



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Datapohjainen päätöksenteko jalkapalloseurojen pelaajahankinnassa

Tietojärjestelmätieteen kandidaattitutkielma

Laatija:

Justus Knuutila

Ohjaaja:

Tohtori Samuli Laato

10.12.2025

Turku

Opiskelijan lausunto tekoälyn käytöstä tähän tutkielmaan liittyen:

En ole käyttänyt tekoälyä hyödyntäviä työkaluja tätä tutkielmaa kirjoittaessani.

Olen käyttänyt tekoälyä hyödyntäviä työkaluja tätä tutkielmaa kirjoittaessani. Tämä käyttö on dokumentoitu tutkielman liitteessä. Vakuutan, että tekoälyä käytettiin yliopiston ohjeistuksen mukaisella tavalla.

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

Kandidaatintutkielma

Oppiaine: Tietojärjestelmätiede

Tekijä: Justus Knuutila

Otsikko: Datapohjainen päätöksenteko jalkapalloseurojen pelaajahankinnassa

Ohjaaja: Tohtori Samuli Laato

Sivumäärä: 31 sivua (+ liitteet 1 sivua)

Päivämäärä: 10.12.2025

Tiivistelmä

Ammattilaisjalkapallo on vuosien mitassa kasvanut kymmenien miljardien kokoiseksi bisnekseksi ja pelkääntään pelaajasiirtoihin käytetään miljardeja vuosittain. Lisäksi data-analytiikka ja koneoppiminen ovat kasvaneet merkitystään usealla osa-alueella mukaan lukien pelaajahankinnassa. Niiden merkitys päätöksenteon tukena ja kilpailukyvyyn säilyttämisessä taloudellisesti ja urheilullisesti on valtava. Tutkielmassa tarkastellaan vertaisarvioidun kirjallisuuden valossa, miten jalkapalloseurat voivat valjastaa moniulotteisen datan käyttöönsä ja miten se voidaan jalkauttaa tukemaan päätöksentekoa pelaajahankinnoissa. Tutkielma on kirjallisuuskatsaus keskittyen erityisesti datamuotoihin, datan keräysmenetelmiin, sen hyödyntämiseen pelaajien arvioinnissa ja osana päätöksentekoa organisaatioissa. Datamuoto riippuu paljon siitä, mitä halutaan tutkia. Tutkielman perusteella usein käytettyjä datamuotoja pelaajahankinnan kontekstissa ovat tapahtuma-, sijainti- ja videodata. Dataa analysoimalla mitataan pelaajan tai joukkueen suoriutumista, tutkitaan taktisia аспекteja sijaintidatan avulla ja kehitetään kokonaisvaltaisia suoriutumisen mittareita. Näitä menetelmiä hyödynnetään pelaajahankinnassa arvioimaan pelaajan sopivuutta joukkueen taktiseen viitekehykseen tai ennustamaan hänen markkina-arvonsa kehittymistä. Haasteet datan hyödyntämisessä liittyvät saatavuuteen, sen keräämismenetelmiin ja lopulliseen käyttötarkoitukseen sekä ymmärtämättömyyteen kontekstista, jossa data on kerätty. Johtopäätöksenä data-analyysistä on tullut osa jalkapalloseurojen pelaajahankintaa. Se auttaa seuroja vähentämään riskiä pelaajahankinnassa ja edistää kilpailukykyistä asemaa markkinoilla.

Avainsanat: jalkapallo, data-analytiikka, koneoppiminen, pelaajahankinta

SISÄLLYS

1	Johdanto	7
2	Datatyypit ja sen kerääminen pelaajahankinnan tueksi	9
2.1	Data – tärkeä osa modernia pelaajahankintaa	9
2.2	Datan eri muodot ja käyttötarkoitukset	10
2.3	Datan saatavuuden ja laadun asettamat haasteet	11
3	Datan käsittely ja hyödyntäminen pelaajahankinnassa	14
3.1	Pelaajahankintaa tukevat analyysit	14
3.1.1	Suoriutumisen analysointi ja siihen käytettävät mittarit	15
3.1.2	Markkina-arvoanalyysi	17
3.1.3	Pelaajien taktinen roolitus ja pelityylianalyysi	18
3.2	Data-analyysien käyttö pelaajahankinnan tukena / Mallien käyttö pelaajien arvioinnissa	20
3.3	Päätöksenteon prosessi	22
4	Yhteenveto ja johtopäätökset	25
5	Lähteet	27

KUVIOT

Kuva 1 Pelaajahankintaprosessin eteneminen (mukaillen Kim ym., 2021)	23
Kuva 2 Datan käsittelyprosessi (mukaillen Mănescu, 2025)	24

TAULUKOT

Taulukko 1 Analyysin kohteet ja menetelmäkuvaus	15
---	----

1 Johdanto

Ammattilaisjalkapallosta on vuosikymmenten mitassa kasvanut miljardiluokan kokoinen bisnes (Aalbers & Van Haaren, 2019). Deloitteen mukaan (2025) ammattilaisjalkapallon kokonaismarkkina-arvoksi on arvioitu noin 38 miljardia euroa, josta Euroopan viiden suurimman sarjan markkina-arvo on yli puolet. Kauden 2023–2024 aikana Real Madrid ylitti ensimmäisenä jalkapalloseuran maailmassa ja toisena urheiluseurana ikinä miljardin liikevaihdon yhden tilikauden aikana (Teitelbaum & Brett, 2025). Pelaajahankintoihin käytetään vuosittain miljardeja ympäri jalkapalloilevaa maailmaa. Fifan mukaan (2025) vuoden 2025 keskipäivän siirtoikkunassa, joka kestää kesäkuun alusta syyskuun toiseen päivään saakka, miesten jalkapallossa käytettiin yhteensä 9,76 miljardia dollaria pelaajasiirtoihin. Tämän on yli 50 % ylitys verrattuna saman ajankohdan siirtoikkunaan vuonna 2024.

Yhä suurempien pääomien siirtely pelaajahankinnoissa on kasvattanut pelaajahankintaprosessin merkitystä. Joukkueet, jotka eivät pysty siirtomarkkinoilla vahvistamaan joukkuettaan vaarantavat seuran niin taloudellisen kuin kilpailullisen menestyksen. Scouteilla eli joukkueiden kykyjenetsijöillä on käytössään ennen näkemätön arsenaali työkaluja löytää joukkueeseen sopiva pelaaja. (Aalbers & Van Haaren, 2019) Kasvava määrä dataa, kehittyneet datan keräysmenetelmät esimerkiksi tekoälytehosteisen tietokonenäkö (computer vision) ja moniulotteisen datan käsittely koneoppimis- ja tilastollisten mallien avulla ovat avanneet pienemmille seuroille mahdollisuuden tehostaa omaa pelaajahankintaa datan avulla ja näin pärjätä kilpailuilla markkinoilla paremmin, mutta samalla esimerkiksi rahakkaat seurat Englannin Valioliigassa palkkaavat kokonaisia datatiimejä tuottamaan analyysejä harjoittelusta, otteluista tai esimerkiksi mahdollisista pelaajahankinnoista (Frost ym., 2025; Fujii, 2025).

Dataa tuotetaan jatkuvasti lisää ja jalkapallon kompleksista olemusta pyritään uusilla tavoilla mallintamaan. Datan keräysmenetelmien kehittyessä tallentuneesta datasta tulee monimuotoisempaa, aineistot saattavat kasvaa niin suuriksi, etteivät perinteiset datan käsittelymenetelmät enää riitä tai tuottavat vähintäänkin haasteita. (Takamido ym., 2025) Lisääntyvä määrä dataa muokkaa pelaajien arviointia, miten valmentajat rakentavat taktiikoita tai miten pelaajahankintaprosesseja toteutetaan (Sharma ym., 2025). Erilaiset koneoppimismallit ovat tuoneet tärkeän elementin suurien data määrien käsittelyyn. Koneoppimiselle hyödynnetään esimerkiksi ennustamaan maalintekijöitä, maaliodottamaa tai tutkimaan yksittäisen pelaajan vaikutusta joukkueen suoriutumiseen. (Cefis & Carpita, 2025; Sharma ym., 2025)

Tutkielmassa puhutaan tekoälystä ja koneoppimismalleista enemmän luvussa kolme. Akateemisessa kirjallisuudessa tekoäly viittaa koneen ja ohjelmiston kykyyn suorittaa ihmismielelle ominaisia toimintoja, kuten oppiminen, ongelman ratkaisu tai päätöksenteko (Cappello ym., 2023; do Rosário ym., 2015; Mason, 2003). Puhekielessä tekoälyllä taas viitataan laajoihin kielimalleihin (large language model, LLM), joita chatbotit, kuten ChatGPT tai Google Gemini edustavat. Koneoppiminen on suosittu tekoälyn muoto (Cappello ym., 2023) ja sitä pidetään yhtenä lähiaikojen merkittävimmistä innovaatioista (Singh & Suguna, 2023; Thantilage ym., 2023). Koneoppiminen on tietokoneohjelman kyky oppia datasta, löytää siitä yhteneväisyyksiä ja korrelaatioita, tunnistaa trendejä ja suorittamaan asetetut tehtävät ilman, että sille erikseen ihmisen toimesta ohjelmoidaan toiminnallisuus tai sääntö. Kun ohjelma tai malli tunnistaa korrelaatioita ja löytää syötteistä eli sille annetusta datasta trendejä, sen kyky ennustaa tai arvioida syötteen ominaisuuksia harjaantuu. (Fujii, 2025; Khanday & Sofi, 2021; Yazici ym., 2023)

Tässä kandidaattitutkielmassa selvitetään vertaisarvioitujen tieteellisten artikkelien pohjalta, millaista dataa pelaajista tulee kerätä edistääkseen jalkapalloseurojen rekrytointipäätöksiä ja miten tekoäly- sekä koneoppimismallit helpottavat datan mallintamista.

Tässä tutkielmassa pyritään vastaamaan seuraaviin tutkimuskysymyksiin:

- Mitä dataa kerätään ja millaisia haasteita liittyy datan laatuun ja saatavuuteen?
- Miten dataa hyödynnetään ja miten se vaikuttaa päätöksentekoon pelaajahankinnassa?

Luvussa kaksi käsittelemme dataa ja sen hyödyntämistä jalkapallon kontekstissa. Luvussa esitellään lyhyesti, miten data on muokannut päätöksentekoa jalkapallossa, mitä dataa pelaajista ja kentän tapahtumista yleisesti kerätään ja millaisia haasteita dataan liittyy. Luvussa kolme tarkastelemme datan käsittelyä tarkemmin. Luvussa käsitellään analyysien laatimista rekrytoinnin näkökulmasta ja miten muun muassa koneoppimismalleja hyödynnetään pelaajien arviointiin. Luvussa katsotaan lisäksi päätöksentekoprosessia pelaajahankinnassa ja miten malleja hyödynnetään pelaajahankintaa tehdessä.

2 Datatyypit ja sen kerääminen pelaajahankinnan tueksi

Data on mullistanut mahdollisuudet ymmärtää jalkapalloa ja siihen liittyviä tapahtumia (Kim ym., 2021). Suuret datavirrat ja niiden tehostunut analysointi tarjoavat päätöksenteolle perusteltavissa olevan pohjan tarvittaessa jopa reaaliajassa (Dubois & Walzak, 2025; Mănescu, 2025). Tässä luvussa käsitellään rekrytointiprosessia ja miten lisääntynyt data ja siitä johdetut tilastot ovat vaikuttaneet rekrytointiin. Luvussa tarkastellaan mistä dataa kerätään ja millä tavoin, vertaillaan eri datakategorioiden ominaisuuksia, pohditaan datan keräämiseen, analysointiin ja ymmärtämiseen liittyviä ongelmia sekä miten dataa voidaan hyödyntää osana jalkapalloseurojen rekrytointipäätöksiä.

2.1 Data – tärkeä osa modernia pelaajahankintaa

Talenttien tunnistaminen ja pelaajahankinta ovat yksiä tärkeimmistä tehtävistä kilpailukykyisen joukkueen rakentamisessa. Pelaajahankintaa pyritään tukemaan data-analytiikan voimin, sillä pelaaja-arvion jättäminen vain muutaman yksilön, yleensä scoutin, oman ”silmätestin” varaan, on lopputulos hyvin altis vinoumille ja usein epäoptimaalinen. (Dubois & Walzak, 2025) Suoriutumista on mitattu yksinkertaisin tilastoin, kuten maalit, maalisyötöt tai nollapelit, mutta nämä kuvaavat heikosti pelaajan todellista arvoa kentällä (Sharma ym., 2025). Pelaajista on kuitenkin olemassa nykyään valtava määrä erilaista dataa. Puettavat liivit ja erilaiset sensorit mittaavat juostua matkaa sekä sijoittumista kentällä ja lukuisat kamerat videoivat joka syötön, vedon tai liukutaklauksen. (Adam, 2022) Datan määrä ja sen valjastaminen päätöksenteon tueksi onkin ohjannut suurimmat jalkapalloseurat palkkaamaan yhden analyytikon sijasta kokonaisia datatiimejä edistämään esimerkiksi data-vetoista pelaajahankintaa (Frost ym., 2025). Englannin pääsarjaa, Valioliigaa, pelaava Liverpool FC on useammassa lähteessä esitetty datavetoisuuden edelläkävijänä. Singh ja Suguna (2023) mainitsevat Liverpoolin nojaavan vahvasti data-analyyseihin tunnistaessaan pelaajien ominaisuuksia ja arvioidessaan hänen sopivuuttansa joukkueen pelityyliin. Wang ja muut (2024) esittelivät artikkelissaan läheisessä yhteistyössä Liverpoolin asiantuntijoiden kanssa kehitetyn TacticAI tekoälymallin, jonka tarkoituksena on optimoida pelaajien liikettä ja sijoittumista kulmapotkutilanteessa.

Monimuotoistuva data ja siitä johdetut mallit auttavat seuroja tekemään rekrytointia hallitummilla riskeillä kuin aikaisemmin. Datan avulla voidaan tarkasti mallintaa pelaajan ominaisuuksia, pelitapaa ja tarkastella hänen sopivuuttansa valmentajan asettamaan pelityyliin. (Singh & Suguna, 2023) Frost ja muut (2025) totesivat tutkimuksessaan, että suorituskykyanalyysin ja pelaajarekrytointin ympäriltä on hyvin vähän kirjallisuutta, vaikka sen vaikutus seurojen toimintaan siirtomarkkinoilla ja lopulta suoriutumiseen kentällä on valtava. Pelaajien arvioinnin ollessa altis scouttien,

valmentajien tai rekrytoinnin ammattilaisten mielen vinoumille ja omille mieltymyksille (Larkin & O'Connor, 2017) Frost ja muut (2025) huomasivat tutkimuksessaan, että pelaajahankintaan erikoistuneet analytiikat kokemuspohjasta riippumatta hyödynsivät data pohjaisia suoriutumisen mittareita ja esimerkiksi videodataa tehdessään pelaaja-arvioita. Historiallisesti datan hyödyntäminen on tullut jalkapalloon suhteellisen myöhään verrattuna esimerkiksi baseballiin, sillä jalkapallon tapahtumia on vaikea selittää tilastollisesti (Kim ym., 2021). Koneoppiminen on ollut tässä suhteessa yksi suurimmista innovaatioista. Erilaiset metriikat pelaajien suoriutumisen mittaamisessa on avannut mahdollisuuksia hyödyntää koneoppimismalleja liiketoiminnan tärkeissä päätöksissä. Koneoppimismallit tai tilastometriikat eivät poista jalkapallosta kontekstiosaamista, mutta toimivat työkaluina päätöksenteon tukena. (Singh & Suguna, 2023)

2.2 Datan eri muodot ja käyttötarkoitukset

Akateemisessa kirjallisuudessa kentäntapahtumien tilastoinnin osalta jako tapahtuu yleensä kahteen kategoriaan, tapahtumadataan (event data) ja sijainti- tai seurantadataan (location data, tracking data) (Anzer & Bauer, 2021; Decroos ym., 2019). Tapahtumadataksi luokitellaan erilaisia kentän konkreettisia tapahtumia, kuten syöttö, laukaisu, virhe, varoitus tai ulosajo. Tapahtumadataan jää aika- ja paikkaleima siitä missä ja milloin tapahtuma on ilmennyt. (Decroos ym., 2019; Fujii, 2025) Tapahtumadataa kerätään suurimmissa sarjoissa manuaalisesti koulutetun henkilöstön toimesta (Anzer ym., 2021; Fujii, 2025). Tapahtumadata on kriittinen työkalu, kun tutkitaan yksilön suoriutumista. Yhdistettynä esimerkiksi sijaintidatan kanssa seurojen rekrytoijat kykenevät kokoamaan kokonaisvaltaisen aineiston pelaajan suorituksesta ja tehdä datalla perusteltavissa olevan päätöksen uuden pelaajahankinnan suhteen. (Fujii, 2025) Tapahtumadata saattaa yksinään olla hyvin ongelmallista ja harhaanjohtavaa (Singh & Suguna, 2023). Vaikka tapahtumadataa on laajasti tarjolla pienillä kustannuksilla (Decroos ym., 2019), kattavin data on kuitenkin saatavilla vain suurimmissa sarjoissa pelaaville joukkueille (Yeung ym., 2025).

Sijaintidata seuraa pelaajien ja pallon liikettä ja pyrkii mallintamaan sitä (Decroos ym., 2019; Fujii, 2025). Sijaintidatan keräämiseen on käytössä useita teknologioita ja niistä yleisimmät ovat optiset seurantajärjestelmät, globaalit ja lokaalit sijaintijärjestelmät (GPS ja LPS) sekä tietokonenäkö (Kovalchik, 2023). Sijaintidataa saatetaan käyttää moneen tarkoitukseen. On määritettävä mitä halutaan tutkia ja tämän perusteella valitaan käyttötarkoitukseen optimaalinen vaihtoehto. (Torres-Ronda ym., 2022) Harper ja muut (2019) hyödynsivät GPS ja LPS teknologiaa ja tutkivat meta-analyysissään eri urheilulajien vaadetta korkean ja hyvin korkean intensiteetin kiihdytyksiin ja pysähdyksiin ottelun aikana. Siinä missä tapahtumadata pyrkii keskittymään ensisijaisesti siihen, mitä

tapahtuu pelivälineen ympärillä tai sen kanssa sijaintidatan avulla voidaan mallintaa, miten muut pelaajat liikkuvat kentällä tai reagoivat kentän tapahtumiin. Niin kuin Yeungin ja muiden (2025) artikkelista käy ilmi on sijaintidatan tuottaminen hyvin kallista ja monesti vain rahakkaimmilla seuroilla ja liigoilla on siihen varaa. Modernit optiset kamerajärjestelmät mallintavat pelaajien liikettä kentällä 2D ulottuvuudessa ja pallon liikettä 3D ulottuvuudessa noin 25 hertsin taajuudella, eli sensorit luovat datapisteen 25 kertaa sekunnissa (Kovalchik, 2023). Kehittynyt tietokonenäkö taas mahdollistaa sijaintidatan tallentamisen ottelulähetyksistä, jopa reaaliaikaisesti. Aikaisemmin pelaajien liikkeiden tallentaminen ja analysointi suoritettiin kuva kuvalta ihmisen toimesta, joka kulutti kohtuuttomasti aikaa ja resursseja. (Fujii, 2025) Optisia kamerajärjestelmiä pidetään toistaiseksi kuitenkin tarkimpana (Kovalchik, 2023) ja esimerkiksi Saksan jalkapallon pääsarjassa Bundesliigassa on käytössä optiset kamerajärjestelmät, jotka mallintavat jokaisen pelaajan liikettä useasta eri kamerakulmasta (Anzer & Bauer, 2021).

Kirjallisuudessa myös videodata tai videoanalyysi yleensä esiintyvät omana datamuotonaan (Frost ym., 2025; Fujii, 2025), mutta esimerkiksi sijaintidataa tuotetaan tietokonenäön avulla videokuvasta (Olthof & Davis, 2025). Suorituskyvyn mittareita ja laaja kattaus erilaista videodataa on saatavilla eri palveluntarjoajilta. Videokuva on kustannustehokas ja paikanpäälle menemiseen verrattuna marginaalisesti aikaa vievä tapa analysoida joukkueen tai yksilön pelaamista. (Frost ym., 2025) Videokuvaa hyödyntävät monipuolisesti niin analyttikot, pelaajat itse kuin valmentajatkin tukeakseen datan pohjalta tehtyjä havaintoja tai arvioissa omaa suoriutumista kentällä (Fujii, 2025).

Urheiludataa on kategorisesti paljon muutakin riippuen siitä, mitä halutaan tutkia. Seuroilla saattaa olla käytössään dataa pelaajien loukkaantumishistoriasta tai fyysisen suorituskyvyn mittareista, joiden avulla voidaan arvioida yksilön loukkaantumisriskiä ja edesauttaa pelaajan ehjänä pysymistä (Singh & Suguna, 2023). Esimerkiksi sykemittarit seuraavat pelaajan sydämen toimintoja antaen reaaliaikaista kuvaa yksilön fyysisestä suorituskyvystä, kestävydestä ja palautumisesta (Fujii, 2025). Mănescu (2025) tutki millä tavoin voidaan optimoida treenimäärät, mutta samalla välttyä takareisivammoilta ammattilaisjalkapallossa. Hän keräsi puettavaa sensoridataa pelaajan suorittamista kovavauhtisista juoksuista pelin aikana samalla tarkkaillen pelaajan lihasten happikylläisyyttä, videokuvasta sijaintidataa tukemaan GPS laitteiden tuottamaa dataa ja lisäksi pelaajilta itseltään kerättiin raportteja omasta voinnistaan.

2.3 Datat saatuuden ja laadun asettamat haasteet

Perinteiset tilastot ovat tarjonneet historiassa katsauksen kentän tapahtumiin, mutta enenevässä määrin on selvää, että kehittyvä laji ja monimuotoiset strategiat tarvitsevat mallintamiseen

edistyneempiä tilastoja (Elsharkawi ym., 2025). Pelkästään pelaajarekrytoinnin tai -arvioinnin näkökulmasta suoriutumisen mallintaminen ei ole yksinkertaista. Pelaajaa voidaan arvioida monella eri tasolla ja jalkapallossa nämä ominaisuudet, kuten nopeus, ketteryys, peliäly, tekninen osaaminen tai psyykkinen ja henkinen kapasiteetti, vaikuttavat yksittäisen pelaajan esiintymiseen kentällä. (Larkin & O'Connor, 2017) Pelitavalla on valtava merkitys siihen, miten pelaajia arvioidaan datan pohjalta. Alempien sarjojen tai junioriliigojen tilastoissa tapahtumadataksi tallentuvat teot korostuvat, koska sijaintidataa on vähän tai ei ollenkaan ja sen yhdistäminen tapahtumadatan kanssa kokonaisvaltaisen analyysin koostamiseksi on mahdotonta, jolloin pelitavallisesti niiden joukkueiden pelaajat erotuvat edukseen, jotka hallitsevat palloa ja luovat maalipaikkoja (Larkin & O'Connor, 2017; Yeung ym., 2025). Esimerkiksi EFL Championshipin eli Englannin toiseksi korkeimman sarjataso Derby Countyn rekrytointipäällikkö on painottanut tilastojen kontekstin ymmärtämistä. Hän on sitä mieltä, että pelitavalla on merkitystä siihen, millaisia tilastoja pelaajista syntyy. Hyvin puolustavalla pelityylillä operoivan joukkueen puolustajilla saattaa olla merkittävästi paremmat puolustuspään tilastot, kuten syötön katkot tai pääpelivoitot, kuin pallonhallintaan keskittyvällä joukkueella. Datasta saattaakin korostua tietyt tilastot pelaajien toiminnasta kertomatta kuitenkaan, kuinka laadukkaasti pelaaja tai joukkue on suoriutunut. (Singh & Suguna, 2023) Tässä korostuu taito ymmärtää, millaisessa kontekstissa data on kerätty ja toisaalta nostaa katsetta yhdestä tilastokategoriasta suuremman kuvan tarkasteluun.

Datan kerääminen tai sen saatavuus saattavat ajoittain tuottaa haasteita. Decroos ja muut (2019) toteavat artikkelissaan yhden merkittävän haasteen olevan tapahtumadatan painottumisen maalinteon ympärille tilastoina, kuten tehdyt maalit, vedot ja maalisyötöt. Anzer ja Bauer (2021) mainitsevat artikkelissa muun muassa jääkiekon ja koripallon, jotka ovat jalkapalloon verrattuna runsasmaalisia pelejä ja pohtivat jalkapallon kontekstissa, kuinka haastavaa on pelkän tapahtumadatan pohjalta luoda kattava aineisto maalinteon analyttiselle tutkimiselle. Alves (2025) totesi tutkimuksessaan, että välittömästi maaliin tai maalipaikkaan johtavat tapahtumat eivät tuota riittävän kattavaa aineistoa jalkapallo-ottelun seuraavan tapahtuman ennustamiseksi. Hän päätyi noteeraamaan tapahtumia, kuten kaksinkamppailut tai pallonriistot, koska niitä tapahtuu jalkapallo-ottelussa huomattavasti useammin ja niillä on merkittävä vaikutus siihen, miten tapahtumat kentällä etenevät. Alves (2025) kuitenkin toteaa niiden havainnoinnin olevan haastavaa sijaintidatan niukan saatavuuden takia. Niukkuus onkin yksi keskeinen haaste sijaintidatan hyödyntämiselle laajemmalla mittakaavalla. Laadukasta sijaintidataa on hyvin vähän saatavilla, sillä sen tallentaminen on äärimmäisen pääomaintensiivistä ja sen takia lähtökohtaisesti käytettävissä vain suurimpien sarjojen seuroilla. Tämä hankaloittaa analyttikoiden ja tutkijoiden työtä, jotka toimivat alempien sarjatasojen tai

juniorituotannon parissa (Yeung ym., 2025). Decroos ja muut (2019) toteavat tutkimuksessaan käyttävänsä ainoastaan tapahtumadataa tutkiessaan pelaajien kokonaisvaltaista suoriutumista kentällä, sillä käyttöoikeuksia sijaintidataan ei saatu. He totesivat kuitenkin mallinsa olevan täydennyskelpoinen sijaintidatalla, mikäli sellaista on saatavilla.

Käytettävissä oleva data ei aina sovellu täysin analyysien tuottamiseen joukkueen tai yksittäisen pelaajan suoriutumisesta. Tapahtumadataa tuotetaan usein moneen eri käyttötarkoitukseen, kuten raportoitavaksi tv-katsojille tai lehdistölle, eikä tästä syystä automaattisesti sovellu parhaalla mahdollisella tavalla analyysien laatimiseen (Decroos ym., 2019). Dataa kohokohtatapahtumista on laajalti tarjolla monesta eri lähteestä, sillä valtaosaa kuluttajista kiinnostaa enemmän montako vetoa heidän lempijoukkueellaan on kohti maalia kuin paljonko laitapakin yksittäinen syöttö on vaikuttanut pelin positiiviseen virtaukseen (Alves, 2025). Ottelun kohokohtia on huomattavasti helpompi mitata ja esittää katsojille niihin liittyviä lukuja reaaliajassa (Alves, 2025), mutta yksittäinen katsojalle kiinnostava luku, esimerkiksi pelaajan huippunopeus pelin tiimellyksessä, ei kerro onko se edesauttanut joukkueen pelaamista tuottamalla fiksun syötön eteenpäin tai voittamaan kaksinkamppailun (Takamido ym., 2025). Tämä saattaa kuitenkin olla merkittävä havainto ottelua ja pelaajaa seuraavalle scoutille. Koska palveluntarjoajat pyrkivät tuottamaan dataa erilaisiin käyttötarkoituksiin, kuten tutkimukseen tai suorituskyvyn analysointiin, käyttävät he tuottamassaan datassa toisistaan eroavaa terminologiaa, joka vaikeuttaa datan yhdenmukaistamista eri lähteiden ja palvelun tarjoajien välillä (Yeung ym., 2025). Yritykset, jotka ovat tuottaneet dataa eri tarkoituksiin vuosikymmenten ajan ovat ikään kuin jumissa omissa tyylivalinnoissaan datan osalta. Tekemällä suuria muutoksia datan monimuotoisuuteen saataisi dataa menettää vertailtavuutensa. Toisaalta palvelun tarjoajat, jotka ovat olleet markkinoilla vähemmän aikaa tarjoavat tiettyihin käyttötarkoituksiin turhankin monimuotoista dataa ja vaikeuttavat kyselyjen tekemistä tietokantaan. Yksinkertaisella tapahtumalla, kuten veto voi olla useista kategorioista riippuen mikä on sen lopputulos. (Decroos ym., 2019)

3 Datan käsittely ja hyödyntäminen pelaajahankinnassa

Aikaisemmassa luvussa käsiteltiin, millaista dataa on saatavilla, mistä sitä kerätään, millä metodeilla ja millaisia haasteita dataan liittyy. Tässä luvussa pääsemme tarkastelemaan, miten dataa todella hyödynnetään. Pääsemme tutustumaan pelaajahankintaa tukeviin analyysihin, miltä näyttää pelaajahankintaprosessi, miten data on siihen vaikuttanut ja miten analyysijä hyödynnetään pelaajahankinnassa.

3.1 Pelaajahankintaa tukevat analyysit

Kerättyä data pitää pystyä analysoimaan ja sitä voidaan hyödyntää pelaajahankinnan kontekstissa moneen osa-alueeseen. Jotta pelaajahankinta on kestäväällä pohjalla taloudellisesti ja sopivuus urheilullisesti olisi taattu, on mahdollista tulevaa hankintaa analysoida kokonaisvaltaisesti. Kentän tason suoritukset, loukkaantumisalttius tai vain sopivuus joukkueen pelityyliin pitää olla tarkoin analysoitu ennen kuin voidaan perustella taloudellinen investointi uuteen pelaajaan.

Mittauskohde	Menetelmä	Hyöty pelaajahankintaan	Lähteet
Maaliiodottaman mittaaminen	Maaliiodottama (expeted goal, xG)	Maalinteon tehokkuuden arviointi, suoriutumisen arviointi, yksittäisen pelaajan kyky tuottaa maalipaikkoja ja maaliiodottamaa.	(Anzer & Bauer, 2021; Cefis & Carpita, 2025; Robberechts & Davis, 2020)
Pelaajan kokonaisvaltaisen suoriutumisen arviointi	VAEP (Valuing Actions by Estimating Probabilities), PIM (player impact metric), PA (performance analysis), videoanalyysi	Pelaajien kokonaisvaltainen arviointi huomioiden myös defensiiviset teot ja tapahtumat.	(Decroos ym., 2019; Elsharkawi ym., 2025; Frost ym., 2025)
Markkina-arvo analyysi	Arvion joukkoistaminen ja XGBoost.	Arviot pelaajan hinnasta antavat suunnan missä pelaajan markkina-arvo suurin piirtein liikkuu tai millaista arvoa pelaajan voidaan olettaa seuraan tuovan.	(Coates & Parshakov, 2022; McHale & Holmes, 2023)

Taktinen analyysi	Tekoäly ja TacticAI, koneoppiminen potentiaalisten tilanteiden tunnistamisessa, VAEP ja VDEP.	Pelitalanteiden analysointi, pelaajien käyttäytyminen ja kyvyt, joukkueen optimaalisen rakenteen analysointi ja mallintaminen	(Fujii, 2025; Toda ym., 2022; Wang ym., 2024)
Pelaajatyypin tunnistaminen	Ryvästämisen (pelitavataan saman kaltaisten pelaajien löytäminen ja vertailu)	Pelaajatyypien tunnistaminen, samankaltaisten pelaajien tunnistaminen ja siitä johdettu hankintahintojen vertailu (aliarvostetut pelaajat)	(Aalbers & Van Haaren, 2019; Seri ym., 2025; Vidal ym., 2025)

Taulukko 1 Analyysin kohteet ja menetelmäkuvaus

3.1.1 Suoriutumisen analysointi ja siihen käytettävät mittarit

Michael Lewisin (2004) kirja *Moneyball* on yksi vaikutusvaltaisimmista teoksista, kun puhutaan numeraali datan hyödyntämisestä pelaajahankinnassa ja sitä pidetään teoksena, joka popularisoi data-vetoisen pelaajahankinnan joukkueurheilussa (Frost ym., 2025). Pelaajien suoritteiden arvioiminen datan pohjalta on mennyt harppauksin eteenpäin koneoppimisen ja tilastollisten mallien myötä (Elsharkawi ym., 2025). Sen avannut ovia pienemmille seuroille tunnistaa aliarvostettuja pelaajia, mutta toisaalta suurimmat seurat ovat voineet palkata kokonaisia datatiimejä rakentamaan analyysijä erilaisiin tarkoituksiin esimerkiksi pelaajahankintaan (Frost ym., 2025).

Vaikka maali on jalkapallossa harvinainen tapahtuma, on se silti yksi tärkeimmistä suoriutumisen mittareista niin joukkue kuin yksilö tasolla. Kuitenkin pelkän laukauksen onnistumisen mittaaminen diskreetillä arvolla maali tai ei maalia on liian pelkistävä arvio, kun halutaan tutkia maalintekoyrityksen todellista potentiaalia. (Anzer & Bauer, 2021) Myöskään vain laukausmäärien tarkastelu ei anna laadukasta kuvaa joukkueen tai yksittäisen pelaajan suorituksista, sillä laukausmäärät tai yksittäiset kirjaukset eivät mallinna maalintekoyrityksen laatua. Koneoppimisen avulla on luotu mittari, maaliodottama (expected goal, xG), joka arvioi laukauksen todennäköisyyttä päätyä tulostaululle. Malli koulutetaan suurella historiallisella datalla, joka ottaa huomioon muun muassa laukaisupaikan ja etäisyyden maaliin sekä millä kehonosalla veto tapahtuu ja antaa yritykselle arvon yhden ja nol-lan väliltä sen perustella, kuinka todennäköisesti maalintekoyritys päättyy maaliin. (Robberechts & Davis, 2020)

Anzer & Bauer (2021) halusivat tutkimuksessaan laajentaa maali-dottamaa ottamaan huomioon pelkän laukauksen sijaan myös muiden pelaajien positiot suhteessa siihen, mistä veto lähtee. Testattuaan useita koneoppimis algoritmeja he totesivat XGBoost (extreme gradient boosting) mallin olevan tehokkain ja tuottavan tarkemman ennusteen kuin aikaisemmin esitetyt maali-dottamamallit. He perustelivat tutkimuksessaan, että useilla tekijöillä oli merkitystä maali-dottamalukeman muodostumiseen, joista merkittävin oli matka maaliin. Tekijöitä, joita vain tapahtumadataksi ei tallennu ovat esimerkiksi muiden pelaajien aiheuttama paine, maalivahdin sijoittuminen ja etäisyys maalista sekä pelaajan liikkeen nopeus laukaisuhetkellä. Näitä ei esimerkiksi ollut aikaisemmin akateemisessa kirjallisuudessa esiintyneissä malleissa otettu huomioon. Cefis & Carpita (2025) halusivat tuoda malliin lisäelementin ja ajattelivat, että sillä kuka palloa potkaisee tai kuka on maalissa, merkitsee. Mallin kouluttamiseen käytettiin logistista regressiomallia ja otoksen tasapainotusmenetelmää (sample-balanced technique, SBT). Tutkimukseen valittiin tämä tilastollinen malli, sillä haluttiin tietää, miten uudet tilastolliset muuttujat vaikuttavat maali-dottama-arvion kehittymiseen. Koska tutkimuksen otoksessa itse tutkittava kohde eli maalit olivat vain pieni osa otoksesta, oli käytettävä otoksen tasapainotusta. Yksinkertaistettuna tasapainotusmenetelmää hyödyntäen otoksesta poistetaan yksittäisiä havaintoja useimmin esiintyvistä luokasta eli laukauksesta, joka ei pääty maaliin ja harvimminkin esiintyvän luokan dataa monistetaan, jotta mallin kouluttamiseen käytettävä data olisi tasapainoisempi ja mallin tulosteet luotettavampia. Tutkimuksessa huomattiin muun muassa laukaisukulman, pelaajan liikkeen ja maalivahdin taitojen olevan merkittäviä tilastollisia muuttujia. Pelaajan suoriutumista arvioivissa metriikoissa usein suositaan offensiivisiä tilastomuuttujia. Esimerkiksi Decroos ja muut (2019) pyrkivät rakentamaan pelaaja-arvioinnin viitekehyksen, joka arvottaa tapahtumia niiden kontekstissa esimerkiksi syöttö paineen alla ei ole sama kuin paineeton syöttö puolustajalta puolustajalle ja toisaalta arvottaa tapahtumaa sen pidemmän aikavälin vaikutuksen perusteella. Heidän mallissaan tunnistettavia haittoja taas olivat, ettei malli ota huomioon puolustuspelillisiä аспекteja. Lisäksi heikompi-tasoisissa sarjoissa positiivisia suoritteita oli helpompi tehdä kuin korkean intensiteetin sarjoissa. Elsharkawi ja muut (2025) esittelivät artikkelissaan pelaajan vaikutusmittari (player impact metric, PIM) viitekehyksen, jonka tarkoitus on huomioida offensiiviset sekä defensiiviset tapahtumat pelaajan suoriutumista arvioidessa. Suorituksia painotettiin sen perusteella, miten ne vaikuttavat ottelun lopputulokseen. Mallin rakentamisessa hyödynnettiin ordinaalista logistista regressiota, sillä se tarjosi selitettävät tulosteet ja mahdollisti usean ennustavan muuttujan käsittelyn. Mallin tuloksia verrattiin WhoScored palvelun pelaaja-arvioon, joka antaa pelaajalle arvon hänen suoriutumisensa perusteella. Mallin tuottamat arvot olivat pitkälti linjassa WhoScored palvelun arvojen kanssa, mutta suurimmat erot tulivat pelaajien kohdalla, jotka

suoriutuivat vahvasti defensiivisesti. Toisaalta tutkimuksessa hyödynnetty malli ei antanut yhtä suurta painoarvoa maalintekemiselle kuin WhoScored palvelun malli. Pelaaja ei siis saanut yhtä helposti hyvää pelaajan vaikutusmittarin arvoa, mikäli hänen ainoat positiiviset kontribuutiot olivat maali tai siihen johtanut syöttö.

3.1.2 Markkina-arvoanalyysi

Ammattilaisurheilu ja pelaajamarkkina on äärimmäisen harvinainen pelikenttä. Se on maailman ainoa ala, jossa palkattavasta henkilöstöstä eli pelaajista tiedetään aivan kaikki fyysisistä ominaisuuksista numeraalisiin todisteisiin työsuoritteesta. Jos voimme olettaa pelaajamarkkinan toimivan tehokkaasti, siirtosummat edustavat arvoa, jonka pelaaja pystyy seuralle tuottamaan. (McHale & Holmes, 2023)

Koska kaikkia pelaajia ei siirrellä seurasta toiseen joka kausi, pelaajilla ei ole jatkuvaa joka kausi realisoituvaa markkina-arvoa. Olemassa on kuitenkin palveluita, jotka tarjoavat estimaatin pelaajan markkina-arvosta. (Coates & Parshakov, 2022) Esimerkiksi suosittu sivusto transfermarkt.com tarjoaa joukkoistetun (crowd-sourced) laskelman ammattilaisjalkapalloilijoiden markkina-arvoista. Sivuston jäsenet antavat oman arvionsa pelaajan arvosta ja niitä painotetaan sen perusteella, kuinka hyvin jäsen on pystynyt arvioimaan pelaajien todellisia siirtosummia menneisyydessä. (McHale & Holmes, 2023) Vaikka Transfermarktin markkina-arvo johtaja Christian Schwarz sanoi *The Guardianin* haastattelussa (Aarons, 2020), etteivät heidän menetelmänsä ole tieteellisiä eikä jalkapalloseurojen pitäisi tehdä pelaajahankintaa perustuen heidän antamiinsa arvioihin, eri tahot alalla, kuten seurat ja agentit, käyttävät tätä arvioita mittarina pelaajan arvosta markkinoilla (Coates & Parshakov, 2022). Vaikka itse pelaajan arvoa ei sidottaisikaan tämän kaltaisen palvelun antamaan estimaattiin, todettakoon kuitenkin, että arvoit pelaajat markkina-arvosta toimivat hyvinä mittareina siirron onnistumisesta (Singh & Suguna, 2023). Mikäli pelaajan markkina-arvo nousee siirron jälkeen, voidaan todeta pelaajan todennäköisesti suoriutuneen hyvin ja jos taas pelaaja on suoriutunut odotuksiin nähden huonosti, markkina-arvo tai siitä tehty arvio todennäköisesti laskee.

McHale ja Holmes (2023) totesivat artikkelissaan, että akateemisessa kirjallisuudessa on hyvin vähän, jos ollenkaan tietoa mikä on luotettava mittari pelaajansuoriutumisesta kentällä ja toisaalta mitkä tekijät vaikuttavat pelaajan markkina-arvoon. McHale ja Holmes hyödynsivät tutkimuksessaan niin joukkoistettua dataa kuin yksilöllistä pelaajadataa. Tutkimuksessa testattiin usean koneoppimismallin suoriutumista ennustettaessa pelaajien markkinahintoja. Lopuksi tuloksia verrattiin Transfermarktin dataan. Keskimääräisen absoluuttisen virhearvon (Mean Absolute Error, MAE) tulokset olivat Transfermarktin palvelun arvojen kanssa samankaltaisia, mutta mitattaessa

keskimääräistä suhteellista virhettä (Mean Absolute Percentage Error, MAPE) McHalen ja Holmesin malli suoriutui huomattavasti paremmin. Mallin voidaan olettaa olevan seuroille käyttökelpoinen työkalu tehdessä pelaajahankintoja ja heidän markkina-arvoansa määritettäessä. Silti on todettava, että mallien antamat tulokset ovat vain ennusteita tai arvioita pelaajien arvoista ja seuran tulee itse määrittellä paljonko sillä on maksukykyä ja mikä on sen riskinsietokyky mahdollisen ylimaksamisen suhteen.

Vaikka pelaajahankintaa ja muuta päätöksentekoa pyritään tehostamaan datan avulla, on hankintapäätökset edelleen seurajohdon intuition varassa. Dataa pyritään monessa päätöksenteon kontekstissa hyödyntämään, mutta pelaajia edelleen satunnaisesti palkataan ”nimen perusteella.” Lähivuosilta esimerkiksi espanjalainen keskushyökkääjä Alvaro Morata on tästä hyvä esimerkki. Hänen suoriutumisen on ollut vähintäänkin keskinkertaista, mutta suuret siirtosummat ovat antaneet ymmärtää toista. Pelaajien arvoa ei voida toisaalta aina mitata vain suoriutumisen perusteella. Pelaajan arvoon vaikuttavat esimerkiksi hänen tunnettavuutensa ja muiden joukkueiden kiinnostus samaa pelaaja kohtaan. (Kim ym., 2021) Pelaajahankinta ja heidän arvonsa määrittäminen ovat kompleksisia kokonaisuuksia. Pelaajan arvoon ei vaikuta yksinään vain suoriutuminen tai edes satunnaisen joukon antama arvio. Talouden kasvuvauhti, lait ja määräykset sekä muut tekijät vaikuttavat edellä mainittujen seikkojen lisäksi siirtomarkkinoilla liikkuviin rahasummiin. (Singh & Suguna, 2023)

3.1.3 Pelaajien taktinen roolitus ja pelityylianalyysi

Videoanalyysi on kasvattanut suosiotaan taktisen analyysin työkaluna ja onkin käytetyin työkalu tässä kontekstissa. Sen huonoja puolia on, että videoanalyysi on hyvin aikaa vievää ja inhimillisille erehdyksille altis. Edistyneet datan käsittelymenetelmät, kuten koneoppiminen ovat avanneet mahdollisuuden automatisoida prosessia. Esimerkiksi maaliiodottama, pelaajien liikkeiden mallintaminen tai erilaiset suoriutumisen mittarit ovat tuoneet uuden ulottuvuuden taktisen analyysin laatimiselle. (Olthof & Davis, 2025) Pelin tapahtumien mallintaminen ja datan seasta toistuvien rakenteiden löytäminen on hyvin haastavaa. Pelikatkot ja erikoistilanteet tarjoavat analyytikoille hallitun ja toistuvamman ympäristön mallintaa pelaajien liikettä. Kulmapotkut tai pitkät rajaheitot ovat pelissä huomattavasti toisteisempia kuin maalit, joten näiden tilanteiden analysointi ja pelaajien liikkeen tai toistuvien rakenteiden tutkiminen on helpompaa. Onnistuneella erikoistilanteella saattaa lisäksi olla merkittävä vaikutus pelin lopputuloksen kannalta. (Wang ym., 2024)

Aikaisimmin luvussa 2.1 oli maininta Wangin ja muiden (2024) artikkelista, jossa he loivat TacticAI tekoälyavusteisen mallin, jonka tarkoitus on tukea erikoistilanteen specialistia kulmapotkutilanteiden analysoinnissa ja kulmapotku rutiinien parantamisessa. Wang ja muiden (2024) mallin

kouluttaminen tapahtuu esikäsittelemällä luokiteltu sijaintidata graafiseen muotoon. Käsitelty data syötetään mallille, jonka tehtävä on prosessoida syötetty data luokittelu- tai regressio-ongelmana. Wang ja muut totesivat tutkimuksessaan TacticAI mallin pystyvän tasaisesti ennustamaan pelaajan, joka ensimmäisenä kulmapotkun jälkeen koskisi palloon, sen avulla pystyttiin uskottavasti luomaan kulmapotkuskenaarioita ja testaamaan erilaisia taktiikoita sekä malli pystyi tunnistamaan samankaltaisia ja toisteisia taktiikoita historiallisesta datasta.

Maaliiodottama on lähtökohtaisesti käytössä suoriutumisen mittarina, mutta lähempää tarkasteltuna käyttökelpoinen myös taktisen päätöksenteon tukena. Anzer ja Bauer (2021) tutkivat maaliiodottamaa eri Bundesliiga joukkueiden tiimoilta ja millaisissa tilanteissa joukkueet luovat maaliiodottamaa. Tarkastelemalla sijaintidatan avulla maaliiodottama suhteita eri pelitilanteissa tai millaisista kohdista kenttää laukaukset tyypillisesti lähtevät voidaan tehdä taktisia valintoja niin hyökkäys- kuin puolustus-suuntaan. Olthof ja Davis (2025) toteavat, että maaliiodottamaa tarkastelemalla voidaan kannustaa pelaajia laukomaan paikoista, joista he eivät välttämättä laukoisi tavallisesti. Cefis ja Carpita (2025) esimerkiksi totesivat, että laukaisuetäisyydellä ja maalivahdin sijoittumisella on merkitystä maalin syntymisen todennäköisyydelle. Pelaajia voidaan valmentaa tunnistamaan tilanteita, joissa puolustava joukkue on heikosti sijoittunut tai maalivahti toistuvasti sijoittuu maalinteon suhteen edullisesti.

Fujii (2025) toteaa, että eri tilanteiden johtamista maaliin on vaikea arvioida ja esittelee Todan ja muiden (2022) kehittämän viitekehysten, joka arvioi hyökkäyspelaamista puolustavan joukkueen pelaajien tekojen ja sijoittumisen kautta. Malli arvioi puolustavan joukkueen toimintaa ja sijoittumista suhteessa hyökkääjiin ja arvioi heidän vaikutustaan syntyvään maalipaikkaan ja hyökkäyksen todennäköisyyteen päätyä maaliin (Valuing Defence by Estimating Probabilities, VDEP). Malli ottaa huomioon enemmän vaaralliset hyökkäykset ja arvio ne epäonnistuneena puolustuspelinä, kuin hyökkäykset, jotka päättyvät laukaukseen (Toda ym., 2022). Malli ottaa huomioon siis myös hyökkäykset, jotka eivät pääty laukaukseen, mutta lasketaan vaarallisiksi maalintekoyrityksiksi. Malli on muunnos samankaltaisesta mallista, joka arvioi pelaajia offensiivisten tekojen todennäköisyyden perusteella päätyä maaliksi (Valuing Actions by Estimating Probabilities, VAEP) (Decroos ym., 2019). Toda ja muut (2022) huomasivat VDEP mallia testatessaan, että sillä on kohtalainen positiivinen korrelaatio tienattujen sarjapisteiden kanssa, mutta alkuperäisellä VAEP mallin arvolla oli vahva positiivinen korrelaatio. VAEP mallilla taas ei ollut minkäänlaista korrelaatiota päästettyjen maalien kanssa, koska se mittaa vain offensiivisia arvoja. Toda ja muut (2022) totesivatkin, että VDEP malli voi olla hyvä työkalu joukkueen kokonaisvaltaisen suoriutumisen seuraamisessa ottaen huomioon puolustuspelilliset kyvykkyydet.

Vidal ja muut (2025) halusivat tutkimuksessaan tunnistaa samankaltaisia pelaajia valvotulla ryväs-tämismallilla. Tarkoituksena on löytää pelitavallisesti samankaltaisia pelaajia, joka voi hyödyttää joukkueita suunnittelemaan kokoonpanoja ja hakea taktisia variaatioita omaan pelaamiseensa. Viite-kehystä voisi mahdollisesti hyödyntää löytämään pelillisiä synergioita tai tunnistaa yhteensopivuuk-sia joukkueen muiden pelaajaprofiilien välillä. Aalbers ja Van Haaren (2019) pyrkivät samalla ta-valla tunnistamaan erilaisia pelaajia heidän pelityylinsä perusteella. Malli jakoi aineiston pelaajat 21 erilaiseen rooliin pelitavan perusteella. Aalbers ja Van Haaren arvioivat artikkelissaan esimerkkinä keskikenttäpelaajien roolituksia. He arvioivat keskikenttäpelaajien sopivuutta viiteen erilaiseen roo-liin ja loivat mittarin yksittäisen pelaajan suoriutumisesta ennalta asetetussa roolissa tai sopivuudesta tietyn roolin asettamiin vaatimuksiin pelitavallisesti. Seri ja muut (2025) vaihtoehtoisesti an-toivat mallille etukäteen kiinteän luvun, moneenko rypäaseen mallin tulee jakaa datan sisältämät pe-laajat. Heidän käyttämä data sisälsi monen eri pelipaikan pelaajia sisältäen maalivahdit. Yksi heidän mainitsemistansa kehityskohdista onkin hyödyntää mallia tunnistamaan eroavia pelityylejä saman paikan pelaajissa. Näin voidaan edistää yksityiskohtaisemman analyysin tekemistä pelaajien omi-naisuuksista ja auttaa päätöksentekoa niin pelaajahankinnassa kuin taktisessa suunnittelussa.

3.2 Data-analyysien käyttö pelaajahankinnan tukena

Data-analyysien perimmäinen tarkoitus on auttaa scoutteja ja seurajohtoa tekemään perustellumpia päätöksiä ja välttämään tärkeiden, urheilullista ja taloudellista menestymistä määrittävien päätösten perustamista vain intuition varaan. Tarkoitus ei kuitenkaan ole hylätä scouttien ja muiden alan am-mattilaisten kartuttamaa kokemusta tai vahvaa kontekstiosaamista vaan tuoda siihen elementtejä, joita ei ole aikaisemmin osattu ottaa huomioon. (Dubois & Walzak, 2025) Pelaajahankintaa teh-dessä dataksi voidaan mallintaa pelaajan suoriutumista historiallisesti ja tarkastella hänen kehitys-tään, analysoida miten pelaaja sopii joukkueeseen taktisesta näkökulmasta ja ennustaa miten han-kinnan markkina-arvo tulee mahdollisesti kehittymään tulevina vuosina korostuen erityisesti nuo-riassa talenteissa (Davis ym., 2024).

Käytössä oleva data, siitä johduttu suoriutumisen mittarit ja tehokkaat koneoppimismallit voivat oi-kein käytettyinä auttaa jalkapalloseuroja säästämään miljoonia ja tuomaan teknisemmän näkökul-man urheiluun ja sen kompleksisen maailman ymmärtämiseen. Jalkapalloseurat voivat tehdä fik-sumpia investointeja laskelmoidummilla riskeillä. (Singh & Suguna, 2023) Frost ja muut (2025) tutkivat, miten jalkapalloseurojen rekrytoinnin ammattilaiset hyödyntävät suoritus analyysiä (per-formance analysis, PA) ja videokuvaa muodostaessaan rekrytointiraporttia. He vertailivat kokemat-tomimpien ja kokeneempien rekrytoijien suhtautumista dataan ja miten he hyödyntävät sitä

raportteja laatiessaan. Tutkimuksessa huomattiin kokemattomimpien rekrytoijien antavan huomattavasti suurempaa painoarvoa datasta johdetuille suoritusanalyysille, mutta suurin osa myös kokeneemmista rekrytoijista hyödynsivät muun muassa videoanalyysiä työssään ja pitivät suoritusmittareita tärkeinä tehtäessä arvioita mahdollisesta hankinnasta. Suoritusanalyysit antavat faktuaalista kuvaa pelaajan suoriutumisen ja auttavat seuran päättäjiä hahmottamaan pelaajan kykyjä sekä punnitsemaan mahdollista sopivuutta valmentajan asettamaan pelityyliin.

Erilaiset tilastot ja mallit voivat auttaa pelaajan suoriutumisen arvioinnissa ja niiden avulla seurat voivat tarkastella minkä tyyppisiä pelaajia he haluavat omaan kokoonpanoonsa. Maaliodottamaa tarkastelemalla voidaan saada hyvä kuva varsinkin hyökkääjän kontribuutioista pelin tärkeimpään osa-alueeseen eli maalintekoon. Anzer & Bauer (2021) vertailivat Bundesliiga pelaajien maaliodottamalukemia toisiinsa ja pystyivät lukujen perusteella tarkastelemaan, kuinka tehokas yksittäinen pelaaja on ollut omien maalipaikkojensa hyödyntämisessä. Heidän tutkimallaan Bundesliiga kaudella 2019/2020 oli yhteensä 11 pelaajaa, jotka olivat tehneet enemmän maaleja kuin heidän maaliodottamansa antaa ymmärtää. Tämä kertoo pelaajan tehokkaasta maalipaikkojen hyödyntämisestä. Cefis & Carpita (2025) tutkivat erilaisten tekijöiden vaikutusta maaliodottaman muodostumiseen muun muassa maalivahdin tai laukojan ympärillä olevien pelaajien sijoittumista. Näiden avulla voidaan arvioida esimerkiksi, kuinka paljon maalivahti on omalla toiminnallaan voinut vähentää vastustajajoukkueelle syntyneitä maaliodottamaa tai kuinka puolustajien sijoittuminen tai pelaaminen hyökkääjää vastaan on vaikuttanut syntyneeseen maaliodottamaan.

Maaliodottama saattaa olla hyvä mittari tietyn profiilin pelaajien arvioinnissa, jos heidän pääasiallinen tehtävänsä kentällä on tehdä maaleja. Suurinta osaa pelaajista ei kuitenkaan voi arvioida ainoastaan sillä, kuinka paljon he tuottavat maaliodottamaa. Pelaajahankinnan näkökulmasta on löydettävissä mittareita, joilla pelaajien suoriutumista voidaan arvioida kokonaisvaltaisemmin. Elsharkawi ja muut (2025) loivat metriikan, joka arvioi pelaajan vaikutusta lopputulokseen ottamalla huomioon niin offensiiviset kuin defensiiviset teot. He huomasivat, että palveluiden, kuten WhoScored pelaaja-arvosanat huomioivat heikosti merkittäviä puolustuspelillisiä tapahtumia, joilla on toisaalta ollut suuri merkitys ottelun lopputulokseen. Tämänkaltaiset löydökset voivat olla merkittäviä tehtäessä hankintoja puolustuspäähän tai mikäli joukkue etsii esimerkiksi monikäyttöistä keskikenttäpelaajaa laajentamaan rosteria.

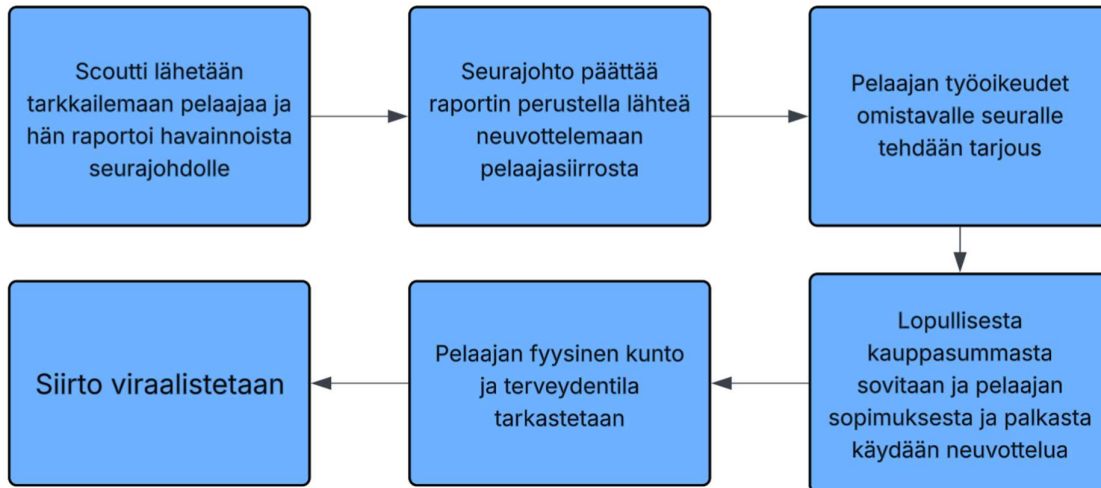
Pelaajien arvon määrittäminen voi olla hyvinkin hankalaa eikä palvelut, jotka tarjoavat arvioita pelaajien hinnoista yleensä perusta arvioitaan tieteellisille menetelmille. McHale ja Holmes (2023) hyödynsivät koneoppimismalleja ennustaessaan pelaajien markkina-arvoja. Mallit arvioivat

historiassa tapahtuneita pelaajasiirtoja ja antoi niille arvon sen perusteella, kuinka paljon pelaaja on tuottanut arvoa joukkueelle suhteessa maksettuun siirtosummaan. Mallit antoivat lisäksi pelaajan mahdollisesta arvosta 5 % ja 95 % kvantiilit, jotka antavat osviittaa pelaajan siirtosummasta. Nämä mallit voivat avustaa siirtoneuvotteluissa antamalla neuvotteluihin raamit mistä summista suurin piirtein lähdetään liikkeelle. Tutkimuksessa todettiin lisäksi pelaajaa tavoittelevalla seuralla olevan merkitystä lopulliseen siirtosummaan. Korkean profiilin joukkueet usein joutuvat maksamaan pelaajista suhteellisesti enemmän kuin matalamman profiilin seurat. Mallien avulla seurat voivat pyrkiä myös ennustamaan hankintansa markkina-arvon kehitystä. Tämä auttaa suunnittelemaan hankintapolitiikkaa ja toisaalta voi rohkaista ottamaan riskejä nuorien lupaavien pelaajien suhteen.

Taktinen analyysi tai pelaajatyypin tunnistaminen videokuvan perusteella on hyvin aikaa vievää ja automatisoitavissa koneoppimismallien avulla (Olthof & Davis, 2025). Koneoppimismallit voidaan valjastaa tunnistamaan datan perusteella pelitavallisesti samankaltaisia pelaajia (Aalbers & Van Haaren, 2019). Mallien avulla valmentajat ja analyttikot voivat pyrkiä ymmärtämään pelaajien profiilia paremmin, tekemään taktisia muutoksia pelin aikana, auttaa pelaajia suoriutumaan omilla vahvuuksillaan tai laajentaa omaa kokoonpanoaan etsimällä samankaltaisia pelaajia, joiden markkina-arvo on syystä tai toisesta alhaisempi, mutta roolitus ja pelityyli vastaavat toisiaan (Seri ym., 2025).

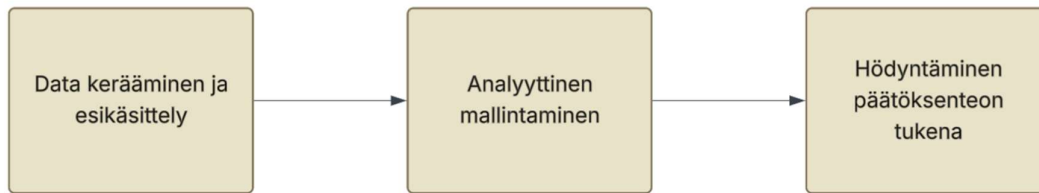
3.3 Päätöksenteon prosessi

Päätöksenteko pelaajahankinnassa on edelleen monessa seurassa seurajohdon ja valmentajien intuition varassa. Kyse ei siis ole siitä, ettei dataa tai päätöksenteon tukijärjestelmiä olisi käytössä vaan niitä ei joko osata käyttää tai ne tarkoituksella sivuutetaan. (Kim ym., 2021) Kuva 1 havainnollistaa pelaajahankintaprosessin etenemistä. Ennen 2000-luvun vaihdetta ja urheiluanalytiikan popularisoidumisen alkua (Fujii, 2025) seurat ovat luottivat scouttien henkilökohtaiseen arvioon pelaajasta ja hänen sopivuudestaan joukkueeseen. Tämä saattoi ajoittain johtaa ylimaksamiseen pelaajasta, joka ei tuonut joukkueelle siirtosumman olettaa arvoa tai kilpailuetua. (Singh & Suguna, 2023) Historiassa on tällaisista pelaajasiirroista monta esimerkkiä. Kim ja muut (2021) nostivat artikkelissaan esiin Andriy Shevchenkon siirron lontoolaisseura Chelsea FC:n riveihin vuonna 2006. Shevchenkon tilastot maalinteon kyvyn osalta olivat olleet laskussa jo ennen siirtoa ja suuri siirtosumma oletti pelaajan suoriutuvan maailman luokan tasolla. Intuiitiivinen supertähtipelaajan hankinta muistetaan seurassa edelleen yhtenä huonoimmista pelaajasiirroista, kun Shevchenkon Chelsea uran saldoksi jäi yhdeksän maalia 58 ottelussa.



Kuva 1 Pelaajahankintaprosessin eteneminen (mukaillen Kim ym., 2021)

Urheiluseurat kasvavissa määrin tunnistavat datan merkittävänä tekijänä kilpailukyvyyn säilyttämisessä niin taloudellisesti kuin kentän tasolla (Fujii, 2025). Mănescu (2025) esittelee artikkelissaan viitekehysten data-analyysin implementointiin urheiluseurojen päätöksenteossa. Kuva 2 visualisoi prosessin etenemistä raa'asta datasta informaatioksi tukemaan päätöksentekoa. Mănescu (2025) kuvaa dataprosessin, joka tarjoaa seurajohdolle käyttökelpoista informaatiota datapohjaiseen päätöksentekoon. Ensimmäisessä vaiheessa dataa kerätään, varastoidaan ja puhdistetaan varmistaen samalla ehjän ja laadukkaan datan taltioiminen myöhempään käyttöön. Prosessin seuraavassa vaiheessa data analysoidaan ja mallinnetaan käyttökelpoiseen muotoon päätöksentekijöille. Dataa käytetään myös koneoppimismallien kouluttamiseen ja niiden ennustavuuden parantamiseen. Viimeisessä vaiheessa dataa hyödynnetään päätösten tukena. Tämä voi koskea taktisia valintoja seuraavassa ottelussa, fyysisen kuormituksen hallinnointia tai päätöstä mahdollisesti tehtävästä pelaajahankinnasta.



Kuva 2 Datan käsittelyprosessi (mukaiillen Mănescu, 2025)

Tarkkaan mietityn ja ehjăn dataprosessin hallinnointi on edellytys datan potentiaalin valjastamiselle. Varsinkin koulutettaessa koneoppimismalleja, datan tulee olla laadukasta ja yhtenäistä, muuten mallien kouluttaminen on mahdotonta, eikä se kykene tunnistamaan toistuvia elementtejä tai rakenteita (Singh & Suguna, 2023). Kun prosessi on kunnossa, data voidaan valjastaa päätöksenteon tueksi lähes reaaliajassa esimerkiksi kesken ottelun. Automaattiset hälytykset pelaajan liiallisesta kuormituksesta voivat reaaliajassa auttaa valmentajaan ottamaan tähtipelaaja vaihtoon, mikäli dataa on viitteitä lisääntyneestä loukkaantumisriskistä. (Mănescu, 2025)

Mănescu (2025) kuvaa artikkelissaan, miten data otetaan mukaan päätöksentekoon. Käyttötarkoituksen mukaan haluttuun analyysiin tulee määrittää avain mittarit kontekstiin liittyvien kyvykkyysien tarkkailuun ja analysointiin. Datan hankintaan ja taltiointiin hyödynnetään kuvan 2 prosessia. Tarkoitus on saada laadukasta, ehjää ja käyttökelpoista dataa liittyen tutkittavaan kontekstiin. Seuraava vaihe on datan analysointi ja mallintaminen informatiiviseksi kojelaudaksi (dashboard), joka mallintaa valittuja suoriutumisen mittareita (key performance indicators, KPI's) ja koneoppimismallien ennusteita tutkittavista ilmiöistä, esimerkiksi markkina-arvon kehittymisestä tai pelaajan kehityspotentiaalista. Kojelaudat implementoidaan osaksi päätöksenteko prosesseja ja voidaan ottaa mukaan esimerkiksi neuvottelujen tueksi. Viimeisenä on jatkuva mallien kehittäminen, testaaminen ja niiden suoriutumisen arviointi.

4 Yhteenveto ja johtopäätökset

Tässä tutkielmassa tutkittiin miten ammattilaisjalkapalloseurat hyödyntävät dataa ja analytiikkamenetelmiä edistääkseen taloudellisesti kestäväää ja urheilullisesti kilpailullista pelaajahankintaa. Tutkielmassa pureudutaan datan eri muotoihin, haasteisiin datan keräämisessä ja käytössä sekä lopulta sen hyödyntämiseen osana pelaajahankintaprosessia. Tutkielmassa tarkasteltiin lisäksi erikategorioita, joissa data auttaa arvioimaan tulevaa pelaajainvestointia ja millä menetelmillä data muutetaan käyttökelpoiseksi informaatioksi.

Ensimmäinen tutkimuskysymykseni liittyi dataan. Mitä dataa kerätään ja mitä haasteita siihen liittyy? Kentän tapahtumien kontekstissa jako tapahtuu usein kahteen kategoriaan, tapahtuma ja sijaintidataan. Niiden avulla tarkastellaan pelaajan suoriutumista kentällä ja liikettä suhteessa pelivälineeseen tai toisiin pelaajiin. Videodata on lisäksi suosittu data muoto taktisessa suunnittelussa ja pelaajahankinnassa. Myös pelaajien fyysistä kuntoa, kuormitusta ja palautumista seurataan, jotta esimerkiksi loukkaantumisriskejä voidaan minimoida. Haasteet datassa liittyvät niiden saatavuuteen, datan painottumiseen pelin tiettyihin aspekteihin tai epäoptimaaliseen rakenteeseen.

Toinen tutkimuskysymykseni oli, miten dataa hyödynnetään ja miten se vaikuttaa päätöksentekoon pelaajahankinnoissa. Suuresta määrästä erilaista dataa voidaan rakentaa kattava tunnuslukujen joukko, joiden avulla pelaajan arviointi ja päätös pelaajahankinnasta voidaan perustella. Suoriutumisen analysointi auttaa seuroja toteamaan pelaajan olevan heidän tarpeeseensa riittävän hyvä, markkina-arvoanalyysi auttaa arvioimaan pelaajan mahdollista siirtosummaa ja taktinen analyysi arvioimaan tarpeen tietyn tyyppiselle pelaajalle tai hänen sopivuuttansa pelisysteemiin. Datan tehtävä ei suinkaan ole korvata alan ammattilaisten kontekstiosaamista esimerkiksi pelaajahankinnassa vaan tukea sitä. Datan avulla voidaan tunnistaa pelaajan ominaisuuksia, katsoa videotallenteita pelisuoritteista ja jopa ennustaa kehityskulkuja. Päättäjät saavat laadukkaan datanprosessin läpi tuotettuja kojelautoja, joissa näkyvät selkeästi määritellyt avain mittarit pelaajahankinnan kannalta. Näiden avulla päättäjät saavat ajankohtaisen tiedon tukemaan pelaajahankintapäätöstä. Koneoppimis- ja tilastolliset mallit auttavat scoutteja ja päättäjiä arvioimaan pelaajia, vertailemaan heitä keskenään ja omien resurssien rajoissa tekemään tietoon perustuvan ratkaisun pelaajahankinnasta. Mallit auttavat heitä myös tunnistamaan omasta kokoonpanosta aukkoja tai tarpeen laajentaa kokoonpanoa. Malleja voidaan kouluttaa ennustamaan hankinnan markkina-arvon muutosta ja tällä tavalla voidaan perustella taloudellista investointia nuoreen ja lupaavaan pelaajaan.

Jalkapalloseurat siis hyötyvät datasta ja sen hyödyntämisestä pelaajahankinnassa. Scouteilla on käytössään ennennäkemätön määrä resursseja pelaajia arvioidessaan, eikä raporteja pelaajasta tarvitse perustaa vain intuition varaan. Organisaation johdon näkökulmasta koneoppimismallein tehostetut ennusteet tai edistyneet analyttiset mallit auttava laskemaan riskitasoja ja punnitsemaan vaihtoehtoja numeerisen faktatiedon pohjalta. Löydökset osoittavat, miten datan analysointi auttaa organisaatioita kehittämään päätöksentekoprosesseja ja tekemään datalla perusteltavissa olevia päätöksiä, jotka laskevat epävarmuutta pelaajahankinnassa.

Tämä tutkielma rajoittuu tarkastelemaan vain olemassa olevaa akateemista kirjallisuutta. Suoriutumisen mittaamisesta ja eri koneoppimismallien hyödyntämisestä moniulotteisen datan analysoinnissa löytyi runsaasti, mutta tutkimusta pelaajahankintaprosessin tai datan hyödyntämisestä osana prosessia on kohtuullisen vähän. Tulevaisuudessa mielenkiintoinen tutkimuskohde olisi, miten data implementoidaan osaksi urheiluorganisaation eri tason päätöksentekoa. Aiheesta ei toistaiseksi ole olemassa tutkimusta. Olemassa ei myöskään ole case-tutkimusta, miten pelaajahankintaa tehdään ammattilaisjalkapallo-organisaatioissa ja millaisia dataratkaisuja todellisuudessa organisaatioissa käytetään. Akateemisessa kirjallisuudessa tutkielmaa laatiessani törmäsin muutamaan mainintaa yleisellä tasolla, että pelaajia arvioidaan datan pohjalta, mutta nämä lähteet eivät paljastaneet tarkempia yksityiskohtia. Tämä saattaa johtua seurojen halusta pitää omat datakyvykkyytensä kilpailevien organisaatioiden saavuttamattomissa.

5 Lähteet

- Aalbers, B., & Van Haaren, J. (2019). Distinguishing Between Roles of Football Players in Play-by-Play Match Event Data. Teoksessa U. Brefeld, J. Davis, J. Van Haaren, & A. Zimmermann (Toim.), *Machine Learning and Data Mining for Sports Analytics* (Vol. 11330, s. 31–41). Springer International Publishing. https://doi.org/10.1007/978-3-030-17274-9_3
- Aarons, E. (2020, joulukuuta 19). Top football clubs relying on transfer valuations made by volunteers. *The Guardian*. <https://www.theguardian.com/football/2020/dec/19/top-football-clubs-relying-on-transfer-valuations-made-by-volunteers>
- Adam, D. (2022). Science and the World Cup: How big data is transforming football. *Nature*, *611*(7936), 444–446. <https://doi.org/10.1038/d41586-022-03698-1>
- Alves, R. (2025). SCORE: A convolutional approach for football event forecasting. *International Journal of Forecasting*, *41*(4), 1636–1652. <https://doi.org/10.1016/j.ijforecast.2025.02.004>
- 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>
- Anzer, G., Bauer, P., & Brefeld, U. (2021). The origins of goals in the German Bundesliga. *Journal of Sports Sciences*, *39*(22), 2525–2544. <https://doi.org/10.1080/02640414.2021.1943981>
- Cappello, G., Defeudis, A., Giannini, V., Mazzetti, S., & Regge, D. (2023). Artificial Intelligence in Oncologic Imaging. Teoksessa E. Neri & P. A. Erba (Toim.), *Multimodality Imaging and Intervention in Oncology* (s. 585–597). Springer International Publishing. https://doi.org/10.1007/978-3-031-28524-0_24
- Cefis, M., & Carpita, M. (2025). A new xG model for football analytics. *Journal of the Operational Research Society*, *76*(1), 1–13. <https://doi.org/10.1080/01605682.2024.2323669>
- Coates, D., & Parshakov, P. (2022). The wisdom of crowds and transfer market values. *European Journal of Operational Research*, *301*(2), 523–534. <https://doi.org/10.1016/j.ejor.2021.10.046>

- 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>
- Deloitte. (2025, kesäkuuta). *Annual Review of Football Finance 2025 | Deloitte UK*. <https://www.deloitte.com/uk/en/services/consulting/research/annual-review-of-football-finance-europe.html>
- do Rosário, C. R., Kipper, L. M., Frozza, R., & Mariani, B. B. (2015). Modeling of tacit knowledge in industry: Simulations on the variables of industrial processes. *Expert Systems with Applications*, 42(3), 1613–1625. <https://doi.org/10.1016/j.eswa.2014.09.023>
- Dubois, L.-E., & Walzak, L. (2025). Blind scouting: Using artificial intelligence to alleviate bias in selection. *Personnel Review*, 54(4), 953–970. <https://doi.org/10.1108/PR-02-2024-0130>
- Elsharkawi, M., Ali, R. H., & Khan, T. A. (2025). Crafting a Player Impact Metric through analysis of football match event data. *Journal of Computational Mathematics and Data Science*, 15, 100115. <https://doi.org/10.1016/j.jcmds.2025.100115>
- Fifa. (2025, syyskuuta 3). *Global transfer market hits new all-time highs in 2025 mid-year window*. <https://inside.fifa.com/transfer-system/media-releases/global-transfer-market-new-all-time-highs-2025-mid-year-window>
- Frost, W., Groom, R., & Nicholls, S. B. (2025). The use of performance analysis and data driven approaches within senior men’s football recruitment. *International Journal of Sports Science & Coaching*, 20(2), 604–616. <https://doi.org/10.1177/17479541251315948>
- Fujii, K. (2025). *Machine Learning in Sports: Open Approach for Next Play Analytics*. Springer Nature Singapore. <https://doi.org/10.1007/978-981-96-1445-5>

- Harper, D. J., Carling, C., & Kiely, J. (2019). High-Intensity Acceleration and Deceleration Demands in Elite Team Sports Competitive Match Play: A Systematic Review and Meta-Analysis of Observational Studies. *Sports Medicine*, *49*(12), 1923–1947. <https://doi.org/10.1007/s40279-019-01170-1>
- Khanday, N. Y., & Sofi, S. A. (2021). Deep insight: Convolutional neural network and its applications for COVID-19 prognosis. *Biomedical Signal Processing and Control*, *69*, 102814. <https://doi.org/10.1016/j.bspc.2021.102814>
- Kim, Y., Bui, K.-H. N., & Jung, J. J. (2021). Data-driven exploratory approach on player valuation in football transfer market. *Concurrency and Computation: Practice and Experience*, *33*(3), e5353. <https://doi.org/10.1002/cpe.5353>
- Kovalchik, S. A. (2023). Player Tracking Data in Sports. *Annual Review of Statistics and Its Application*, *10*(Volume 10, 2023), 677–697. <https://doi.org/10.1146/annurev-statistics-033021-110117>
- Larkin, P., & O'Connor, D. (2017). Talent identification and recruitment in youth soccer: Recruiter's perceptions of the key attributes for player recruitment. *PLOS ONE*, *12*(4), e0175716. <https://doi.org/10.1371/journal.pone.0175716>
- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. W. W. Norton & Company.
- Mănescu, D. C. (2025). Big Data Analytics Framework for Decision-Making in Sports Performance Optimization. *Data*, *10*(7). <https://doi.org/10.3390/data10070116>
- Mason, R. O. (2003). Ethical Issues in Artificial Intelligence. Teoksessa H. Bidgoli (Toim.), *Encyclopedia of Information Systems* (s. 239–258). Elsevier. <https://doi.org/10.1016/B0-12-227240-4/00064-2>
- McHale, I. G., & Holmes, B. (2023). Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *European Journal of Operational Research*, *306*(1), 389–399. <https://doi.org/10.1016/j.ejor.2022.06.033>
- Olthof, S., & Davis, J. (2025). Perspectives on data analytics for gaining a competitive advantage in football: Computational approaches to tactics. *Science and Medicine in Football*, *0*(0), 1–13. <https://doi.org/10.1080/24733938.2025.2533784>

- Robberechts, P., & Davis, J. (2020). How Data Availability Affects the Ability to Learn Good xG Models. Teo-
kessa U. Brefeld, J. Davis, J. Van Haaren, & A. Zimmermann (Toim.), *Machine Learning and Data
Mining for Sports Analytics* (s. 17–27). Springer International Publishing.
https://doi.org/10.1007/978-3-030-64912-8_2
- Seri, E., Rocci, R., & Murphy, T. B. (2025). Partial membership models for soft clustering of multivariate
football player performance data. *Computational Statistics*. <https://doi.org/10.1007/s00180-025-01655-w>
- Sharma, T., Bagga, P., Ahuja, K., & Sharma, S. (2025). Comparative Analysis of Machine Learning Models for
Predicting Top Goal Scorer and Goalkeeper Performance in Football. *2025 3rd International Confer-
ence on Disruptive Technologies (ICDT)*, 1618–1622.
<https://doi.org/10.1109/ICDT63985.2025.10986354>
- Singh, A. P., & Suguna, M. (2023). Data-Driven Player Recruitment in Football. *2023 2nd International Con-
ference on Automation, Computing and Renewable Systems (ICACRS)*, 844–850.
<https://doi.org/10.1109/ICACRS58579.2023.10404860>
- Takamido, R., Ota, J., & Nakamoto, H. (2025). PassAI: An Explainable Machine Learning Framework for Pre-
dicting Soccer Pass Outcomes Using Multimodal Match Data. *IEEE Access*, *13*, 132884–132898.
<https://doi.org/10.1109/ACCESS.2025.3589903>
- Teitelbaum, J., & Brett, K. (2025, heinäkuuta 28). *Forbes 2025 Global Soccer Team Valuations—Ranked List*.
[https://www.forbes.com/sites/justinteitelbaum/2025/05/30/the-worlds-most-valuable-soccer-
teams-2025/](https://www.forbes.com/sites/justinteitelbaum/2025/05/30/the-worlds-most-valuable-soccer-teams-2025/)
- Thantilage, R. D., Le-Khac, N.-A., & Kechadi, M.-T. (2023). Healthcare data security and privacy in Data
Warehouse architectures. *Informatics in Medicine Unlocked*, *39*, 101270.
<https://doi.org/10.1016/j.imu.2023.101270>
- Toda, K., Teranishi, M., Kushiro, K., & Fujii, K. (2022). Evaluation of soccer team defense based on prediction
models of ball recovery and being attacked: A pilot study. *PLOS ONE*, *17*(1), e0263051.
<https://doi.org/10.1371/journal.pone.0263051>

- Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Medicine - Open*, 8(1), 15. <https://doi.org/10.1186/s40798-022-00408-z>
- Vidal, A. S., Sanchez-Anguix, V., & Alberola, J. M. (2025). A Supervised Clustering Approach to Detect Similar Soccer Players. Teoksessa V. Julian, D. Camacho, H. Yin, J. M. Alberola, V. B. Nogueira, P. Novais, & A. Tallón-Ballesteros (Toim.), *Intelligent Data Engineering and Automated Learning – IDEAL 2024* (Vol. 15347, s. 109–119). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-77738-7_10
- Wang, Z., Veličković, P., Hennes, D., Tomašev, N., Prince, L., Kaisers, M., Bachrach, Y., Elie, R., Wenliang, L. K., Piccinini, F., Spearman, W., Graham, I., Connor, J., Yang, Y., Recasens, A., Khan, M., Beauguerlange, N., Sprechmann, P., Moreno, P., ... Tuyls, K. (2024). TacticAI: An AI assistant for football tactics. *Nature Communications*, 15(1), 1906. <https://doi.org/10.1038/s41467-024-45965-x>
- Yazici, İ., Shayea, I., & Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. *Engineering Science and Technology, an International Journal*, 44, 101455. <https://doi.org/10.1016/j.jestch.2023.101455>
- Yeung, C., Ide, K., Someya, T., & Fujii, K. (2025). OpenSTARLab: Open approach for spatio-temporal agent data analysis in soccer. *Complex & Intelligent Systems*, 11(8), 342. <https://doi.org/10.1007/s40747-025-01965-y>

Liite 1

Tekoälytyökalujen käyttö kandidaattitutkielmassa

Kandidaattitutkielmassani olen hyödyntänyt generatiivisen tekoälyn työkaluja (ChatGPT, Scopus AI, Microsoft Word automaattinen kielentarkastus) muun muassa luonnostellessani tutkielman rakennetta, etsiessäni tieteellisiä lähteitä, terminologian kääntämiseen ja selkeyttämiseen sekä käännöstyöhön. Tekoälyä ei ole käytetty tekstin tuottamiseen, aineiston analysointiin tai vastamaan tutkimuskysymyksiin. Tekoälyä on hyödynnetty hyvien akateemisten käytäntöjen mukaisesti avoimesti ja läpinäkyvästi. Tekoäly on toiminut prosessia tukevana työkaluna, mutta ei korvaa tieteellistä analyysiä tai itsenäistä ajattelua.

Tekoälyä on käytetty seuraavilla tavoilla:

- Tutkielman rakenteen suunnittelussa ja hahmottelussa (esimerkiksi 2. ja 3. lukujen rakenne, hyvän johdannon ja yhteenvedon rakenne).
- Tutkimuskysymysten muotoilussa ja tarkentamisessa.
- Oikean suomenkielisen termistön löytämisessä ja käännöstyön tukena.
- Alan termistön selittämisessä ymmärrettävään muotoon (MAE, MAPE, ordinaalinen logistinen regressio).
- Kirjallisuuden haun tukena ja oikeiden hakusanojen löytämisessä.
- Kielen tarkastus Microsoft Wordin automaattisella kielentarkastuksella

Tekoälyä on hyödynnetty yksittäisen kappaleiden sisällön luonnostelussa, mutta kaikki tutkielman teksti on tekijän itse kirjoittamaa.

Tekoälyä ei ole käytetty seuraavaan:

- Tekstin tuottamiseen.
- Tutkimusaineiston generointiin.
- Tutkimuskysymyksiin vastaamiseen.

Tutkielman sisältö ja väittämät perustuvat vertaisarvioituun kirjallisuuteen ja tekijän omaan analyysiin.