

Samuli Peura

**Koneoppimisen hyödyntäminen lupaavien
jalkapalloilijoiden etsinnässä**

Tietotekniikan kandidaatintutkielma

7. kesäkuuta 2024

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Samuli Peura

Yhteystiedot: , samuli.o.peura@student.jyu.fi

Työn nimi: Koneoppimisen hyödyntäminen lupaavien jalkapalloilijoiden etsinnässä

Title in English: Football Player Scouting Using Machine Learning

Työ: Kandidaatintutkielma

Sivumäärä: 18+0

Tiivistelmä: Koneoppiminen jaetaan karkeasti valvomattomiin ja valvottuihin lähestymistapoihin. Tutkielma keskittyy vertailemaan näitä koneoppimisen lähestymistapoja määrittääkseen, kumpi niistä on tehokkaampi jalkapalloilijoiden arvioinnissa. Tutkimuksessa selvitetään myös, mitkä koneoppimisalgoritmit toimisivat parhaiten. Tulokset viittaavat siihen, että koneoppiminen tarjoaa tehokkaita työkaluja pelaajien arviointiprosessiin. Lisäksi erilaiset lähestymistavat toimivat eri tilanteissa riippuen käytettävissä olevasta datasta ja sen eheydestä. Tulokset kuitenkin painottavat valvomattoman lähestymistavan käyttöä pelaajien arviointiprosessissa.

Avainsanat: \LaTeX , kandidaatintutkielmat, koneoppiminen, jalkapallo, pelaaja-arviointi

Abstract: Machine learning is roughly divided to unsupervised and supervised approaches. The study focuses on comparing these machine learning approaches to determine which is more effective given the imbalanced and incomplete nature of football player data. The study also investigates which machine learning algorithms would work the best. Findings suggest that machine learning offers powerful tools for player evaluation process. Furthermore, different approaches work in different situations depending on available data and its integrity. However, study leans towards the use of unsupervised approach in the process of player evaluation.

Keywords: \LaTeX , Bachelor's Theses, Machine learning, Football, Player evaluating

Kuviot

Kuvio 1. (Sarker 2021)	4
Kuvio 2. (Mehmert ja Rein 2018).....	7

Sisällys

1	JOHDANTO	1
2	KONEOPPIMINEN	3
2.1	Koneoppimisalgoritmit.....	3
2.1.1	Tukivektorikone.....	3
2.1.2	Klusterointi.....	4
2.2	Koneoppimismallit	4
2.2.1	Valvottu oppiminen	5
2.2.2	Valvottoman oppiminen	5
3	JALKAPALLOILJOISTA KERÄTTY DATA JA PELAAJIEN ETSINTÄ	6
4	KONEOPPIMISEN SOVELTAMINEN PELAAJIEN KYKYJEN ARVIOINNISSA	9
4.1	Lähestymistavan valinta.....	9
4.2	Koneoppimisalgoritmin valinta	10
5	YHTEENVETO.....	11
	LÄHTEET	13

1 Johdanto

Tutkielmassa on tarkoitus tutkia koneoppimisen hyödyntämistä lupaavien jalkapalloilijoiden etsinnässä. Huippujalkapalloilijoiden tunnistaminen on monimutkainen ja haastava prosessi, sillä huippupelaajat voivat olla ominaisuuksiltaan ja pelaajaprofileiltaan hyvinkin erilaisia. Jalkapalloilijoista on alettu keräämään enemmän ja monipuolisempaa dataa (Pappalardo ym. 2019), joka on mahdollistanut koneoppimisen soveltamisen lupaavien jalkapalloilijoiden tunnistamisessa. Saatavasta datasta on kuitenkin hankalaa päätellä, että kuinka lahjakas pelaaja on, joten koneoppimisesta voisi olla merkittävää hyötyä prosessissa.

Ammattilaisjalkapalloilijoita on rajattu määrä, joten saatavilla oleva data on rajattua (Jauhiainen ym. 2019). Tästä johtuen ongelmaksi muodostuu se, että mitä koneoppimismallia sovelletaan. Tutkielmassa keskitytään etsimään parhaita tapoja soveltamaan koneoppimista, huippupelaajien potentiaalinen mallintamiseen vertailemalla eri lähestymistapoja ongelmaan. Koneoppimisen lähestymistavat voidaan karkeasti jakaa valvottuun ja valvomattomaan oppimiseen (Kelleher ja Tierney 2018). Tässä tutkielmassa vastataan tutkimuskysymykseen: Kumpi koneoppimisen lähestymistavoista valvottu vai valvoton, soveltuu paremmin jalkapalloilijoiden arviointiin. Tutkielmassa tutkitaan myös koneoppimisalgoritmin valinnan vaikutusta tässä prosessissa. Tutkimalla edellä asetettuja kysymyksiä tutkielman tavoitteena on tarjota ymmärrystä siitä, miten koneoppimista tulee soveltaa, jotta se tehostaa potentiaalisten pelaajien tunnistamista optimaalisella tavalla ja näin ollen tukea data-analyytikkoja ja scoutteja tekemään parempia päätöksiä pelaajien arvioinnissa.

Tutkielmassa tutkimusmenetelmänä on systemaattinen kirjallisuuskartoitus, sillä se soveltuu hyvin tutkimuksen tavoitteen saavuttamiseen. Tällä menetelmällä saadaan ratkaisu asetettuun tutkimusongelmaan, rakentamalla ensin kokonaiskuva olemassa olevasta tutkimuksesta liittyen aiheeseen ja sitten johtamalla tutkimuksen tulokset teoreettisen taustan perusteella. Tutkimuksen tausta liittyy vahvasti olemassa olevaan tutkimukseen, joka tutkii koneoppimisen roolia urheilijoiden tunnistamisessa. Aiemmat julkaisut kuten (Jauhiainen ym. 2019) ja (Pappalardo ym. 2019), tarjoavat arvokasta tietoa tutkimukselle.

Tutkielman rakenne on seuraava: johdannon jälkeen toisessa luvussa syvennyttään koneop-

pimisen perusteisiin ja -menetelmiin, kuten algoritmeihin ja koneoppimismalleihin. Luvussa kolme keskitytään jalkapalloilijoista tilastoituun dataan sekä tutkimme erilaisia lähestymistapoja pelaajien suoritusten analysointiin ja potentiaalin arviointiin. Neljäs luku syventyy koneoppimisen soveltamiseen lupaavien jalkapalloilijoiden etsinnässä. Luvussa tutkitaan erilaisia tapoja hyödyntää koneoppimista pelaajien suorituskyvyn estimoimisessa. Viidennessä luvussa tehdään johtopäätöksiä tutkielman tuloksista ja havainnollistetaan, miten koneoppimista voidaan hyödyntää scoutausprosessissa. Lopuksi tiivistetään yhteen tutkielman keskeiset havainnot ja tulokset sekä esitetään jatkotutkimuksen suunnat.

2 Koneoppiminen

Tässä luvussa käydään läpi tutkimukselle keskeisin käsite koneoppiminen, joka perustuu algoritmeihin sekä esitellään eri koneoppimismallivaihtoehdot.

Koneoppimisessa keskiössä on data ja koneoppimisalgoritmit. Data on suuressa roolissa koneoppimisessa, sillä se määrittää kehitetyn mallin lopputuloksen. Koneoppiminen on yksinkertaisuudessaan tietojenkäsittelyn osa-alue, joka on suunniteltu matkimaan ihmisen älykkyyttä oppimalla datasta. Koneoppimisessa tavoite on luoda ohjelma, joka sovituu ohjelmalle annettuun dataan etsimällä siitä säännönmukaisuuksia (Alpaydin 2016). Riskinä tässä on erityisesti ylisovittuminen, joka tarkoittaa ohjelman liaallista mukautumista opetusdataan (Nasteski 2017). Tällöin koulutettu koneoppimismalli ei kykene yleistymään eri syötedatoille. Tämä huomataan, kun testataan kehitettyä mallia testidatalla. Oppiva ohjelma eroaa tavallisesta ohjelmasta siten, että se on muokattavissa koneoppimisalgoritmin parametrejä säättämällä (Alpaydin 2016). Koneoppimisen soveltamisessa, suurin haaste on löytää algoritmi, jonka oppimisvinouma vastaa tiettyä datajoukkoa (Kelleher ja Tierney 2018). Seuraavaksi pohditaan, mitä koneoppimisalgoritmit ovat ja käydään läpi tutkielmalle oleelliset algoritmit.

2.1 Koneoppimisalgoritmit

Algoritmit ovat sarja sääntöjä, jotka ohjaavat toimintaa ja ratkaisevat halutun ongelman. Koneoppimisalgoritmit tarjoavat vaihtoehdon perinteisille ennalta määritetyille matemaattisille algoritmeille, sillä koneoppimisalgoritmeilla on mahdollisuus oppia joustavasti suoraan syötedatasta (Simeone 2018). Seuraavaksi käydään läpi kaksi tälle tutkimukselle merkittävää koneoppimisalgoritmityyppiä.

2.1.1 Tukivektorikone

Tukivektorikonetta voidaan soveltaa eri tavoin, esimerkiksi tunnistamaan anomalioita (Barman ja Demir 2021). Anomalioiden tunnistaminen datasta tarkoittaa poikkeavien tapahtumien havaitsemista. Jos syötedata sisältää paljon kohinaa, tukivektorikone suoriutuu heikosti (Sarker 2021). Kohinalla datassa viitataan puuttuviin tai satunnaisiin tietoihin, jotka häirit-

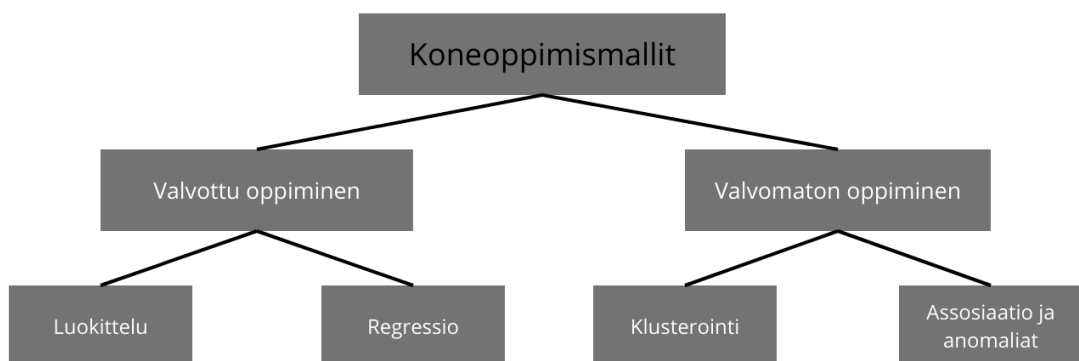
sevät mallin oppimista koulutusvaiheessa.

2.1.2 Klusterointi

Klusterointialgoritmit jakavat syötedatan erillisiin ryppäisiin eli klustereihin. Nämä algoritmit jaetaan kahteen pääkategoriaan: hierarkisiin ja partitiivisiin menetelmiin. Datatieteessä K-lähin naapuri (engl. K-mean) on yleisesti käytetty partitiivinen klusterointialgoritmi. Algoritmissa data jaetaan k määrään ryppäitä valitsemalla ensin k määrä satunnaisia keskiöitä (engl. centroid) ja seuraavaksi toistamalla prosessia niin kauan kunnes löytyy optimaaliset keskiöt dataryppäiden muodostamiseen. K lähin naapuri algoritmi on nopeampi kuin hierarkkinen menetelmä, mikäli muuttujia on monta. (Samreen ym. 2013)

2.2 Koneoppimismallit

Koneoppimisessa koulutetaan koneoppimismalleja hyödyntämällä dataa ja etsimällä siitä säännönmukaisuuksia. Koneoppimismallit voidaan karkeasti jakaa valvotun ja valvomattoman oppimisen malleihin (Kuvio 1) (Sarker 2021). Koneoppimisalgoritmeja voidaan soveltaa eri koneoppimistyyppeihin, kuten esimerkiksi luokitteluun ja regressioon, riippuen siitä mikä soveltuu parhaiten olemassa olevaan syötedataan (Sarker 2021). Myös haluttu lopputulos vaikuttaa koneoppimisalgoritmien valintaan.



Kuvio 1. Koneoppimismallien karkea jako. Kuvio tehty Sarkerin (2021) artikkelissa olevan pohjalta.

Yksi koneoppimisen kriittisimmistä vaiheista on koneoppimismallin valinta (Alpaydin 2016). Koneoppimismallin tulee edustaa mahdollisimman tarkasti valittua data-aineistoa (Nasteski 2017). Tämä mahdollistaa, että malli kykenee yleistymään ja tuottamaan luotettavia ennusteita syötteille, joita se ei ole ennen nähnyt.

2.2.1 Valvottu oppiminen

Valvotussa oppimisessa luodaan koneoppimismalli kouluttamalla opetusdatalla, jota on käsitelty etukäteen. Malli luodaan kuvaamaan syötteiden ja tulosten välistä suhdetta. Yksi yleinen valvotun oppimisen lähestymistapa on luokitteluongelma, jossa opetettu malli luokittelee syötteen tiettyyn luokkaan (Nasteski 2017). Toinen yleinen valvotun oppimisen tyyppi on regressio, jossa toisin kuin luokittelussa tulokset ovat jatkuvia ja saavat numeerisen arvon tietyllä välillä (Sarker 2021).

Valvotulla lähestymistavalla koneoppimismallia kouluttaessa saatavilla oleva data jaetaan opetusdataan ja testidataan. Ohjelma koulutetaan opetusdatalla ja koulutetun mallin virhetä voidaan estimoida vertaamalla sitä koskemattomaan testidataan. Valvottu lähestymistapa suoriutuu heikosti mikäli data ei ole täydellistä tai sitä on niukasti saatavilla. (Alpaydin 2016)

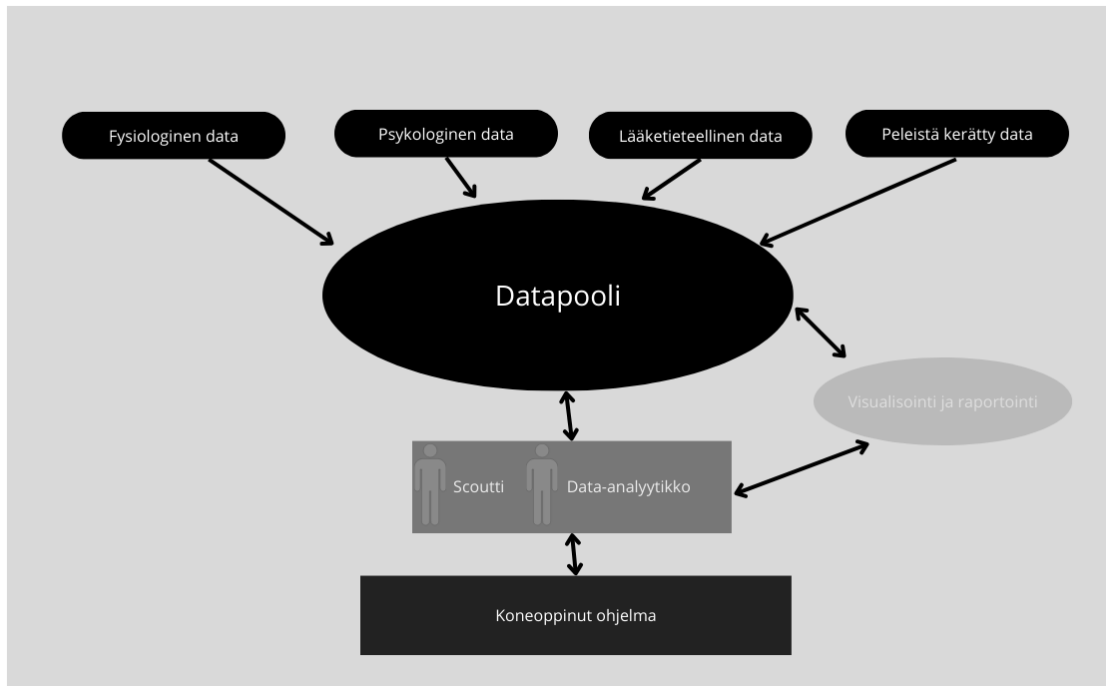
2.2.2 Valvomaton oppiminen

Valvomaton oppiminen on lähestymistapa koneoppimisessa, jossa koneoppimisalgoritmit tutkivat datasta yhteyksiä ilman selkeää ohjausta tai ihmisen syöttämiä esimerkkejä (Samreen ym. 2013). Toisin kuin valvotussa oppimisessa, mikä perustuu ihmisen luokittelman datan tutkimiseen, valvomaton oppiminen sukeltaa syvemmälle itse datan rakenteeseen paljastaakseen yhteyksiä ja anomaliaita. Valvomattoman oppimisen yleisin koneoppimistyyppi on klusterointi eli ryväsanalyysi. Klusterointialgoritmit etsivät datasta ryppäitä, jotka ovat lähempänä toisiaan kuin datan muita tapauksia (Kelleher ja Tierney 2018). Valvomaton lähestymistapa suoriutuu kohinaa sisältävästä datasta huomattavasti valvottua tapaa paremmin.

3 Jalkapalloilijoista kerätty data ja pelaajien etsintä

Pelaajien tilastollinen arviointi ja ominaisuuksien tulkinta ovat erittäin tärkeitä jalkapallossa, jokaisella tasolla ja eri ikäluokissa (Memmert ja Rein 2018). Luvussa pohditaan sen merkitystä pelaajien potentiaalin arvioinnissa sekä miten koneoppiminen voisi olla hyödyllistä tässä prosessissa.

Jalkapallossa on saavutettu digitaalinen vallankumous jo jokin aikaa sitten. Perinteisen otteluista kerätyn tiedon ja pelaajien suoritusten analysoinnin sijasta, on alettu tallentamaan tarkasti pelaajien jokainen suoritus kentällä (Memmert ja Rein 2018). Tämä on luonut uusia tärkeitä mittareita (engl. Key Performance Indicators), esimerkiksi tilan hallinta ja prässäysindeksi. Saatavasta datasta voidaan muodostaa entistä tarkempia visualisointeja. Pelaajista on myös saatavilla entistä enemmän fyysistä, psyykkistä sekä urheilulääketieteellistä dataa (Pappalardo ym. 2019). Saatavilla olevan datan lisääntyminen ja monipuolistuminen mahdollistaa entistä syvemmän datapohjaisen pelaajien arvioinnin. Tämä auttaa valmentajia tekemään paremmin informoituja päätöksiä pelaajien peluutuksessa ja harjoittelujen suunnitteluissa. Modernissa jalkapallo-ottelun analysoinnissa täytyy olla perillä niin tietoteknisestä kuin urheilullisesta puolesta. Täten analyttikko kykenee pysymään tilastoinnin kehityksen perässä sekä kehittämään uusia innovatiivisia kontribuutioita (Memmert ja Rein 2018). Koneoppimisen hyödyntäminen jalkapalloilijoiden kykyjen arvioinnissa mahdollistaa laajempien datamäärien analysoinnin, mikä tuo merkittävää etua joukkueelle. Neljännessä luvussa syvennytään tarkemmin siihen, miten koneoppimisen avulla tehostetaan pelaajien arviointia datapohjaisesti.



Kuvio 2. Esimerkikki data-analyysille jalkapallossa. Kuvio tehty Memmertin ja Reinin (2018) artikkelissa olevan pohjalta.

Kuvio kuvaa data-analyysia jalkapalloilijoiden arvioinnissa. Kuvio esittää, minkä tyyppistä dataa kuuluu datapooliin ja tarjoaa tavan miten hyödyntää koneoppimista prosessissa. Jalkapallossa käytettävän data-analyysimallin tulisi sisältää monipuolisesti erilaista tietoa, jotta voidaan tehdä luotettavia päätelmiä pelaajien potentiaalista. (katso Kuvio 2) Monipuolisen datan saatavuus muodostuu ongelmaksi. Tähän ongelmaan pohditaan ratkaisuja tutkielman neljännessä luvussa.

Potentiaalisten pelaajien etsintä on moniulotteinen prosessi, jossa yritetään tunnistaa parhaat pelaajat, jotka voisivat menestyä tulevaisuudessa (Tom ym. 2022). Tätä työtä tekevät scoutit, jotka arvioivat tehtyjen havaintojen perusteella pelaajan todennäköisyyttä menestyä tulevaisuudessa, sekä sitä miten pelaajat sopisivat tiettyyn joukkueeseen ja sen pelityyliin. Scoutit ovat perinteisesti luottaneet enemmän intuitioon ja kokemukseen pelaajien valinnoissa, jättäen vähemmälle dataan perustuvan analyysin ja dataan pohjautuvien arviointimenetelmien käytön (Tom ym. 2022). Pelaajien kykyjen analysoimisen kompleksisuuden vähentämisek-

si, voidaan käyttää koneoppimista luomaan data-ohjattuja malleja (Memmert ja Rein 2018). Täten koneoppimisen hyödyntäminen scouttausprosessissa helpottaa dataan pohjautuvaa lupaavien jalkapalloilijoiden etsintää ja voi säästää merkittävästi aikaa, mikä aikaisemmin on kulunut suurien datamäärien analysointiin.

4 Koneoppimisen soveltaminen pelaajien kykyjen arvioinnissa

Koneoppimisen lähestymistavan valinta pelaajien kykyjen arvioinnissa on haastavaa. Tässä luvussa vertaillaan valvotun ja valvomattoman lähestymistavan eroja tässä prosessissa sekä tuodaan esille mahdollisia haasteita. Luvussa pohditaan myös mitä koneoppimisalgoritmeja tulisi käyttää kouluttaessa koneoppimismalleja.

Jalkapallossa pelaajia on kentällä kerralla 11 ja erilaisia pelipaikkoja on useita. Tästä johtuen tulee kehittää eri koneoppimismalleja eri pelipaikoille. Huippupelaajia on rajattu määrä, joten saatavilla olevat datasetit eivät ole tasapainoisia (Jauhiainen ym. 2019). Tästä johtuen ongelmaa voitaisiin lähestyä valvomattoman oppimisen kautta. Täten päästään dataan syvemmin käsiksi ja vähennetään koulutettavan koneoppimismallin ylisovittumisriskiä käyttämällä ihmisten käsittelemää dataa. Usein yksinkertaisin on paras, joten on hyvä myös tutkia miten perinteinen valvottu koneoppimisen lähestymistapa voisi toimia. Etenkin tilanteissa joissa dataa on saatavilla runsaasti ja data-aineisto ei sisällä kohinaa, voisi valvottu lähestymistapa olla parempi.

4.1 Lähestymistavan valinta

Soveltamalla valvottua lähestymistapaa eli käyttämällä ihmisten käsittelemää ja luokittelemaan dataa, suurin ja merkittävin riski on koulutetun koneoppimismallin ylisovittuminen opetusdataan (Jauhiainen ym. 2019). Tätä riskiä voidaan pienentää hankkimalla enemmän tasapainoista syötedataa.

Valvomattoman lähestymistavan etuja on datan syvempi tutkinta ja se soveltuu paremmin erilaisille dataseiteille. Todennäköisesti valvomaton lähestymistapa on paras vaihtoehto monessa tapauksessa, sillä jalkapalloilijoista kerätty data on usein epätäydellistä, eikä kaikista jalkapalloilijoista ole saatavilla samanlaista dataa. Toinen merkittävä etu valvomattomassa oppimisessa on kyky löytää piileviä ominaisuuksia datasta, joita ei ole mahdollisesti aikaisemmin osattu arvostaa. Nämä uudet oivallukset huippupelaajien kyvyistä voivat olla merkit-

täviä tulevaisuuden pelaaja-arvioinneissa. Valvotun oppimisen lähestymistapaa puoltaa sen suoraviivaisuus ja erityisesti selkeys tulosten tulkinnassa. Kun on luotu selkeät luokitukset ja merkattu data, valvottu oppiminen tarjoaa selkeämmän vaihtoehdon koneoppimismallien tuottamiseen.

4.2 Koneoppimisalgoritmin valinta

Koneoppimisalgoritmin valinta perustuu hyvin pitkälti saatavilla olevan datan laatuun ja määrään. Barmanin ja Demirin (2021) mukaan valvotuista lähestymistavoista tukivektori-kone suoriutuu parhaiten mallintamaan urheiludataa. Tukivektori-kone kuitenkin toimii heikosti, mikäli data ei ole täydellistä. Tämän ehkäisy onnistuu muokkaamalla dataa ennen koneoppimismallin koulutusvaihetta. Mikäli tämä ei ole mahdollista tulisi valita koneoppimisalgoritmi, joka ei ole niin altis epätäydelliselle datalle, kuten esimerkiksi klusterointialgoritmit. Klusterointialgoritmeista k lähin naapuri on yksi tehokkaimmista ja se soveltuu hyvin tilanteisiin, joissa data on epätäydellistä ja epätarkkaa, kuten jalkapallossa yleensä on.

Näiden tietojen pohjalta koneoppimisalgoritmin valinta on kriittinen vaihe, joka vaikuttaa suoraan mallin kykyyn arvioida pelaajien potentiaalia tarkasti ja luotettavasti. On tärkeää ottaa huomioon sekä datan ominaisuudet että saatavilla olevat resurssit valittaessa sopivaa algoritmia. Tämän vuoksi iteratiivinen ja kokeileva lähestymistapa, jossa testataan useita eri algoritmeja ja hienosäädetään niitä jatkuvasti, voi olla paras strategia optimaalisten tulosten saavuttamiseksi.

5 Yhteenveto

Ammattilaisjalkapalloilijoista kerätyn datan lisääntyessä, koneoppimista on alettu yhä enemmän hyödyntämään pelaajien suoritusten arvioinnissa. Tutkielman pääkysymyksenä oli, että miten eri koneoppimistyylin ja koneoppisalgoritmien valinta vaikuttaa. Tavoitteena oli luoda tietoa ja ymmärrystä miten lähteä lähestymään koneoppimismallin kehittämistä jalkapalloilijoiden arviointiin.

On selvää, ettei ole mahdollista kehittää yhtä koneoppimismallia, joka pystyisi yleistymään toimivaksi jokaisen eri pelipaikan pelaajille. Ratkaisu tähän ongelmaan on kouluttaa eri pelipaikoille omat erilliset koneoppimismallit. Koneoppimismallit voitaisiin erikseen luoda esimerkiksi maalivahdeille, puolustaville kenttäpelaajille ja hyökkääville kenttäpelaajille. Näiden erillisten mallien avulla voidaan saavuttaa parempi tarkkuus ja luotettavuus pelaajien arvioinnissa, sillä jokaisen pelipaikan vaatimukset ja ominaisuudet voivat olla merkittävästi erilaisia.

Valvotun ja valvomattoman oppimisen välillä tehtävä valinta riippuu suuresti käytettävissä olevan datan määrästä ja laadusta. Pienillä syötedatoilla valvotussa oppimisessa on suuri riski ylisovittumiselle, jolloin malli oppii liikaa koulutusdatan erityispiirteitä eikä yleisty hyvin uusiin tapauksiin. Tällaisissa tilanteissa valvottoman oppimismenetelmä on parempi vaihtoehto, sillä se mahdollistaa datan syvemmän analyysin ja vähentää ylisovittumisriskiä. Jos taas dataa on saatavilla runsaasti ja se on tasapainoista, valvottu oppimistapa on tehokkaampi, koska se hyödyntää ihmisten muokkaamaa ja luokittelemaa dataa, joka sisältää arvokasta asiantuntemusta.

Yleensä parhaan ratkaisun löytämiseksi on hyvä kokeilla erilaisia koneoppimisalgoritmeja (Kelleher ja Tierney 2018), joten koneoppimismallin koulutusvaiheessa tulisi kokeilla useampaa eri koneoppimisalgoritmivaihtoehtoa. Nämä vaihtoehdot voisivat olla valvotun oppimisen kohdalla luokittelualgoritmi ja regressioon perustuva tukivektorikone. Valvomattoman oppimisen algoritmeja voisi olla eri klusterointialgoritmit. Tärkeää on kokeilla useampaa vaihtoehtoa ja vertailla niiden suorituskykyä, jotta löydetään parhaiten soveltuva ratkaisu.

Koneoppimisen soveltaminen tuo paljon arvokasta informaatiota scouttaus- ja arviointiprosesseihin. Se ei kuitenkaan tule täysin korvaamaan pelaajascoutteja, vaan pikemminkin tarjoamaan arvokasta lisäinformaatiota, ja tukemaan heitä pelaajien etsintä- ja arviointiprosesseissa. Koneoppimismallit voivat auttaa tunnistamaan piileviä kykyjä ja potentiaalia, joita perinteisillä menetelmillä ei välttämättä havaittaisi. Sekä myös mahdollistaa useamman pelaajan arvioinnin saman aikaisesti.

Tässä tutkielmassa tehty kirjallisuuskatsaus on vasta pintaraapaisu aiheeseen, ja se vaatii syvempää tarkastelua. Jatkossa olisi hyödyllistä tutkia tarkemmin, lisääkö erillisten koneoppimismallien käyttö hyökkäävillä ja puolustavilla pelaajilla mallien tarkkuutta. Lisäksi neuroverkkojen ja syväoppimismallien soveltamisen tutkiminen voisi tarjota uusia näkökulmia ja mahdollisuuksia pelaajien arvioinnissa. Näiden menetelmien avulla voitaisiin saavuttaa entistä parempia tuloksia suurten ja monimutkaisten datamäärien käsittelyssä, mikä voisi merkittävästi parantaa scouttausprosessien tehokkuutta ja tarkkuutta.

Lähteet

Alpaydin, Ethem. 2016. *Machine Learning: The New AI*. 1. painos. ProQuest Ebook Central: MIT Press. Viitattu 19. helmikuuta 2024. <https://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/detail.actiondocID=4714219>.

Barman, I. ja I. Demir. 2021. “Modelling Sport Events with Supervised Machine Learning”. *Fundamental Journal of Mathematics and Applications* 4 (4): 232–244. <https://doi.org/10.33401/fujma.951665>.

Jauhiainen, S., S. Äyrämö, H. Forsman ja J-P. Kauppi. 2019. “Talent identification in soccer using a one-class support vector machine.” *International Journal of Computer Science in Sport* 18 (3): 125–136. <https://doi.org/10.2478/ijcss-2019-0021>.

Kelleher, J. D. ja B. Tierney. 2018. *Datatiede*. Kääntänyt K.Pietiläinen. Terra Cognita.

Memmert, D. ja R. Rein. 2018. “Match analysis, big data and tactics: current trends in elite soccer.” *German Journal of Sports Medicine* 69 (3).

Nasteski, V. 2017. “An overview of the supervised machine learning methods”. *Horizons. b* 4:51–62. <https://doi.org/10.20544/HORIZONS.B.04.1.17.P05>.

Pappalardo, L., P. Cintia, E. Massucco, D. Pedreschi ja F. Giannotti. 2019. “PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via Machine Learning Approach”. *ACM Transactions on Intelligent Systems and Technology* 10 (59): 1–27. <https://doi.org/10.1145/3343172>.

Samreen, N., A. Aqib, A. Sania ja A. Muhammad Munawar. 2013. “An Unsupervised Machine Learning Algorithms: Comprehensive Review”. *International Journal of Computing and Digital Systems* 13 (1). <https://doi.org/10.12785/ijcds/130172>.

Sarker, I.H. 2021. “Machine learning: Algorithms, Real-World Applications and Research Directions”. *SN COMPUT.SCI* 2 (160). <https://doi.org/10.1007/s42979-021-00592-x>.

Simeone, O. 2018. "A Very Brief Introduction to Machine Learning With Applications to Communication Systems". *IEEE Transactions on Cognitive Communications and Networking* 4 (4): 648–664. <https://doi.org/10.1109/TCCN.2018.2881442>.

Tom, L., G. Bergkamp, Wouter G. P. Frencken, A. Susan M. Niessen, Rob R. Meijer ja Ruud. J. R. den Hartigh. 2022. "How soccer scouts identify talented players". *European Journal of Sport Science* 22 (7): 994–1004. <https://doi.org/10.1080/17461391.2021.1916081>.