi arvioita, jotta todenmukainen keskiarvo voitaisiin muodostaa. Sivuston markkina-arvot ovat myös julkisia, mikä tarkoittaa, että yksittäinen seura ei saa markkina-arvioista etulyöntiasemaa siirtoneuvotteluissa.

3.2.3 Datapohjaiset hinnoittelumallit

Datapohjaiset (engl. data-driven) hinnoittelumallit välttävät monet joukkopohjaisten arvioiden heikkoudet. Datapohjaiset hinnoittelut käyttävät tilastollisia malleja, joilla on selkeät indikaattorit. Indikaattoreilla on empiirisesti johdetut painoarvot ja ne ovat yhdenmukaisia kaikkien pelaajien välillä. Pelaajien markkina-arvot voivat päivittyä jokaisen ottelun jälkeen ja se antaa tarkkoja hinnoitteluarvioita tuntemattomammistakin pelaajista, kunhan heistä on dataa saatavilla. Pelaajan määritelty hinta ei ole julkista tietoa, joten seurat voivat hyödyntää arvioita lisähyödyn saamiseen pelaajamarkkinoilla. Lisähyöty voi olla alihinnoitellun pelaajan ostaminen tai ylihinnoitellun pelaajan myyminen. (Müller ym. 2017)

Pelaajien markkina-arvon arviointi on ollut kiinnostuksen kohteena jo pitkään. 2000-luvun alussa Tunaru ym. (2005) rakensivat ensimmäisen viitekehyksen pelaajien markkina-arvon mallintamiseen. Vaikka malli tunnistaa erilaiset suorituskykyyn ja ominaisuuksiin perustuvat indikaattorit, ei se pohjaa suoranaisesti indikaattoreihin vaan Opta Indexiin, yksityisenä toimivan yrityksen pelaajien suorituksia arvioivaan järjestelmään. Opta Index ylläpitää laajaa tietokantaa pelaajien suorituksista ja käyttää monista kenttätapahtumista muodostuvaa dataa pelaajien suorituksien arviointiin. Tunaru ym. (2005) käyttävät indeksiä kvalitatiiviseen pelaajan kyvykkyyden arvioimiseen. Malli käyttää indeksiä ja optioiden hinnoitteluun käytettyjä matemaattisia malleja pelaajan arvon määrittämiseen. Mallin erikoisuutena se ottaa huomioon eri seurojen taloudellisen ja urheilullisen kyvykkyyden hinnoittelussa. Malli antaa eri markkina-arvoja samalle pelaajalle ostavasta ja tämänhetkisestä seurasta riippuen.

Cohen ja Risk (2024) rakensivat oman mallinsa pelaajan markkina-arvon arvioimiseen, tavoitteenaan parantaa Tunarun ym. (2005) mallia. Malli on kaksiosainen ja se yhdistelee taloudellista matematiikkaa sekä verkostoteoriaa (engl. Network theory). Hinnoittelumalli tarkastelee pelaajien suorituksia Opta Indexin sijasta Duchin ym. (2010) tässä tutkielmassa jo aiemmin käsiteltyyn Flow Centrality malliin perustuvalla mallilla. Uusi malli hyödyntää verkostoteoriaa tarkastellakseen pelaajien

suorituksia osana syöttelyketjuja ja ”virtauksia”, mutta lisää siihen syöttelyketjun päättävät toiminnot, kuten pallonmenetykset ja laukaukset. Malli antaa pelaajille arvot, jotka kuvaavat pelaajan tarjoamaa osaa joukkueen kokonaisesta suorituksesta sekä heidän suoritusten vaihtelevuutta otteluiden välillä. Malli määrittää pelaajien taloudellisen arvon käyttämällä Tunarun ym. (2005) tapaan rahoitusinstrumenttien arvon määrittämiseen käytettyjä matemaattisia malleja. Se kuitenkin parantaa aiempaa mallia yhdistämällä siihen aiempaa taloudellista teoriaa määritelläkseen pelaajien monivuotisten sopimusten arvoja. Malli nostaa esille, miten pelaajille maksettu palkka vastaa heidän suorituksiaan. Mallista käy ilmi esimerkiksi Englannin Valioliigan tähtipelaajan Mohamed Salahin alipalkkaus hänen suoritustasoonsa nähden vuosina 2018–2021. Salah saikin palkankorotuksen 2022 kaudesta eteenpäin. Malli on muissakin tapauksissa tunnistanut alipalkattuja pelaajia, jotka ovat myöhemmin saaneet palkankorotuksia. Mallia voi käyttää urheilujohdon näkökulmasta lähinnä seuran pelaajien jatkosopimusten suunnittelemiseen ja mahdolliseen pelaajista luopumiseen. Cohen ja Risk (2024) tunnistavatkin, että tulevaisuudessa voisi tutkia vastaavan mallin toimimista vapaiden agenttien tai siirtokohteiden etsimiseen.

Yalçınkayan ja Işıkin (2024) malli pyrkii auttamaan seuroja pelaajien arvioinnissa, jotta ne voisivat toimia pelaajamarkkinoilla mahdollisimman tehokkaasti. Malli pohjaa pelaajien suorituskykyyn liittyviin indikaattoreihin. Mallissa koulutettiin lukuisia eri koneoppimismetodeja ja niiden tuottamia arvioita pelaajien arvoista vertailtiin Transfermarkt hintoihin. Monella eri koneoppimismetodilla päästiin hyvin lähelle Transfermarktin antamia hinta-arvioita. Neuroverkostoon perustuva malli arvioi pelaajan arvon 97 % tarkkuudella. Monella yksinkertaisemmalla mallilla (kuten logistisella regressiolla) oli hankaluuksia arvioida pelaajien hintoja, koska eri indikaattoreilla voi olla monimutkaiset keskinäiset suhteet. Malli tuotti myös korrelaatiokertoimet eri suorituskykyindikaattoreille markkina-arvon suhteen. Eniten positiivisesti markkina-arvoon vaikutti tehdyt maalit, laukaukset per ottelu, kuljetukset ja läpisyötöt (syötöt, jotka pelataan hyökkääjälle puolustuslinjan taakse). Negatiivisesti markkina-arvoon vaikutti odotetun iän lisäksi puolustukselliset toiminnot, kuten syötönkatkot, voitetut pääpallot ja purut (pallo pelataan pois mahdollisimman kauas oman maalin alueelta). Tyypillisesti hyökkääjille ominaiset toiminnot korreloivat siis voimakkaasti markkina-arvon kanssa, kun taas puolustajille tyypilliset toiminnot negatiivisesti. Tämä on linjassa Hen ym. (2015) sekä Müllerin ym. (2017) tutkimusten kanssa, mitkä määrittelivät hyökkääjien arvojen olevan korkeammalla kuin muiden pelipaikkojen pelaajien. Toimintojen korrelaation eroihin vaikuttaa kuitenkin myös eri liigojen väliset dynamiikat ja pelaajien eri roolit. Yalçınkaya ja Işık (2024) tunnistavatkin mahdolliseksi tulevaisuuden tutkimuksen aiheeksi eri sarjojen ja pelipaikkojen suhteen normalisoidun pelaaja-arvioinnin. Eri pelipaikoille tehdyt mallit olisivat tehokkaampia eri pelaajien vertailuihin, mutta malli ei olisi yhtä yleistettävä. Yalçınkayan ja Işıkin (2024) mukaan malli sellaisenaan kuitenkin voisi auttaa seuroja tekemään tehokkaampia datavetoisia päätöksiä markkinoilla, joilla ne välttäisivät pelaajista ylimaksamisen.

C. Lin ym. (2022) koneoppimista hyödyntämässä hinnoittelumallissa käytetään suorituskykyyn liittyviä indikaattoreita selvittämään pelaajien arvoa. Malli on logiikaltaan samanlainen kuin Yalçınkayan ja Işıkin (2024) malli, mutta se käyttää pelaajille maksettua palkkaa vertauksena pelaajan arvosta. Malli hyödyntää myös pelaajien ominaisuuksiin liittyviä indikaattoreita, kuten pelaajan ikää ja aiempia saavutuksia uran aikana. Mallin avulla voidaan tarkastella mitkä pelaajat ovat yli- tai alipalkattuja. Mallin ennustamien palkkojen perusteella tärkeimmät suorituskyvyn indikaattorit ovat maalinteon ja iän lisäksi kuljetusten määrä ja laukaukset maalia kohti. Mallin heikkoutena on pelaajan suosion ja kuuluisuuden huomiotta jättäminen.

Müller ym. (2017) kehittivät oman mallinsa tutkittuaan joukkopohjaisten markkina-arvojen heikkouksia. He kokosivat pelaaja-arvoihin vaikuttavat indikaattorit perustuen aiempaan tutkimustietoon. Malli hyödyntää kaikkia kolmentyyppisiä indikaattoreita. Mallin erikoisuus on sen muodostamien markkina-arvojen perustuminen aiemman kauden pelaajien Transfermarkt-arvoihin, johon malli ennustaa muutokset kauden tapahtumiin perustuen. Mallin muodostamia markkina-arvoja ei kuitenkaan verrata Transfermarkt-arvoihin tarkkuuden arvioimiseksi vaan toteutuneisiin siirtosummiin. Malli muodostaa ennustetun hinnan koneoppimismetodeja hyödyntämällä, perustuen pelaajien ominaisuuksiin, suorituskykyindikaattoreihin ja pelaajan kuuluisuutta mittaaviin indikaattoreihin, kuten Reddit julkaisuihin ja Wikipedia-sivun katselukertoihin. Malli tuotti matala- ja keskihintaisista pelaajista hinta-arvioita, jotka olivat lähempänä toteutuneita siirtosummia kuin Trasfermarkt-arviot. Transfermarkt-arviot olivat kuitenkin lähempänä toteutuneita siirtosummia koskien korkeahintaisia pelaajia. Korkeahintaisten pelaajien virhearvioinnit johtuvat supertähtien kuuluisuudesta johtuvan lisäarvon mallintamisen vaikeudesta. Myös seurojen käyttäytyminen pelaajamarkkinoilla on poikkeavaa koskien korkeahintaisia supertähtiä. Seurat saattavat maksaa pelaajasta yli hänen todellisen arvonsa, jos he kokevat, että pelaaja on ainut laatuaan tai kyvyiltään ylivertainen. Mallin heikkoutena onkin toteutuneiden siirtosummien ja markkina-arvon välinen häilyvä suhde. Pelaajan todellinen markkina-arvo on pelaajan omakohtainen näkymätön tekijä, jota ei voida määrittää samalla tavalla kuin siirtosummaa, johon vaikuttaa pelaajan ominaisuuksien ja kykyjen lisäksi joukkueiden väliset taloudelliset tekijät ja markkinoiden tilanne.

3.3 Pelaajien alihinnoittelun hyödyntäminen

Monet hinnoittelumallit antavat seuroille mahdollisuuden hyödyntää löydetyt markkinoiden epätehokkuudet. Hinnoittelumallien avulla löydetyt pelaajat, jotka ovat mallin perusteella arvokkaampia, kuin heidän Transfermarkt-arvonsa ovat esimerkkejä markkinoiden epätehokkuuksista, joita seurat voivat hyödyntää. Mallit ovat jäävät kuitenkin usein teoreettiselle tasolle, eikä niissä ole selkeää viitekehystä itse seuran toiminnalle.

Lechner ja Gudmundsson (2012) määrittelevät tutkimuksessaan urheiluseuran arvonluontia ja kilpailukyvyn syntymistä. Ihmisresurssit synnyttävät lisäarvon jalkapallojoukkueissa. Ihmisresurssit (jalkapallon tapauksessa pelaajat) ovat liikkuvia, joten iso osa seurojen arvonluonnista muodostuu näiden resurssien hankkimisen tehokkuudesta. Tutkimuksessaan Lechner ja Gudmundsson löysivät, että vaikka sisäisesti tuotetut resurssit (seuran omat kasvattipelaajat) voivat olla tärkeitä, heidän käytöllään joukkueessa on vahva, suora negatiivinen yhteys arvonluomiseen. Tämä selittyy sillä, että joukkueiden omien resurssien tuottamista pitää verrata kaikkien muiden seurojen resurssien tuottamiseen. Joukkueet, jotka tuottavat omat resurssinsa tekevät usein niin taloudellisen pakon vuoksi. Joukkueet, jotka eivät osallistu pelaajamarkkinoille jäävät siis paitsi ylimääräisestä arvonluomisesta. Pelaajamarkkinoilla seurat voivat hankkia resursseja, jotka ovat paitsi korkeatasoisia, myös valikoituja yhteensovittamisen kannalta. Tutkimuksessa myös todettiin, että managerin (engl. Manager, jossain tapauksissa suomennettavissa myös päävalmentajaksi) taidot ja kokemus ovat tärkeitä seuran arvonluomiseen. Managerit voivat lisätä joukkueen resurssien panosta tarjoamalla opetusta ja neuvontaa pelaajille. Managerit ovat myös oleellinen osa seuran resurssiportfolion strukturointia (pelaajaprofiilien ja taktiikoiden yhteensopivuus). Selkeä visio mielessä koottu joukkue on isompi kuin sen osasten summa.

Pantuso (2017) määrittelee seurojen strategisen ongelman olevan käytettävissä olevan budjetin sijoittaminen pelaajiin, jotta seuran odotettu arvo maksimoituu. Seuroilla on tarve hankkia pelaajia, joilla on tietyt ominaisuudet ja kyvyt sekä mahdollisuus kasvattaa omaa arvoaan huolimatta riskitekijöistä. Tämä ongelma on kuvailtu Jalkapallojoukkueen kokoamisongelmaksi (engl. Football Team Composition Problem, FTCP). Pantuso kehitteli mallin, jonka tavoitteena on optimoida seuran sijoitukset pelaajiin, ottaen huomioon mahdolliset riskitekijät. Manageri on yleensä vastuussa joukkueen kokoamisesta ja täten määrittelee minkälaisen pelaajan seura ostaa. Seura muodostaa listan kohdepelaajista, jotka täyttävät tietyt tilastolliset kriteerit. Kohdepelaajista on tiedossa heidän nykyinen seuransa, hinta ja palkka. FTCP:n osana on näistä kohdepelaajista parhaan valinnan tekeminen. Paras valinta on sellainen, joka sopii budjettiin, kilpailun sääntöihin ja managerin vaatimuksiin. FTCP malli mittaa onnistunutta strategiaa seuran pelaajien markkina-arvojen kasvuna. FTCP mallia testattiin vertailemalla sen antamia pelaajahankintavalintoja toteutuneisiin vuoden 2013/2014 Valioliigassa. Malli paransi useimpien seurojen kohdalla markkina-arvoa realistisella aikajanalla. Mallin ehdottamien hankintojen taloudellinen menestys perustui laajalti nuorien lupaavien pelaajien ostamiseen.

4 YHTEENVETO JA JOHTOPÄÄTÖKSET

Tutkielman tutkimusongelma oli:

- Miten jalkapalloseurat voivat parantaa urheilullista menestystään löytämällä alihinnoiteltuja pelaajia datan avulla?

Tutkimusongelmaan löydettiin vastaus tarkastelemalla seuraavia tutkimuskysymyksiä:

- Minkälaista dataa on saatavilla pelaajista?

Teknologian kehitys on mahdollistanut kehittyneiden datamallien muodostamisen pelaajista. Datamalleja voidaan käyttää tunnistamaan tietynlaisia ominaisuuksia tai piilotettuja kyvykkyyksiä pelaajista.

- Miten joukkueet voivat löytää alihinnoiteltuja pelaajia datan avulla?

Joukkueet voivat löytää kehittyneillä datamalleilla tai hinnoittelumalleilla pelaajia, jotka ovat alihinnoiteltuja. Pelaajat ovat tärkeä henkilöstöresurssi seuralle, joka pyrkii muodostamaan kestävää kilpailuetua eli urheilullista menestystä.

Jalkapallo kautta voidaan tarkastella data-analytiikan kehittymistä ja sen käyttökohteita. Jalkapallossa teknologian kehitysaskeleet ovat vaikuttaneet datavetoisen lähestymistavan mahdollistamiseen enemmän kuin monissa muissa urheilulajeissa. Vaikka jalkapallossa pelaajamarkkinoilla liikkuvat summat ovat nykypäivänä kymmenissä tai jopa sadoissa miljoonissa, monet joukkueet maksimoivat taloudellisen voiton sijaan urheilullista menestystä. Urheilullisen menestyksen ensisijaseksi asettaminen tarkoittaa, että monet joukkueet ovat velkavetoisia sekä taloudellisesti kannattamattomia. Datavetoinen pelaajahankinta mahdollistaa tehokkaan ja taloudellisesti kannattavan toiminnan äärimmäisen kilpailullisilla markkinoilla. Huippujalkapallossa yksi ottelu tai yksi maali voi tarkoittaa miljoonien tuloja tai tulojen menetystä. Vaikka urheilullinen ja taloudellinen hyöty kulkevatkin jalkapallossa käsi kädessä, on tämän tutkielman kontekstissa arvioitu pelaajahankintoja puhtaasti urheilullisen menestyksen kannalta, se tarkoittaa, että esimerkiksi pelaajien suorituskykyyn liittyviä mittareita voidaan tarkastella oleellisena (ja välittömänä) osana pelaajahankintaprosessia tutkimuskysymyksiin vastattaessa.

Tutkielman toisessa luvussa tarkastelin pelaajamittareita ja datamalleja, joille on tieteellisesti muodostettu yhteys urheilulliseen menestykseen. Tavoitteeni oli vastata tutkimuskysymykseen: Minkälaista dataa pelaajista on saatavilla? Tutkin seurojen motivaatioita ja tavoitteita. Monet eurooppalaiset jalkapalloseurat eivät noudata tavanomaisen yritystoiminnan lainalaisuuksia, vaan koettavat taloudellisen voiton sijaan maksimoida urheilullista menestystä. Pelaajat ovat oleellinen osa järjestelmää, joka tuottaa lisähyötyä (voittomaksimisoivissa seuroissa lisäävät voittoja). Olen muodostanut yhteyden ja perustan pelaajien kyvyille ja urheilulliselle menestykselle teorian kautta tarkastelemalla pelaajia Barneyn resurssiperusteisella näkökulmalla. Resurssiperusteisessä näkökulmassa yritysten kilpailuetu muodostuu yrityksen käytettävissä olevien resurssien hyödyntämisestä. Tietynlaisilla aloilla

(esimerkiksi joukkueurheilussa) henkilöstöresurssit eli työntekijöiden kyvykkyydet ovat fyysisiä resursseja tärkeämpiä lisäarvon tuottamisessa. Kilpailu näistä resursseista on erittäin kovaa. Eurooppalaisten jalkapalloseurojen filosofiat sekä pelaajamarkkinoiden luonne tarkoittaa kasvaneita siirtosummia ja taloudellisia riskejä.

Seurojen strateginen ongelma on huippukykyisten pelaajien hankkiminen erittäin kilpailullisilla markkinoilla. Menestyminen edellyttää joko isoa budjettia tai kykyä tunnistaa ja hankkia kyvykkäitä pelaajia alihintaan. Toisessa luvussa käsittelin datamalleja pelaajien mittaamiseen. Data-analytiikkaa voidaan hyödyntää jalkapallossa monella tavalla, mutta tämän tutkielman kontekstissa tarkastelin sen merkitystä pelaajahankintaan. Data-analytiikka pelaajahankinnassa pyrkii ratkaisemaan seurojen strategisen ongelman tunnistamalla kykyjä pelaajissa, joita perinteiset kykyjenetsintämetodit kuten asiantuntijan mielipiteet tai kykyjenetsijöiden intuitio eivät löydä. Data-analytiikan historia urheilussa on melko lyhyt. Sen juuret ovat amerikkalaisessa urheilussa, erityisesti baseballissa. Jalkapallopelaajista on baseballpelaajia vaikeampaa muodostaa datamalleja ja mittareita, jotka ovat selkeitä, mutta eivät ole yleisessä tiedossa. Jalkapallo onkin ollut muita urheilulajeja jäljessä data-analytiikan hyödyntämisessä. Teknologian kehitys on kuitenkin mahdollistanut uudenlaisten mittareiden ja datamallien kehittämisen.

Pelaajista saatava data on nykyään paikkaan ja aikaan sidottua, mikä mahdollistaa pelaajien tilastoissa pelitapahtumien kontekstin huomioimisen. Dataa voidaan tallentaa enemmän ja se on tarkempaa ja monimuotoisempaa kuin koskaan ennen. Koneoppimisen kehittyminen mahdollistaa uudenlaisten mallien kouluttamisen tämän datan perusteella. Mallit antavat painotetut arvot tietyille kenttätapahtumille perustuen aiempaan dataan. Maaliodottama (xG) on koneoppimisella koulutettu malli, joka kuvastaa laukauksen odotettua todennäköisyyttä päätyä maaliin. xG on laajalti sovellettavissa erilaisiin datamalleihin, joiden kautta voidaan tarkastella maalintekopaikoille pääsyä, maalintekopaikkojen luomista toisille ja maalintekotaitoa. xG on tärkeä mittari varsinkin hyökkääjille, joiden tehtävä on tehdä maaleja.

Laukauksien lisäksi myös syötöille voidaan määrittää konkreettinen arvo. Monet mallit tarkastelevat syöttöjen vaikutusta pallonpitojaksoihin ja todennäköisyyksiin, että pallonpitojakso tai tietty syöttö päättyy laukaukseen. Syötöt syöttöketjuissa voidaan arvottaa todennäköisyyksien muutosten avulla. Syöttöketjut voidaan nähdä myös virtauksina, joita voidaan tarkastella verkostoteorian kannalta. Pelaajien keskeisyys näille virtauksille toimii indikaattorina pelaajien osallisuudesta joukkueen menestykseen. Pelaajien merkitystä joukkueelle voidaan myös tarkastella tilankäytön ja -luomisen kannalta. Jalkapallossa pelaajat ovat useimmiten ilman palloa. Pallotonta tekemistä on ollut aiemmin vaikea kvantifioida, mutta tilankäytön mallit antavat siihen työkalut. Pelaajat voivat pallottomalla liikkeellään luoda tilaa itselleen tai joukkuetovereille.

Suurin osa malleista suosii hyökkääjiä. Hyökkääjien toiminnoista jää selkeä datajälki, ja toiminnot voidaan painottaa selkeän lopputuleman (maalin) kannalta. Puolustajat voivat oikeanaikaisella sijoittautumisella ja reagoinnilla suojella omaa maaliaan ilman, että tapahtumasta syntyy minkäänlaista konkreettista dataa. Jotkut mallit kuten PlayeRank ja VAEP pyrkivät kuitenkin pelaajien objektiiviseen arviointiin pelipaikasta riippumatta, antamalla kaikenlaisille kenttätapahtumille tietyt arvot.

Tutkielman toinen luku on siis ollut laajalti erilaisten mallien tutkimista ja arviointia. Tavoitteenani oli eri mallien yleiskatsauksen lisäksi tuoda ymmärrystä data-analytiikan historiasta, teknologian kehityksestä ja sen merkityksestä. Luvun jälkeen lukijalla on hyvä yleistietämys erilaisista datamalleista, joita pelaajista voidaan muodostaa ja miksi nämä mallit ovat hyödyllisiä ja tutkimisen arvoisia.

Vaikka näitä malleja voidaan itsessään käyttää Moneyball-tyyliseen kyvykkäiden ja halpojen pelaajien tunnistamiseen, koen tarpeelliseksi tutkimusongelman kokonaisvaltaisen selvittämisen kannalta tutkia myös malleja, jotka hinnoittelevat pelaajia ja selvittää miten alihinnoiteltujen pelaajien löytämistä voidaan hyödyntää joukkueenkokoamisessa.

Tutkielman kolmas luku keskittyy pelaajien hinnoitteluun liittyviin kysymyksiin. Tavoitteena on saada vastaus tutkimuskysymykseen: Miten alihinnoiteltuja pelaajia voidaan löytää datan avulla? Pohjustin aihetta käymällä läpi jalkapallon pelaajamarkkinoiden toimintaa ja sääntöjä. Suurin osa pelaajista vaihtavat joukkuetta siirtokorvausta vastaan. Kun seura on maksanut toiselle seuralle siirtosumman, pelaajan ja uuden seuran väliset sopimusneuvottelut voidaan aloittaa. Bosman-tapaus oli merkittävä pelaajamarkkinoiden luonteen kannalta. Vähentynyt säännöstely pelaajien työntekijäaseman myötä tarkoitti lisääntynyttä aktiviteettiä pelaajamarkkinoilla. Pelaajien jäljellä olevilla vuosilla oli tapauksen myötä merkittävä vaikutus pelaajan markkina-arvoon. Seurat eivät halua menettää pelaajaa ilmaiseksi, joten ne usein myyvät pelaajan halvalla, jos hänellä on alle vuosi sopimusta jäljellä, eivätkä uuden sopimuksen neuvottelut ole edenneet.

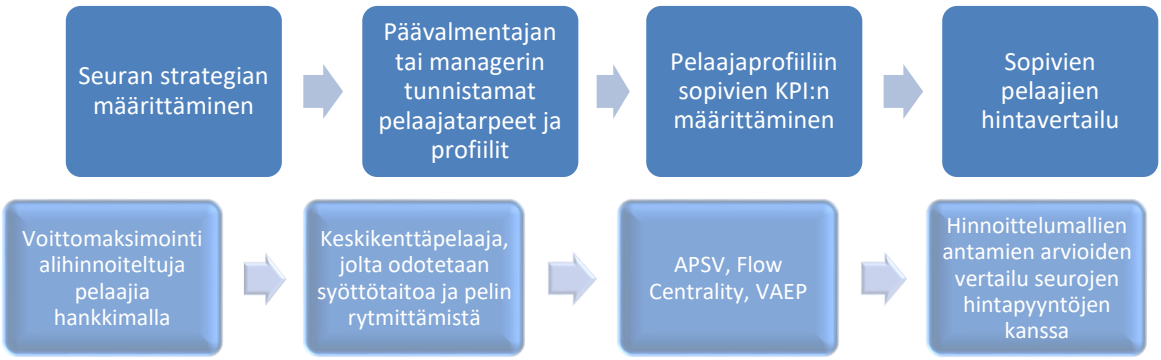
Pelaajan markkina-arvo on arvio pelaajasta maksettavasta siirtokorvauksesta. Markkina-arvoon vaikuttaa pelaajan suorituskyvyn ja ominaisuuksien lisäksi markkinavoimat ja taloudelliset tekijät. Pelaajan markkina-arvolle on määritelty tiettyjä indikaattoreita. Monessa tutkimuksessa on tunnistettu samat ominaisuudet tai kyvykkyydet markkina-arvon ajajina. Indikaattorit voidaan jakaa kolmeen kategoriaan: pelaajan kyvykkyydet, pelaajan ominaisuudet ja pelaajan kuuluisuus. Pelaajan kyvykkyyksiä kuvastaa erilaiset tilastot ja mittarit. Suurin yhteys markkina-arvoon on pelaajan tehdyillä maaleilla, maaliin johtaneilla syötöillä ja kuljetuksilla. Pelaajan ominaisuudet ovat faktatietoa pelaajasta, kuten pelaajan ikä, pelipaikka, pituus ja kansallisuus. Pelaajan markkina-arvo yleisesti nousee, kunnes pelaaja on noin 26-vuotias, jonka jälkeen se laskee. Pelaajan pituus korreloi positiivisesti markkina-arvon kanssa. Kuuluisat pelaajat ovat markkina-arvoltaan korkeampia. Pelaajan julkisuus

tuo joukkueelle imago ja markkinointihyötyä ja lisää lippu- ja paitamyyntiä. Kuuluisuutta voidaan mitata esimerkiksi internetissä tapahtuvien mainintojen, klikkausten ja artikkeleiden kautta.

Monet mallit vertaavat omia hinnoittelujaan pelaajien Transfermarkt arvoihin. Transfermarkt arvot perustuvat yleisön, eli nettisivun käyttäjien, antamien hinta-arvioiden painotettuun keskiarvoon. Transfermarkt arvot ovat yleisesti olleet melko tarkkoja ennustamaan toteutuneita siirtosummia, ja ne ovatkin usein mediassa esillä. Transfermarkt arvojen tarkkuus tarkoittaa, että niitä käytetään monissa tutkimuksissa vertailun kohteena tai pohjana omille arvioille. Itse hinnoittelumallit perustuvat taloustieteellisiin laskelmiin, pelaajien markkina-arvon indikaattoreiden painoarvojen määrittämiseen tai niiden yhdistelmiin. Malleissa nousevat tietyt markkina-arvojen ajajat esille. Ikä, maalintekotaito ja Englannin Valioliigassa pelaaminen korreloivat voimakkaasti markkina-arvojen kanssa useimmissa malleissa. Malleja voidaan käyttää määrittämään epätarkkuudet pelaajien markkina. Epätarkkuuksia voidaan hyödyntää alihinnoiteltuja pelaajia palkkaamalla tai ylihinnoiteltuja pelaajia myymällä.

Mallit ovat kuitenkin usein teoreettisia, eikä sellaista kattavaa tieteellistä katsausta tai viitekehystä ole tehty, missä eri malleja voitaisiin hyödyntää seuran kannalta. Voin tutkielmassani käytyjen tutkimusten valossa määrittämään mahdollisen viitekehyksen urheilujohtajalle, managerille tai kykyjenetsintäpäällikölle, siitä miten pelaajahankintaa voitaisiin tehdä datavetoisesti. Datavetoisen lähestymistavan tarkoitus on löytää muille piilotettuja kyvykkyyksiä markkinoilta. Vaikka tietyt joukkueet ovatkin voittomaksimoivia, on niiden strateginen ongelma linkittynyt oman budjetin maksimoimiseen pelaajamarkkinoilla, tämä tarkoittaa, että myös hinnoittelumallit ovat oleellinen osa pelaajahankintaprosessia.

Taulukko 3 Viitekehys datavetoisen pelaajahankintaprosessin vaiheista esimerkeillä



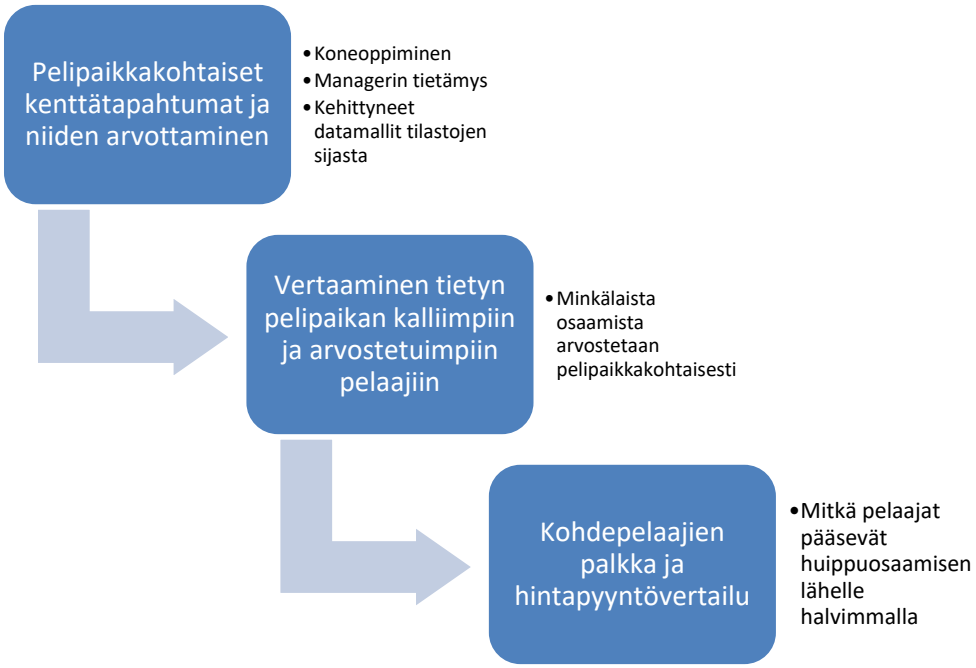
Pelaajahankintaprosessi alkaa seuran strategian määrittämisellä. Seuralla, jonka tavoite on taloudellisen voiton maksimointi, on todennäköisesti erilainen pelaajahankintaprosessi. Tämän tutkielman kontekstissa tarkastelin seuraa, joka voittomaksimoi alihinnoiteltuja pelaajia hankkimalla. Jotta seura hyödyntäisi sen resursseja mahdollisimman tehokkaasti, on sen joukkue koottava järkevästi pelipaikoittain. Manageri tai päävalmentaja on tästä vastuussa, ja se on osittain tämän tutkielman laajuuden ulkopuolella. Käytetään kuitenkin esimerkkinä tilannetta, jossa manageri tai päävalmentaja on tunnistanut tarpeelliseksi lisätä joukkueeseen keskikenttäpelaajan, jolta odotetaan syöttötaitoa ja pelin rytmittämistä. Eri pelaajarooleille ja pelipaikoille on tärkeää löytää sopivat ja omanlaiset KPI:t. Esimerkkitilanteessamme keskikenttäpelaajan syöttötaitoja, sekä pelinrytmittämistä on valittu mittaamaan APSV ja Flow Centrality mallit. Näiden lisäksi voidaan tarkastella esimerkiksi pelaajan syötöistä luotua maaliodottamaa. VAEP on yleisempi malli, jolla saadaan kuva pelaajan taitotasosta ja merkityksestä nykyiselle joukkueelle. Näille tietyille mittareille asetettuihin tavoitteisiin yltää todennäköisesti useampi pelaaja. Tässä vaiheessa on todennäköisesti kannattavaa, että pelaajien suorituksia käydään katsomassa paikan päällä. Niin sanotulla silmätestillä (engl. eye test) voidaan havaita, onko datassa vääristymiä tai vaikuttaako jokin muu asia, kuin pelaajan omat ominaisuudet, hänen tuottamaan dataan. Silmätestillä voidaan sen lisäksi selvittää mittaamattomia asioita, kuten pelaajan asennetta, luonnetta ja käyttäytymistä kentällä. Datan ja silmätestin perusteella sopivien pelaajien välillä tehdään sen jälkeen hintavertailut eri hinnoittelumalleja käyttämällä. Tavoitteena on selvittää, onko jokin ehdokkaista alihinnoiteltu verrattuna hänen nykyisen seuransa hintapyyntöön nähden. Kun paras ehdokas on löydetty ja siirtosummasta sovittu, voidaan aloittaa sopimusneuvottelut. Vastaavanlaista viitekehystä voitaisiin tarkastella empiirisesti seurakohtaisesti tai esimerkiksi Pro Gradu-tutkimuksessa.

Kahden sisältöluvun ja johtopäätösosiossa olevan viitekehyksen kautta olen saanut vastauksen tutkimusongelmaan. Seurat voivat parantaa sarjasijoitustaan valjastamalla käyttöönsä lisähyödyn, jota monimuotoiset data- ja hinnoittelumallit mahdollistavat. Pelaajat ovat oleellinen osa kilpailuetua, johon joukkueet pyrkivät. Data-analytiikka antaa seuroille mahdollisuuden kilpailla hyödyntämällä heidän pelaajasiirtoihinsa varatun pelaajabudjetin mahdollisimman tehokkaasti. Tutkielmani rajoitteena olivat teoriapohjaiset ja matemaattiset mallit, joiden ymmärtäminen ja soveltaminen pelaajahankintaprosessiin on hankalaa ja työlästä. Tutkimani mallit keskittyivät hyökkääjien pelaajien arviointiin. Puolustavien pelaajien onnistumisesta ei välttämättä seuraa minkäänlaista selkeää datajälkeä.

Useat mallit pohjaavat joko pelaajien kyvykkyyksien tai markkina-arvon kohdalla mittareihin, jotka ovat liian yksiselitteisiä, ja siksi ne eivät ole käytännönläheisiä tai tarkkoja. Esimerkiksi pelaajien kyvykkyyksiä on mitattu tutkimuksissa perustuen FIFA-videopelisarjan tarjoamiin arvoihin pelaajan taitojen osalta. Joissain malleissa on päästy hyvin lähelle mallia, josta jalkapalloseura voisi

konkreettisesti hyötyä, mutta jatkotutkimuksen osalta onkin ehdotettu laajentamista toisiin urheilulajeihin. Pelaajan arviointiin käytettäviä malleja ei pitäisi laajentaa, vaan kaventaa. Lajien erot ovat niin suuret, että on mahdotonta muodostaa mitään yleistä mallia, joka toimii eri lajeissa. Nykyään olemassa olevista datamalleista huomaamme, että ne ovat painottuneet hyökkäävien pelaajien arviointiin. Tulevaisuuden tutkimusala voisi olla puolustajien arviointiin keskittyvä malli. Pelaajahankintaprosessia koskevan viitekehyksen heikkoutena onkin puolustajille ja puolustaville keskikenttäpelaajille määritettävät pelaajamittarit. Lisätutkimus, joka toisi lisähyötyä alan tutkimukseen olisi malli, joka määrittelee pelaajien arvot arvottamalla yksityiskohtaisempia kenttätapahtumia ja pelaajamittareita, perinteisten tilastojen käyttämisen sijaan. Tämänlainen tutkimus voisi tuoda esille myös hinnoittelumallien konkreettisen hyödyn, eli alihinnoiteltujen pelaajien ostamisen.

Taulukko 3 Kehittyneiden suorituskykymittareiden yhdistäminen pelaajahankintaprosessiin



Monien mallien keskeisenä ongelmana oikean elämän käyttötarkoituksen kannalta on kykenemättömyys nostaa esille tuntemattomia pelaajia datan avulla. Mallit nostavat esille pelaajia, joiden kyvyt ovat yleistä tietoa (nämä pelaajat ovat erittäin kalliita). Toisaalta malli toimii todistetusti pelaajan kykyjen arviointiin, jos se nostaa esille näitä supertähtiä, mutta yhtä lailla se ei anna mitään lisätietämystä tai hyötyä kenellekään. Tulevaisuuden tutkimuksissa lisähyötyä itse alan ihmisille tarjoaisi selkeät esimerkit pelaajista, jotka ovat alihinnoiteltuja tai tuntemattomia, mutta jotka malli tunnisti kyvykkäiksi. Myös erilaisten viitekehysten ja prosessikaavioiden kehittäminen olisi seurojen päättäjille hyödyksi.

LÄHTEET

Antonioni, P. (2000). *The Bosman Ruling and the Emergence of a Single Market in Soccer Talent*. European Journal of Law and Economics, 9:2; 157]173

Anzer, G., & Bauer, P. (2021). A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.624475

Aurenhammer, F. (1991). Voronoi diagrams—A survey of a fundamental geometric data structure. *ACM Computing Surveys*, *23*(3), 345–405. https://doi.org/10.1145/116873.116880

Behravan, I., & Razavi, S. M. (2021). A novel machine learning method for estimating football players' value in the transfer market. *Soft Computing*, *25*(3), 2499–2511. https://doi.org/10.1007/s00500-020-05319-3

Brooks, J., Kerr, M., & Guttag, J. (2016). Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49–55. https://doi.org/10.1145/2939672.2939695

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, *25*(12), 1467–1476. https://doi.org/10.1080/00036849300000150

Cohen, A., & Risk, J. (2024). *European Football Player Valuation: Integrating Financial Models and Network Theory* (No. arXiv:2312.16179). arXiv. https://doi.org/10.48550/arXiv.2312.16179

Davis, J., Bransen, L., Devos, L., Jaspers, A., Meert, W., Robberechts, P., Van Haaren, J., & Van Roy, M. (2024). Methodology and evaluation in sports analytics: Challenges, approaches, and lessons learned. *Machine Learning*, *113*(9), 6977–7010. https://doi.org/10.1007/s10994-024-06585-0

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals: Valuing Player Actions in Soccer. *Proceedings of the 25th ACM SIGKDD International*

Conference on Knowledge Discovery & Data Mining, 1851–1861. https://doi.org/10.1145/3292500.3330758

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the Performance of Individual Players in a Team Activity. *PLOS ONE*, *5*(6), e10937. https://doi.org/10.1371/journal.pone.0010937

Fernandez, J., Barcelona, F. C., & Bornn, L. (2018). *Wide Open Spaces: A statistical technique for measuring space creation in professional soccer*. Sloan sports analytics conference

Franceschi, M., Brocard, J.-F., Follert, F., & Gouguet, J.-J. (2024). Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys*, *38*(3), 577–600. https://doi.org/10.1111/joes.12552

Franck, E., & Nüesch, S. (2012). Talent and/or Popularity: What Does It Take to Be a Superstar? *Economic Inquiry*, *50*(1), 202–216. https://doi.org/10.1111/j.1465-7295.2010.00360.x

Frick, B. (2007). The Football Players' Labor Market: Empirical Evidence from the Major European Leagues. *Scottish Journal of Political Economy*, *54*(3), 422–446. https://doi.org/10.1111/j.1467-9485.2007.00423.x

Garcia-del-Barrio, P., & Pujol, F. (2007). Hidden monopsony rents in winner-take-all markets—Sport and economic contribution of Spanish soccer players. *Managerial and Decision Economics*, *28*(1), 57–70. https://doi.org/10.1002/mde.1313

Goes, F. R., Kempe, M., Meerhoff, L. A., & Lemmink, K. A. P. M. (2019). Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data*, *7*(1), 57–70. https://doi.org/10.1089/big.2018.0067

Gudmundsson, J. (2012). *Football analysis using spatio-temporal tools*. SIGSPATIAL '12: Proceedings of the 20th International Conference on Advances in Geographic Information Systems

Pages 566 - 569

https://doi.org/10.1145/2424321.2424417

Gudmundsson, J., & Horton, M. (2018). Spatio-Temporal Analysis of Team Sports—A Survey. *ACM Computing Surveys*, *50*(2), 1–34. https://doi.org/10.1145/3054132

Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). *Testing Causality Between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer*. Economics: Faculty Publications Smith College

He, M., Cachucho, R., & Knobbe, A. (2015). *Football Player's Performance and Market Value*. LIACS, Leiden University, the Netherlands

Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, *17*(4), 484–492. https://doi.org/10.1016/j.smr.2013.12.006

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, *14*(6), 798–817. https://doi.org/10.1177/1747954119879350

Kang, C., Hwang, J., & Li, K. (2006). Trajectory Analysis for Soccer Players. *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)*, 377–381. https://doi.org/10.1109/ICDMW.2006.160

Késenne, S. (2006). The Win Maximization Model Reconsidered: Flexible Talent Supply and Efficiency Wages. *Journal of Sports Economics*, *7*(4), 416–427. https://doi.org/10.1177/1527002505279347

Lechner, C., & Gudmundsson, S. V. (2012). Superior value creation in sports teams: Resources and managerial experience. *M@n@gement*, *15*(3), 284–312.

Lehmann, E. E., & Schulze, G. G. (2008). What Does it Take to be a Star? – The Role of Performance and the Media for German Soccer Players. *Applied Economics Quarterly*, *54*(1), 59–70. https://doi.org/10.3790/aeq.54.1.59

Li, C., Kampakis, S., & Treleaven, P. (2022). *Machine Learning Modeling to Evaluate the Value of Football Players* (No. arXiv:2207.11361). arXiv. https://doi.org/10.48550/arXiv.2207.11361

Li, Y. (2021). *When Moneyball Meets the Beautiful Game: A Predictive Analytics Approach to Exploring Key Drivers for Soccer Player Valuation*. Submitted in partial fulfillment of the requirements for the degree of Master of Science in Management (Operations and Information Systems) Goodman School of Business, Brock University

McNamara, P., Peck, S. I., & Sasson, A. (2013). Competing Business Models, Value Creation and Appropriation in English Football. *Long Range Planning*, *46*(6), 475–487. https://doi.org/10.1016/j.lrp.2011.10.002

Mustafi, Z., Bayle, E., & Terrien, M. (2025). Non-Big five football clubs' strategies for generating transfer revenues: The case of Switzerland's Super League. *Soccer & Society*, *26*(1), 62–79. https://doi.org/10.1080/14660970.2024.2374254

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, *263*(2), 611–624. https://doi.org/10.1016/j.ejor.2017.05.005

Oberstone, J. (2009). Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. *Journal of Quantitative Analysis in Sports*, *5*(3). https://doi.org/10.2202/1559-0410.1183

Pantuso, G. (2017). The Football Team Composition Problem: A Stochastic Programming approach. *Journal of Quantitative Analysis in Sports*, *13*(3), 113–129. https://doi.org/10.1515/jqas-2017-0030

Pantuso, G., & Hvattum, L. M. (2020). *Maximizing performance with an eye on the finances: A chance-constrained model for football transfer market decisions* (No. arXiv:1911.04689). arXiv. https://doi.org/10.48550/arXiv.1911.04689

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach. *ACM Transactions on Intelligent Systems and Technology*, *10*(5), 1–27. https://doi.org/10.1145/3343172

Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series A (General)*, *131*(4), 581. https://doi.org/10.2307/2343726

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, *5*(1), Article 1. https://doi.org/10.1186/s40064-016-3108-2

Schlenger, J., Wunderlich, F., Raabe, D., & Memmert, D. (2023). Systematic Analysis of Position-Data-based Key Performance Indicators. *International Journal of Computer Science in Sport*, *22*(1), 80–101. https://doi.org/10.2478/ijcss-2023-0006

Sæbø, O. D., & Hvattum, L. M. (2019). Modelling the financial contribution of soccer players to their clubs. *Journal of Sports Analytics*, *5*(1), 23–34. https://doi.org/10.3233/JSA-170235

Taki, T., & Hasegawa, J. (2000). Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings Computer Graphics International 2000*, 227–235. https://doi.org/10.1109/CGI.2000.852338

Tax, N. & Yme Joustra. (2015). *Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach*. Unpublished. https://doi.org/10.13140/RG.2.1.1383.4729

Tunaru, R., Clark, E., & Viney, H. (2005). An option pricing framework for valuation of football players. *Review of Financial Economics*, *14*(3–4), 281–295. https://doi.org/10.1016/j.rfe.2004.11.002

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Yalçınkaya, M. A., & Işık, M. (2024). The Role of Performance Metrics in Estimating Market Values of Footballers in Europe's Top Five Leagues. *Pamukkale Journal of Sport Sciences*, *15*(3), 455–485. https://doi.org/10.54141/psbd.1489554