

Frans Murto

TUOMARIEN KAAVAMAIKUUS SUOMALAIKISSA LIIGAJÄÄKIEKKOILUSSA

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Toukokuu 2020

TIIVISTELMÄ

Frans Murto: Tuomarien kaavamaisuus suomalaisessa liigajääkiekkoilussa
Kandidaattitutkielma
Tampereen yliopisto
Tietojenkäsittelytieteiden tutkinto-ohjelma
Toukokuu 2020

Tuomarien odotetaan toimivan joukkueurheilussa puolueettomina oikeudenvalvojina, jotka mahdollistavat toiminnallaan otteluiden pelaamisen lajisääntöjen puitteissa. Tuomaritoiminnassa on kuitenkin perinteisesti ollut lajien seuraajien mielestä havaittavissa piirteitä, jotka saattavat tämän oikeudenmukaisuuden kyseenalaiseen valoon. Tällaisia stereotyyppioita ovat esimerkiksi kotijoukkueen suosiminen, hyvitykset virheellisten päätösten jälkeen, rangaistusmäärien tasoittelu ja taipumus olla rankaisematta pelaajia otteluiden loppuhetkillä. Tässä tutkielmassa on tarkoitus selvittää suomalaisen liigajääkiekkoilun näkökulmasta, onko näiden kaavamaisuuksien olemassaoloa mahdollista todistaa konkreettisesti. Tutkielma on tiettävästi ensimmäinen, joka tarkastelee eurooppalaisen jääkiekkosarjan tuomaritoimintaa tästä näkökulmasta.

Tuomaritoiminnan kaavamaisuutta tarkastellaan mallintamalla ottelun jäähyjakaumaa soveltamalla eri koneoppimismenetelmiä jäähyjen kontekstidataan, joka hankittiin ja esikäsiteltiin haravointiohjelmalla Liigan verkkosivujen otteluseurannoista kausien 2014–2015 ja 2019–2020 välisen ajan 2663 runkosarjaottelusta. Malleissa käytetty lopullinen data koostui 15918 jäähystä ja niiden 18 kontekstimuuttujasta. Parhaiten dataan sovittuva malli, gradient boosting, valittiin yleistyvyyttä mittaaviin toimenpiteisiin pohjautuen vähäisimmän tappiofunktion perusteella.

Mallin rakennetta tulkitsemalla voidaan arvioida ja havainnollistaa eri muuttujien vaikutusta jäähyjakaumaan ja osoittaa, että peliaika, johto- ja tappioasema sekä erityisesti jäähysuhde ja aikaisemmat peräkkäiset jäähyt vaikuttavat tuomaritoimintaan. Kotietu on läsnä jokaisessa kuvailussa vaikutuksessa ja äärimmillään vierasjoukkueen todennäköisyys saada seuraava jäähy on tietyissä tilanteissa jopa yli 70 %. Tuomareista aiheutuva kotietu ja heidän päätösten ennakoitavuus sekä kontekstiriippuvaisuus aikaisempiin ottelutapahtumiin ovat tutkimuskirjallisuudessa laajalti tunnustettuja ilmiöitä ja tämän tutkielman havainnot tukevat näitä näkemyksiä. Tietyt urheilulajit vaikeiksi tuomaroitaviksi tekevä tulkinnanvaraisuus, inhimilliset tekijät kuten paineensietokyky ja ennakkoluulot, sekä teoria tuomareista vuorovaikuttavina pelinjohtajina osoittavat kohti mahdollista selitystä kaavamaisuuksille tiedostettuina tai tiedostamattomina tuomaritoimintaa helpottavina metasääntöinä.

Avainsanat: jääkiekko, tuomaritoiminta, koneoppiminen, data-analyysi

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Sisällysluettelo

1	Johdanto	1
2	Jäähyjen kontekstidata	2
3	Tutkimusmenetelmä	4
3.1	Koneoppiminen	6
3.1.1	Päätöspuut	7
3.1.2	Gradient boosting	7
3.1.3	Satunnaismetsä	8
3.1.4	Neuroverkot	8
3.2	Mallin valinta	8
4	Tulokset	10
5	Keskustelu	16
6	Yhteenveto	19
	Lähdeluettelo.....	19

1 Johdanto

Jääkiekko on Suomen suosituin urheilulaji ja sen miesten pääsarjataso Liiga on yleisömäärissä mitattuna katsotuin yksittäinen sarja. Jääkiekko-ottelu kestää 60 minuuttia ilman jatkoajoja ja kentällä voi yhdeltä joukkueelta olla kerrallaan viisi kenttäpelaajaa ja yksi maalivahti. Nopeatempoisena kontaktilajina jääkiekossa tapahtuu usein sääntörikkomuksia, jotka johtavat jäähyiksi kutsuttuihin rangaistuksiin. Suurin osa jäähyistä on kahden minuutin pituisia, mutta vakavammista rikkomuksista voidaan antaa viiden minuutin rangaistus. Rangaistun pelaajan joukkue menettää jäähyn ajaksi yhden kenttäpelaajan kuitenkin niin, että kentällä on vähintään kolme kenttäpelaajaa. Tätä tilannetta kutsutaan vastustajajoukkueen kannalta ylivoimaksi. Ylivoimat ovat ottelun lopputuloksen kannalta merkityksellisiä tilanteita, sillä maalin tekemisen todennäköisyys kasvaa tällöin huomattavasti. Esimerkiksi Liigan kaudella 2018–19 maaleja tehtiin tasakentällisin noin 12 minuutin välein siinä missä vastaava luku ylivoimatilanteissa oli noin 8 minuuttia.

Sääntöjen noudattamisesta vastaavat Liigan otteluissa kaksi päätuomaria. Heidän lisäksi otteluissa on myös kaksi avustavaa linjatuomaria, joilla ei kuitenkaan ole valtuuksia määrätä rangaistuksia pelaajille. Kausittain Liigassa toimii noin 40 pää- ja linjatuomaria puoliammattilaispohjalta. Ilmiselvät tai törkeät rikkomukset, kuten kiekon sulkeminen käteen tai päähän kohdistunut taklaus, ovat tuomarin kannalta suoraviivaisia tilanteita, mutta sääntökirjassa määriteltyjen rangaistusten kuvaukset ovat pitkälti tulkinnanvaraisia ja tuomarien antamissa tai antamatta jääneissä jäähyissä voi olla tilanteesta riippuen huomattavan suuri subjektiivisuuden aste.

Jääkiekkoa seuraavan yleisön keskuudessa tiedostetaan useita tilanteita, joissa nämä tulkinnanvaraisuudet vaikuttavat kallistuvan toisen joukkueen puolelle. Tuomari saattaa alitajuisesti suosia kotijoukkuetta antamalla tämän vastustajalle suuremmalla todennäköisyydellä jäähyjä miellyttääkseen häntä äänekkäästi kritisoivaa kotiyleisöä tai rangaista helpommin joukkuetta, jonka vastustaja on saanut useamman peräkkäisen jäähyn ylläpitääkseen jäähysuhdetta ja vaikutelmaa puolueettomuudesta. Jos tuomari on tehnyt aikaisemmin virheen, hän saattaa hyvittää siitä kärsinyttä joukkuetta antamalla tämän vastustajalle jäähyn tilanteesta, josta hän ei sitä tavallisesti tuomitsisi. Otteluiden ratkaisuhetkillä ja pudotuspeleissä annettujen jäähyjen määrä putoaa, koska tuomarit eivät halua heidän päätöksiansä tulkittavan ratkaiseviksi tekijöiksi lopputuloksen kannalta, vaikka sääntöjen mukaan he olisivat oikeutettuja antamaan rangaistuksia tietyistä lopussa tapahtuvista pelitilanteista. Tuomarien oletetaan olevan inhimillisissä rajoissa puolueettomia ja objektiivisia toimijoita, mutta ottelun kontekstista riippuen he vaikuttavat toimivan johdonmukaisesti tavoin, jotka eivät ole tasapuolisia molempia joukkueita kohtaan.

Aikaisempi tutkimus tuomarien kaavamaisuuteen joukkueurheilussa on kohdistunut eurooppalaiseen seurajalkapalloon. Buraimon ja muiden (2010) tutkimustulosten perusteella tuomarit rankaisevat systemaattisesti kotijoukkueita vähemmän kuin vierasjoukkueita vaikutuksen ollessa suurempi silloin, kun katsomot ovat lähempänä kenttää. Dohmen (2008) löysi tuomarien antavan enemmän lisäaikaa otteluissa, joissa kotijoukkue oli tappiolla ja suosivan kotijoukkueita rangaistuspotkujen ja maalien hyväksymisen määrän osalta. Page ja Page (2010) tutkivat tuomarien kykyä käsitellä yleisön aiheuttamaa painetta ja havaitsivat huomattavia yksilötason eroja tuomareista aiheutuvassa kotiedussa yleisömäärästä riippuen.

Jääkiekkoa käsittelevien tutkimusten osuus on selvästi pienempi ja se on keskittynyt pitkälti maailman suurimpaan jääkiekkosarjaan National Hockey Leagueen (NHL). Lopez ja Snyder (2013) löysivät huomattavaa näyttöä siitä, että tuomarit tasoittavat ottelun jäähysuhdetta kotijoukkueiden hyötyen tästä enemmän. Abrevaya ja McCulloch (2014) käyttivät koneoppimisen menetelmiä osoittamaan kotiedun, pelitilanteen sekä peliajan ja jäähyjen välisen ajan vaikuttavan siihen, kumpi joukkue saa seuraavan jäähyä. He löysivät myös näyttöä tuomarien taipumuksesta rangaista joukkueita, joiden vastustajat ovat saaneet useita peräkkäisiä jäähyjä. Beaudoin ja muut (2016) rakensivat näiden tulosten pohjalta ja havaitsivat tuomarien antavan jäähyjä vähemmän joukkueille, joita on rangaistu enemmän ottelussa ja enemmän jäähyjä johdossa oleville joukkueille.

Tässä tutkielmassa tarkastelen tuomarien käyttäytymistä tehokkaiksi osoitetuun koneoppimisalgoritmeihin Liigasta kerätyn otteludatan perusteella löytääkseni pelitilanteiden ulkopuolisia tekijöitä, jotka vaikuttavat rangaistusten määräytymiseen ottelun sisällä. Käyttämäni metodologia perustuu pitkälti Abrevayan ja McCullochin (2014) sekä Beaudoin ja muiden (2016) menetelmiin. Tutkielma on tietävästi ensimmäinen suomalaisen urheilusarjan dataa hyödyntävä kvantitatiivinen tutkimus tuomarikäyttäytymiseen. Hyödynnän eri koneoppimisalgoritmeja ja vertailen niiden tehokkuutta muodostaakseni jäähyjen jakautumista parhaiten kuvastavan mallin. Pystyn osoittamaan parhaan mallin avulla tuomaritoiminnan kaavamaisuuksien olemassaolon.

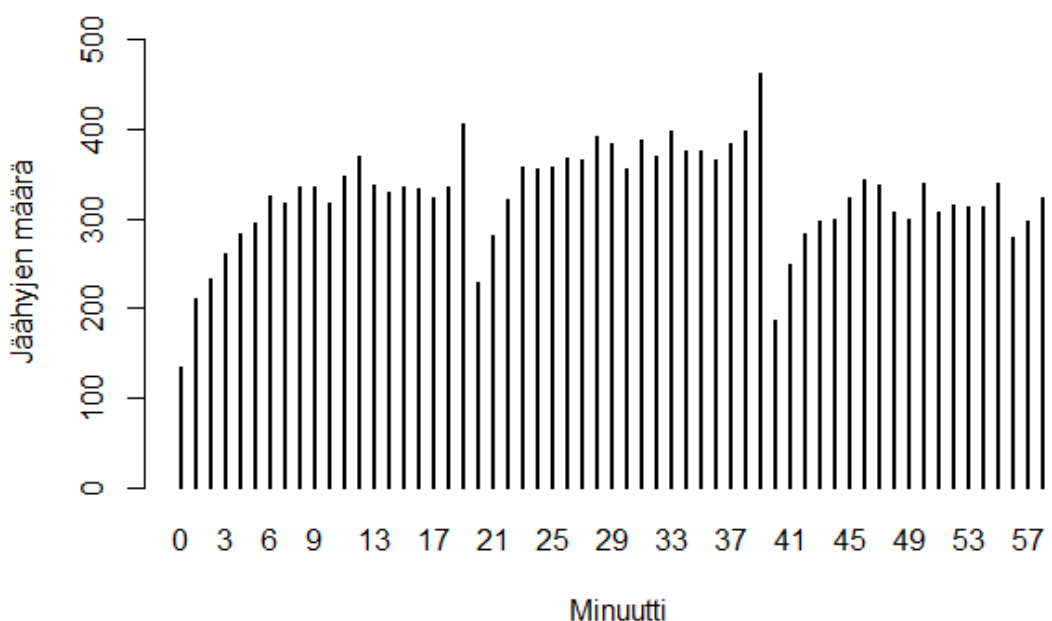
Tutkielma on jaettu neljään osioon: data, menetelmä, tulokset ja keskustelu. Data-osiossa kerron, miten käyttämäni data on kerätty ja esittelen sen ominaisuuksia. Menetelmässä erottelen käyttämäni koneoppimisalgoritmit ja kuvailen, miten mallien valintaprosessi suoritetaan. Tuloksissa tulkitsen parhaan mallin muuttujien vaikutuksia ja keskusteluosiossa käyn läpi tulosten merkityksellisyyttä ja niiden mahdollisia selityksiä.

2 Jäähyjen kontekstidata

Tässä tutkielmassa käytetty data on kerätty kaikista saatavilla olevista Liigan verkkosivujen (Liiga, 2020) otteluiden liveseurannoista käyttämällä Python-kielellä kirjoitettua, Scrapy-kirjastoon (GitHub, 2020) perustuvaa haravointiohjelmaa.

Verkkosivujen haravointi eli *web scraping* tarkoittaa (useimmiten automatisoitua ja ohjelmallista) tiedon keräämistä verkossa sijaitsevista lähteistä tavoilla, jotka eivät hyödynnä valmiita rajapintoja (Mitchell, 2018). Jäähyjen tiedot ja niiden konteksti kerätään otteluseurannan tapahtumalistasta ja data muotoillaan taulukoksi mahdollistaen helpon jatkokäsittelyn (Murto, 2020). Lopullisesta datasta karsitaan otteluiden ensimmäiset jäähyt, joista puuttuu tutkimuksessa käytettäviä muuttujia sekä jäähyt, joista ei aiheudu ylivoimaa kummallekkaan joukkueelle näiden ollessa epäolennaisia tutkimuskysymyksen kannalta. Tällaisia ovat tilanteet, joissa molemmat joukkueet ovat saaneet saman määrän jäähyminuutteja.

Datan muodostamiseen on tarkasteltu 2663 runkosarjaottelun tietoja kausien 2014–2015 ja 2019–2020 väliseltä ajalta yhteensä kuuden kauden verran. Tällä aikavälillä vain kolmeen otteluun ei ollut kirjattu yhtään jäähyä. Tarkasteltavina olleiden kausien aikana annettiin yhteensä 19838 rangaistusta, mikä tarkoittaa noin 7,45 rangaistuksen keskiarvoa per ottelu. Yleisimmät jäähyjen syyt olivat kampitus (20,29 %), koukkaaminen (13,96 %), kiinnipitäminen (9,97 %) ja korkea maila (9,46 %). Ensimmäisten, samanaikaisten ja puuttellisten tietojen omaavien jäähyjen poistamisen jälkeen mallinnettava data koostuu 15918 rangaistuksesta.



Kuva 1. Jäähyjen jakautuminen minuuteittain.

Jäähyt jakautuivat epätasaisesti joukkueiden kesken, sillä vierasjoukkueiden osuus rangaistuksista varsinaisella peliajalla oli 53,62 %. Jatkoajalla annetuissa jäähyissä tämä

osuus kasvaa 56,28 %. Beaudoin ja muut (2016) raportoivat samanlaisesta jakaumasta NHL:n osalta. Kuvasta 1 voi huomata jäähyjen jakaantuvan varsinaisella peliajalla riippuen siitä, kuinka paljon ottelua on pelattu. Jääkiekko-ottelun varsinainen peliaika kestää 60 minuuttia jaettuna kolmeen 20 minuutin mittaiseen erään. Jäähymäärät ovat matalimmillaan erien aluissa, tasaantuvat keskivaiheilla ja pysyvät samoina tai laskevat hieman loppua kohti. Jäähyjä annetaan eniten toisessa erässä ja niiden määrä laskee selvästi kolmanteen erään, mikä tukee käsitystä tuomarien haluttomuudesta näyttäytyä ratkaisevana tekijänä ottelun lopputuloksen kannalta viheltämällä rangaistuksia pelien loppuhetkillä. Tuomarilinjan voi siis olettaa tiukentuvan toiseen erään asti ja muuttuvan kolmannessa erässä kohti rikkomuksia sallivampaan suuntaan. Samansuuntaista näyttöä jäähymäärän vähenemisestä erien aluissa ja kolmannessa erässä on löydetty myös NHL:n otteluista kerätystä datasta (Abrevaya & McCulloch, 2014; Schuckers & Brozowski, 2012) kuin myös jalkapallo-otteluiden aluissa varoitusten osalta (Unkelbach & Memmert, 2008).

Erien viimeisillä minuuteilla esiintyvät poikkeuksellisen korkeat jäähymäärät myötäilevät Abrevayan ja McCullochin (2014) minuuttijakaumaa NHL:n rangaistuksista ja he ehdottavat, että pelaajat ottavat enemmän rangaistuksia kolmansien erien lopuissa tilanteissa, joissa peli on jo ratkennut eikä jäähy vaikuta merkityksellisesti lopputulokseen. Tämä ei kuitenkaan selitä ensimmäisen ja toisen erän loppujen kohonneita jäähymääriä. Mahdollisia syitä tälle ovat väsyneiden pelaajien tekemät virheet erien lopuissa tai joukkueiden pyrkimys puolustaa joukkueen johtoasemaa sääntöjen rajamailla olevin keinoin niin, ettei vastustaja onnistu tasoittamaan ottelua juuri ennen ottelun loppua tai tauolle siirtymistä.

3 Tutkimusmenetelmä

Tämän tutkielman tutkimuskysymys pohjautuu Beaudoinin ja muiden (2016) artikkeliin, jossa he pyrkivät logistista regressiota ja koneoppimismenetelmää käyttäen selittämään muuttujan $y = 1(0)$, eli annetaanko jäähy koti- vai vierasjoukkueelle. He hyödyntävät otteludatasta kerättyjä jäähyjen kontekstimuuttujia x tarkastelemaan, mitkä muuttujat vaikuttavat eniten jäähyjen jakautumiseen. Siinä missä Beaudoin ja muut sisällyttävät malleihinsa vain neljää muuttujaa (jäähysuhde, maalisuhde, jäähyn aika ja joukkueen paremmuusindikaattori), yhdistän näihin tutkielmassani Abrevayan ja McCullochin (2014) ”käänteisjäähyjen” mallintamiseen käyttämiä muuttujia pienin erotuksin. Käytän datan mallintamiseen neljää koneoppimisalgoritmia – päätöspuita, gradient boostingia, satunnaismetsiä sekä neuroverkkoja – ja vertailen näitä keskenään parhaan mallin valitsemiseksi. Sisällytän mukaan myös logistisen regressiomallin verrokkina koneoppimismallien soveltuvuudelle.

Mallintamisessa käytetyt muuttujat esitetään kokonaisuudessaan taulukossa 1. Korvaan peräkkäisyyttä ilmaisevat osoitinmuuttujat yksittäisellä *inRow*-muuttujalla vähentämään mallin kompleksisuutta ja kauden päivää ilmaisevan muuttujan sijasta käytän kuukausimuuttujaa, koska tämä kuvastaa paremmin mahdollisesti muuttuvaa tuomarilinjaa runkosarjan lähentyessä loppuaan ja pudotuspelien alkua. Joukkueen vahvuusindikaattori korvautuu joukkueiden hyökkäys- ja puolustuskykyä sekä jäähyjen ottamista ja hankkimista kontrolloivilla muuttujilla. Sisällytän malleihin myös Lopezin ja Snyderin (2013) tutkimukseen perustuen *capacity*-muuttujan kuvaamaan mahdollista täyden kotiyleisön lisävaikutusta tuomarin päätöksiin.

Taulukko 1. Tutkimuksessa käytetyt muuttujat.

Muuttuja	Muuttujan kuvaus
<i>Riippuva muuttuja</i>	
home	Kumpi joukkue on saanut jäähyn (1=koti, 0=vieras)
<i>Muut muuttujat</i>	
capacity	Areenan täyttöaste (1=95% tai yli paikoista täynnä, 0=alle 95% täynnä)
ppgoal	Johtiko edellinen ylivoima maaliin (1=kyllä, 0=ei)
season	Millä kaudella ottelu pelattiin (esim. 2014–2015)
numpen	Jäähyn järjestysluku ottelussa
time	Milloin jäähy on annettu ottelussa
timebetpens	Aikaväli edelliseen jäähyyn
month	Kuinka monta kuukautta kaudesta on pelattu
goaldiff	Ottelun maaliero (positiivinen jos kotijoukkue johtaa)
pendiff	Ottelun jäähyero (positiivinen jos kotijoukkueella enemmän jäähyjä)
inRow	Peräkkäisjäähyjen määrä (positiivinen jos edellinen jäähy kotijoukkueelle)
gf1	Edellisen jäähyn saaneen joukkueen tehdyt maalit per ottelu kaudella
ga1	Edellisen jäähyn saaneen joukkueen päästetyt maalit per ottelu kaudella
pf1	Edellisen jäähyn saaneen joukkueen alivoimat per ottelu kaudella
pa1	Edellisen jäähyn saaneen joukkueen ylivoimat per ottelu kaudella
gf2	Vastakkaisen joukkueen tehdyt maalit per ottelu kaudella
ga2	Vastakkaisen joukkueen päästetyt maalit per ottelu kaudella
pf2	Vastakkaisen joukkueen alivoimat per ottelu kaudella
pa2	Vastakkaisen joukkueen ylivoimat per ottelu kaudella

3.1 Koneoppiminen

Koneoppiminen on tietojenkäsittelytieteiden osa-alue, joka käsittelee datasta tai aikaisemmasta kokemuksesta oppivia ja sen perusteella toimintaansa optimoivia algoritmeja ja malleja. Näillä menetelmillä luotuja malleja voidaan käyttää uuden tiedon muodostamiseen tekemällä ennustuksia tai päättelemällä sitä jo olemassa olevasta datasta. (Alpaydin, 2014)

Koneoppimisalgoritmit voidaan jakaa karkeasti ohjattuun, ohjaamattomaan ja vahvistettuun oppimiseen. Ohjatussa oppimisessa mallin opettamiseen käytetyn datan ennustettavat muuttujat ovat tiedossa ja ohjaamattomassa oppimisessa niitä ei tiedetä. Vahvistetussa oppimisessa malli pyrkii löytämään optimaalisia arvoja vuorovaikuttamalla ympäristönsä kanssa ja muuttamalla toimintaansa vasteen perusteella. (Bishop, 2006)

Koneoppiminen mainittiin terminä ensimmäisen kerran Arthur Samuelin toimesta jo vuonna 1959 (Samuel, 1959), mutta se alkoi saavuttamaan varsinaista suosiota ja tunnustusta tieteenalana vasta 1990-luvulla paradigman muuttuessa tekoälymallinnuksesta konkreettisten ongelmien ratkaisemiseen (Langley, 2011). Nykypäivänä koneoppimismenetelmiä sovelletaan laaja-alaisesti dataan liittyvissä ongelmissa niin tietojenkäsittelytieteissä kuin muillakin tieteenaloilla. Tavallisia käyttökohteita ovat esimerkiksi tiedonlouhinta, suosittelijajärjestelmät sekä konenäön ja luonnollisen kielen käsittelyn kaltaiset tunnistustehtävissä. (Alpaydin, 2014)

Koneoppimisalgoritmien vahvuus ilmenee niiden kyvyssä mallintaa tehokkaasti suuria määriä dataa löytämällä muuttujien välisiä ei-parametrisiä suhteita tavallisimpiin data-analyysissä käytettäviin lineaarisiin malleihin verrattuna. Koneoppimismallit saattavat kuitenkin olla käytetystä menetelmästä riippuen hankalasti tulkittavia tai kompleksisuudeltaan laskennallisesti vaativia ja aikaavieviä. (Hastie ja muut, 2009)

Mallien sopivuutta arvioidaan tavallisesti jakamalla mallinnettava data harjoitus-, validointi- ja testausjoukkoihin. Mallit opetetaan harjoitusjoukon perusteella ja keskinäinen vertailu suoritetaan sovittamalla ne validointijoukkoon. Parhaaksi valikoituva malli sovitetaan lopulta testausjoukkoon osoittamaan sen yleistyvyttä eli kykyä sovittua uuteen dataan. Suurimmassa osassa tapauksista tarkasteltavana olevaa dataa ei kuitenkaan ole riittävästi tarpeeksi suurien joukkojen muodostamiseen, joten mallin valinta suoritetaan ristivalidointimenetelmällä. Ristivalidoinnissa harjoitusjoukko jaetaan satunnaisesti samankokoisiin osiin, joista yksi eriytetään validointijoukoksi. Malli opetetaan ja sovitetaan tätä menetelmää toistaen, kunnes kaikki osat ovat toimineet validointijoukkona ja sopivuus määritetään tulosten keskiarvon perusteella. (Alpaydin, 2014)

Sopivuuden mittaamiseen käytetään yleisimmin tappiofunktioita, jotka kuvaavat mallin tuloksen ja todellisen arvon välistä virhettä. Jos mallin virhe on pieni

harjoitusjoukossa, mutta suuri testausjoukossa, on kyse ylisovittumisesta, joka osoittaa heikkoa yleistyvyyttä ja harjoitusjoukon piirteiden liiallista oppimista. Suuri virhe molemmissa joukoissa kuvaa alisovittumista ja huonoa yleistyvyyttä mallin matalan kompleksisuuden takia. (Hastie ja muut, 2009)

Mallin valinnassa parhaan sopivuuden saavuttamiseen vaaditaan usein optimaalisten hyperparametrien löytämistä. Nämä parametrit määrittävät mallin kompleksisuuden ja huonot yhdistelmät johtavat usein yli- tai alisovittumiseen. Hyperparametreiksi voidaan luokitella kaikki mallin algoritmin käyttämät parametrit, joita ei pystytä laskennallisesti pääättelemään tai arvioimaan käytetyn datan perusteella. (Kuhn & Johnson, 2013)

Kaikki mallit toteutetaan hyödyntäen R-ohjelmistoa ja sen kirjastoja *rpart*, *gbm*, *rForest*, *nnet*, ja *glm* (R Core Team, 2013).

3.1.1 Päättöspuut

Päättöspuut ovat luokittelu- ja regressiopuihin jaettavia binääripuita, jotka osittavat dataa pienempiin ja homogeenisempiin ryhmiin luokkien tai riippuvan muuttujan perusteella. Puut rakennetaan käyttäen rekursiivista algoritmia, joka muodostaa lapsisolmun parhaimmalle osituskohdalle jakokriteerin perusteella. Luokittelupuissa jakokriteeri määritetään solmun puhtaudella eli vähäisimmällä luokitteluvirheellä tai Gini-kriteerillä tunnetulla jakauman poikkeavuudella. Regressiupuissa data jaetaan riippumattomien muuttujien yhteenlasketun neliövirheen perusteella. Riippuva muuttuja selitetään kulkemalla reitti juuresta johonkin lehtisolmuun asti. Päättöspuiden suurimmat edut ovat niiden helppo toteutus, mallien tulkittavuus ja kyky sopeutua dataan. Ne eivät kuitenkaan usein yllä parhaaseen mahdolliseen yleistyvyyteen ja päättöspuumallien rakenne saattaa muuttua merkittävästi, jos datassa tapahtuu vähäisiäkin muutoksia. (Hastie ja muut, 2009; Kuhn & Johnson, 2013)

3.1.2 Gradient boosting

Gradient boosting (Friedman, 2001) on oppimisen tehostamismenetelmä, jossa riippuvan muuttujan selittävä malli muodostetaan useasta heikosta oppijasta (yleensä regressiopäättöspuu) virhefunktion perusteella. Tehostettu malli rakennetaan gradienttimenetelmää (*gradient descent*) käyttäen luomalla heikkoja oppijoita ja lisäämällä niitä siihen, jos ne parantavat sopivuutta eli vähentävät mallin virhettä. Gradient boosting -mallien kyky sovittua on erinomainen ja useat hyperparametrit mahdollistavat niiden muokkaamisen monenlaiseen datajoukkoon sopivaksi. Vähäiseen virheeseen pyrkivä toimintaperiaate tekee gradient boosting -malleista kuitenkin taipuvaisia ylisovittumiseen ja oikeiden hyperparametrien löytäminen sekä monimutkaisten mallien muodostaminen ovat mahdollisesti aikaavieviä operaatioita. (Kuhn & Johnson, 2013)

3.1.3 Satunnaismetsä

Satunnaismetsät (Breiman, 2001) ovat kokoelma toisistaan erillisiä päätöspuumalleja, jotka muodostetaan satunnaisotannalla käytetystä datasta kuitenkin niin, että jokaisessa otanta jakautuu samalla tavalla. Päätöspuut valitaan niiden sopivuuden perusteella ja satunnaismetsä selittää riippuvan muuttujan sen sisältämien päätöspuiden tulosten keskiarvon perusteella niin luokittelussa kuin regressiossa. Hastien ja muiden mukaan (2009) satunnaismetsien rakenne suojaa niitä jossakin määrin ylisovittumiselta ja ne pystyvät saavuttamaan hyviä tuloksia, mutta samanlainen heikkoja oppijoita yhdistelevä gradient boosting -menetelmä pystyy heidän esittämiensä vertailujen perusteella sovitumaan yleisesti ottaen paremmin erilaisiin datajoukkoihin. Vähäisestä hyperparametrien määrästä johtuen satunnaismetsät ovat myös yksinkertaisia käyttää.

3.1.4 Neuroverkot

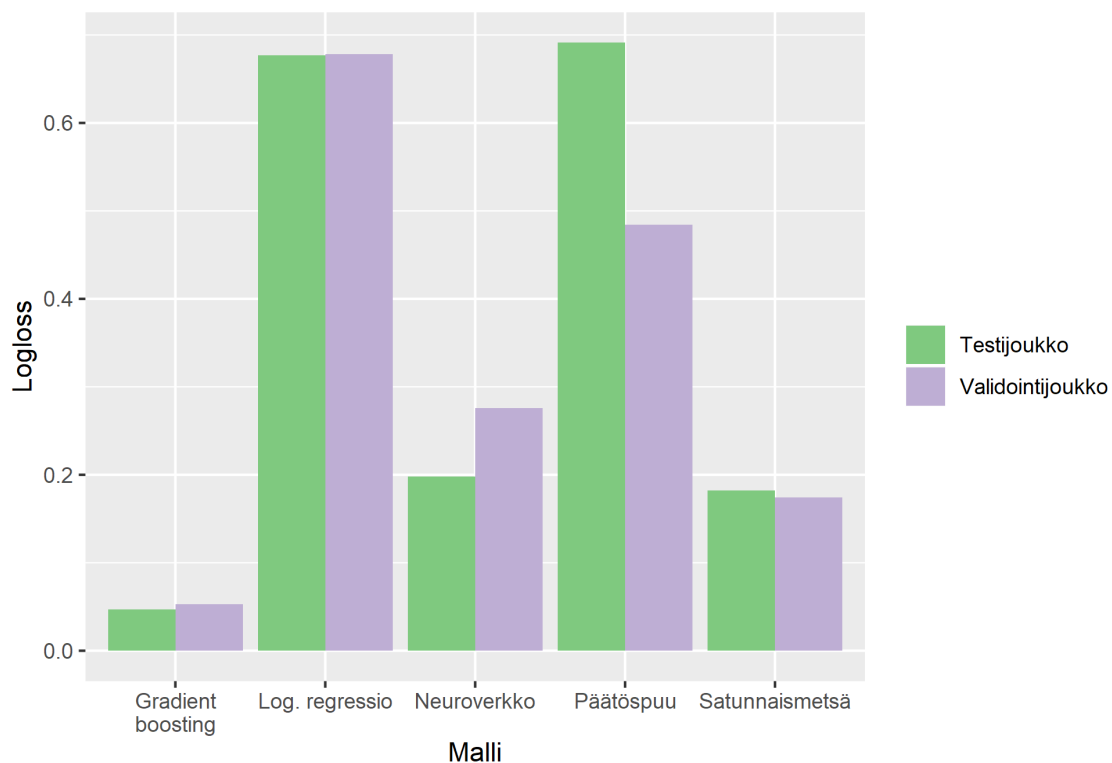
Neuroverkot ovat ihmisaivojen toimintaan perustuvia, epälineaarisia luokitteluun tai regressioon käytettyjä menetelmiä, joissa riippuva muuttuja mallinnetaan käyttämällä välikätenä piilotetuksi kerrokseksi (*hidden layer*) nimitettyä kokoelmaa tarkastelemattomista muuttujajoukkojen yksiköistä (Alpaydin, 2014). Piilotettuja kerroksia voi olla useampia ja niiden yksiköt voivat sisältää minkä tahansa yhdistelmän riippumattomia muuttujia, joihin on sovellettu epälineaarista funktiota (esim. logistinen funktio). Riippuva muuttuja selitetään lopulta näiden yksikköjen lineaarisella yhdistelmällä. Neuroverkot ovat tehokkaita, mutta niillä on taipumus ylisovittua ja optimaalisen hyperparametrien yhdistelmän löytäminen saattaa osoittautua hankalaksi (Kuhn & Johnson, 2013).

3.2 Mallin valinta

Jäähydata jaetaan ensin harjoitus- ja testausjoukkoon jakaumalla 80 %–20 %. Aikaisemmin kuvailtuja malleja sovitetaan harjoitusjoukkoon eri hyperparametriyhdistelmin ja lopulliseen vertailuun valitaan parhaan tuloksen 10-kertaisessa ja kolmesti toistetussa ristivalidoinnissa saavuttanut malli. Ristivalidoinnin määrä on valittu Kohavin (1995) tulosten perusteella ja toistaminen toteutetaan luotettavamman arvion saamiseksi (Alpaydin, 2014). Parhaat mallit sovitetaan lopulta testausjoukkoon yleistyvyyden tarkastelemiseksi. Virheen arviointimenetelmänä käytetään binäärimuuttujiin yleisesti sovellettua logistista tappiofunktiota (*log loss*). Logistisen tappiofunktion virhe on kasvaa mallin luokitellessa väärin tapauksia, joista se on varma. (Buja ja muut, 2005)

Parhaimmat tulokset saavutettiin gradient boosting -mallissa hyperparametrien yhdistelmällä $n.trees = 1500$ (puiden määrä), $interaction.depth = 8$ (puiden haarautumisten määrä) ja $shrinkage = 0,1$ (puiden kutustumiskerroin); satunnaismetsässä $mtry = 17$ (käytettävien muuttujien määrä puiden haarautumisessa); neuroverkossa $size$

= 6 (piilotetun kerroksen yksikköjen määrä), $decay = 0,1$ (painoheikennys) ja päätöspuussa $cp = 0,007653061$ (puun kompleksisuus). Lopullisten mallien virheet validointi- ja testausjoukossa ovat nähtävissä kuvassa 1. Pienimmän virheen molemmissa joukoissa saavutti lopulliseen analyysiin valikoitunut gradient boosting -malli (testaus 0,053 ja validointi 0,047) ja hyviin tuloksiin pääsivät myös satunnaismetsä (0,174 ja 0,182) sekä neuroverkko (0,276 ja 0,198). Yksinkertainen päätöspuu (0,484 ja 0,691) pärjasi validoinnissa paremmin kuin logistinen regressiomalli (0,678 ja 0,677), mutta yleistyi yksinkertaisen rakenteensa takia odotetusti huonommin (Kuhn & Johnson, 2013). Logistisessa regressiomallissa ei erikseen määritelty parametrisiä suhteita muuttujien välille, mikä olisi todennäköisesti parantanut mallin tulosta.



Kuva 1. Mallien log loss -virheet testi- ja validointijoukoissa.

Tulokset myötäilevät Abrevayan ja McCullochin (2014) tekemää vertailua eri koneoppimismenetelmien soveltumisesta jäähydatan analysointiin gradient boosting-mallin ja satunnaismetsän pärjätessä hyvin. Hastien ja muiden (2009) sekä Kuhnin ja Johnsonin (2013) esittämät vertailut koneoppimismenetelmien tehokkuudesta eri datajoukkoihin sovitettuina vahvistavat myös käsitystä siitä, että nämä kaksi menetelmää ja neuroverkot kykenevät sovittumaan hyvin monenlaiseen dataan.

4 Tulokset

Koska koneoppimismallien tulokset ovat usein vaikeaselkoisia ja niiden tulkitseminen on harvoin yksiselitteistä, käytän tässä tutkielmassa useita eri tapoja arvioimaan muuttujien merkityksestä lopputulokseen. Tarkastelen mallin sisäistä metriikkaa muuttujien suhteellisesta vaikutuksesta ja vertailen muuttujien yhteisvaikutusta laskemalla mallin ennustusten keskiarvon niissä tapauksissa, joissa yhdelle muuttujalle annetaan tietty arvo sekä muuttujien osittaisvaikutusta Friedmanin (2001) kuvailemalla tavalla, missä mallin ennustusten keskiarvo lasketaan tietyn muuttujan arvoille jokaisen muun muuttujan yhdistelmän kanssa.

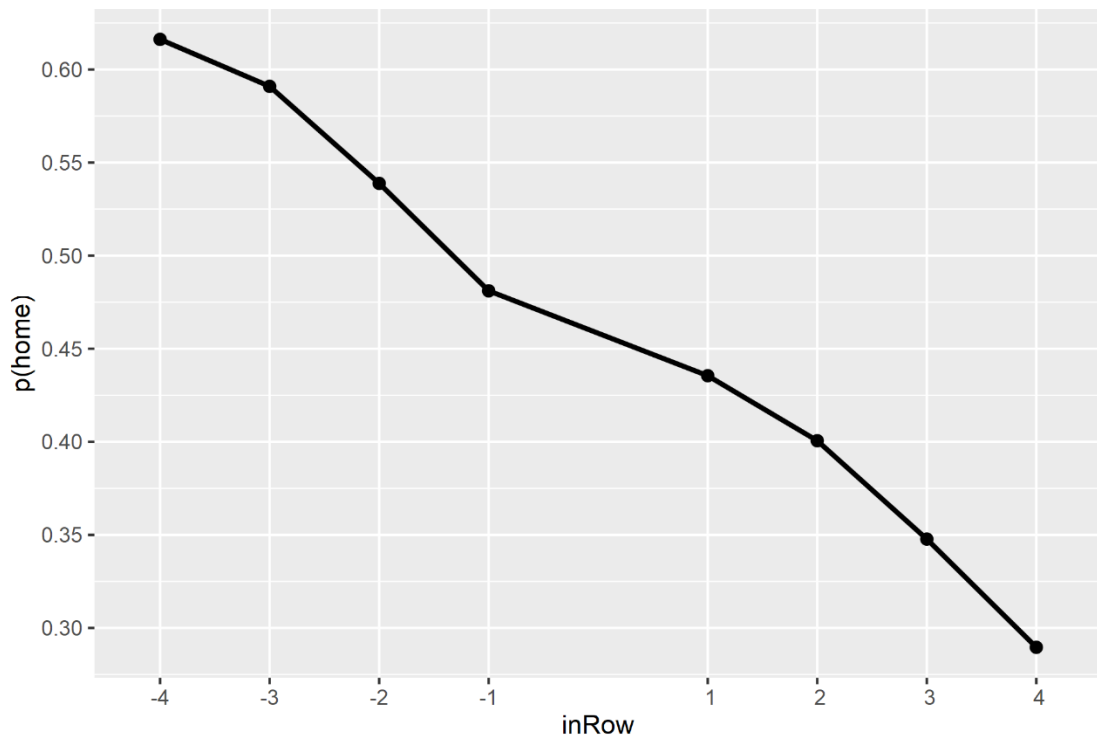
Taulukossa 2 esitetään kohdassa 3.2 parhaaksi valikoituneen gradient boosting -mallin muuttujien suhteellinen vaikutus (*relative influence*). Suhteellinen vaikutus kuvaa sitä, kuinka suuri osuus muuttujan käyttämisessä puumallin jakokriteerinä on sen virheen määrän vähentämisessä (Ridgeway, 2007). Kaikki mallissa käytetyt muuttujat olivat merkittäviä ja etenkin peräkkäisiä jäähyjä kuvaava *inRow*-muuttujan vaikutus malliin oli suuri. Myös joukkueiden tehtyjä ja päästettyjä maaleja sekä yli- ja alivoimien määrää kuvaavat muuttujat vaikuttivat malliin merkityksellisesti, mikä on odotettavaa näiden korreloidessa suoraan joukkueiden paremmuuden ja jäähyjen määrän kanssa.

Taulukko 2. Gradient boosting -mallin muuttujien suhteellinen vaikutus.

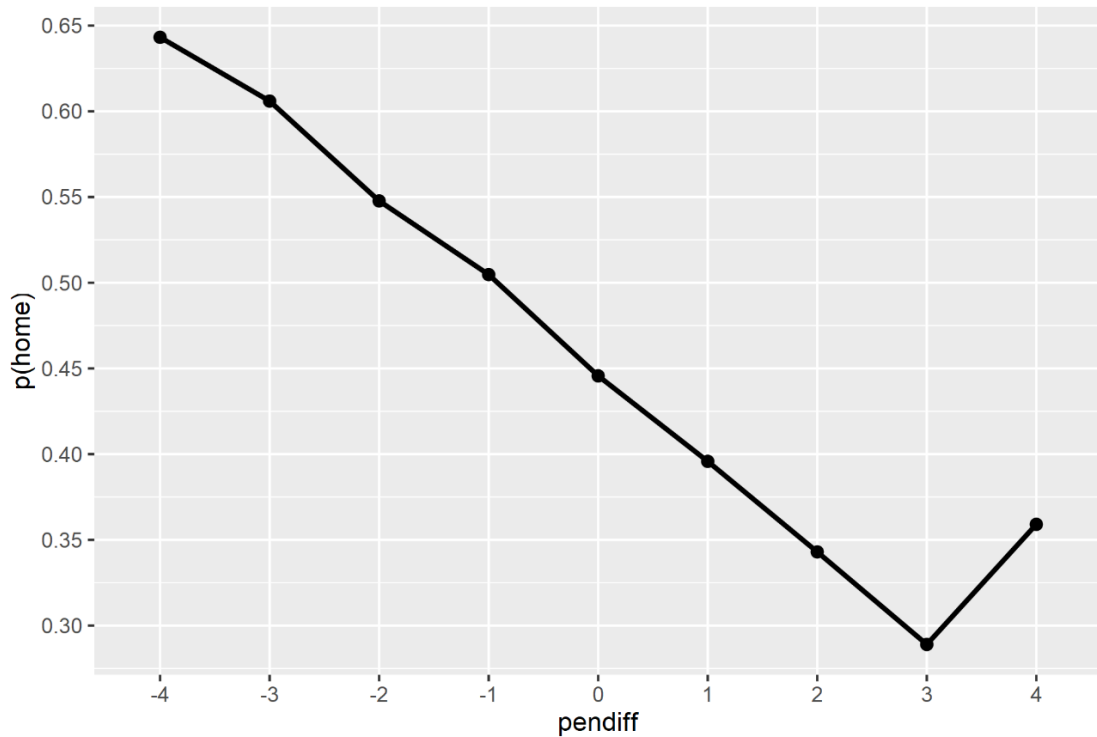
Muuttujan nimi	Suhteellinen vaikutus (%)
inRow	37,17
pa2	7,54
ga2	7,52
pa1	7,44
gf1	6,59
gf2	6,34
pf2	5,94
ga1	5,00
pf1	4,57
pendiff	2,95
season	2,17
time	2,07
timebetpens	2,07
goaldiff	1,40
month	0,48
numpen	0,46
capacity	0,22
ppgoal	0,06

Tulokset myötäilevät olemassaolevan kirjallisuuden havaintoja aikaisemman tuomaritoiminnan vaikutuksesta tuleviin päätöksiin ottelun sisällä. Saman joukkueen peräkkäisten jäähyjen ennustavuus seuraavan jäähyn antamisesta vastajoukkueelle on osoitettu Abrevayan ja McCullochin (2014) tutkimuksessa ja myös jalkapallossa ottelutapahtumien kontekstin (Unkelbach & Memmert, 2008) sekä aikaisemmin annettujen varoitusten määrän ja viimeaikaisuuden (Buraimo ja muut, 2010) on havaittu vaikuttavan tuomarin tuleviin päätöksiin. Myös Beaudoin ja muut (2016) sekä Lopez ja Snyder (2013) ovat löytäneet näyttöä tuomarien tasoittelevan jäähyjen suhdetta otteluiden sisällä. On kuitenkin huomioitava, että Beaudoinin ja muiden käyttämässä yksinkertaisessa gradient boosting -mallissa jäähysuhteen (*pendiff*) suhteellinen vaikutus oli suurin, mutta käyttämäni mallin perusteella peräkkäisjäähyt (*inRow*) näyttäisi ennustavan tuomarien päätöksiä herkemmin koko ottelun jäähysuhteeseen verrattuna.

Muuttujien yhteisvaikutusta tarkastellessa selviä trendejä oli havaittavissa muuttujien *inRow*, *penDiff* ja *goaldiff* vaihtelussa. Kuvissa 2–4 on esitetty pystyakselilla mallin riippuvalle muuttujalle (*home*) antamien tulosten keskiarvo silloin, kun muuttujalle annetaan vaaka-akselilla oleva arvo. Edellä kuvatut tutkimustulokset peräkkäisjäähyjen ja jäähysuhteen vaikutuksesta tuomaritoimintaan ja jäähyjen jakautumisen todennäköisyyteen on huomattavissa kuvissa 2 ja 3 vielä selkeämmin kuin aikaisemmin. Mitä enemmän jäähyjä joukkue on ottanut peräkkäin tai saanut niitä toiseen joukkueeseen verrattuna, sitä todennäköisempää on seuraavan jäähyn viheltäminen vastakkaiselle joukkueelle. Vaikutus on myös suurempi kotijoukkueen hyödyn näkökulmasta tarkasteltuna ja neljän perättäisen kotijoukkueelle annetun jäähyn jälkeen seuraavan jäähyn ennustetaan jaettavan vierasjoukkueelle jopa yli 70 % ajasta. Myös eräkohtaisen jäähysuhteen vaikutusta on tarkasteltu Lopezin ja Snyderin (2013) toimesta, mutta tässä mallissa sen jakaminen eriin ei tuottanut parempia tuloksia.



Kuva 2. Muuttujien yhteisvaikutus *inRow*-muuttujan eri arvoilla.

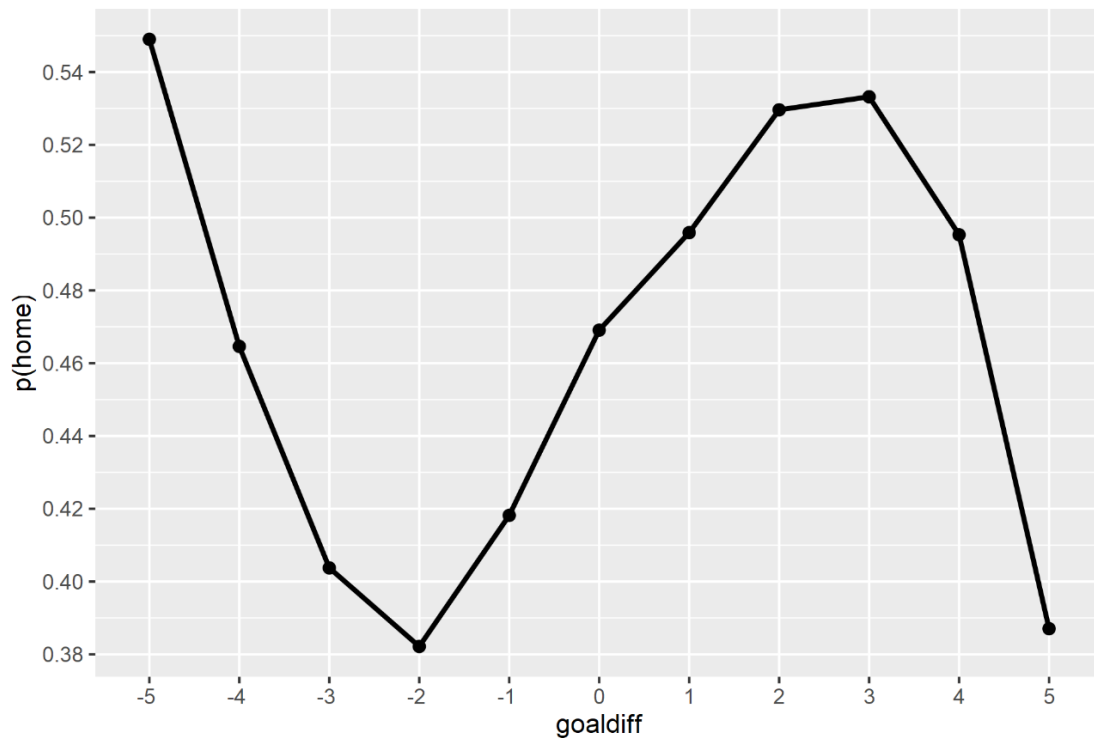


Kuva 3. Muuttujien yhteisvaikutus *pendiff*-muuttujan eri arvoilla.

Kuvasta 4 voidaan huomata, että johdossa olevilla joukkueilla on kasvava todennäköisyys saada jäähy aina kolmen maalin johtoon saakka, jonka jälkeen suunta muuttuu laskevaksi. Kuten peräkkäisjäähyjen ja jäähysuhteen osalta, maalisuhteen

(*goaldiff*) vaikutus on jälleen suurempi kotijoukkueen hyödyksi. Pelitilanteen vaikutuksesta tuomaritoimintaan löytyy näyttöä Schuckersiin ja Brozowskiin (2012) sekä Abrevayan ja McCullochin (2014) tuloksista, jotka kuvaavat samanlaista maalisuhteen vaikutusta. Jalkapallon osalta Dohmen (2008) havaitsi, että tuomarit hyväksyvät tappiolla olevien joukkueiden väärin tehtyjä maaleja suuremmalla todennäköisyydellä. Tuomarit antavat myös enemmän lisäaikaa otteluissa, joissa kotijoukkue on niukasti tappiolla tai tilanteen ollessa tasan (Sutter & Kochera, 2004).

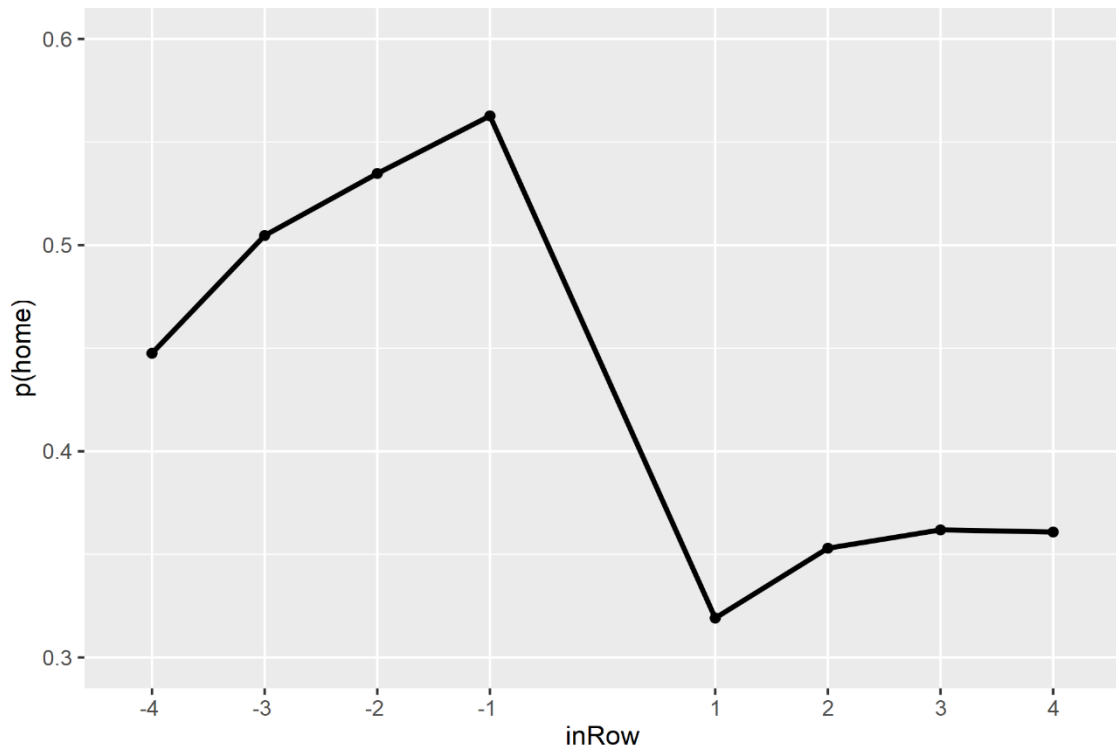
Maalisuhteen vaikutusta jäähyjakaumaan voi mahdollisesti selittää joukkueiden pelitilanteesta riippuvalla pelityylillä. Olisi loogista olettaa, että johtoasemaa puolustava joukkue pelaisi konservatiivisemmin ja ottaisi enemmän pelin kulkua rikkovia ja vaarallisia tilanteita estäviä rangaistuksia. Toisaalta, tappiolla olevien joukkueiden voisi vastavuoroisesti olettaa ottavan enemmän riskejä ja pelaavan aggressiivisemmin, hankkien näin enemmän rangaistuksia. Buraimo ja muut (2010) osoittivat, että jalkapallossa tappiolla olevat joukkueet hankkivat enemmän varoituksia yrittäessään tasoittaa ottelua. Beaudoinin ja muiden (2016) NHL-otteluista keräämän datan perusteella tappioasemassa olevat joukkueet laukovat keskimäärin enemmän, mikä puoltaisi niin johtavan joukkueen vetäytymistä kuin tappiolla olevan joukkueen riskinottoa. Mahdollista on myös se, että tuomarit antavat useammin rangaistuksia johdossa olevalle joukkueelle tasapuolisuuden vaikutelman luomiseksi etenkin, jos tappiolla oleva joukkue pelaa kotonaan äänekkäästi tuomareita painostavan yleisön edessä.



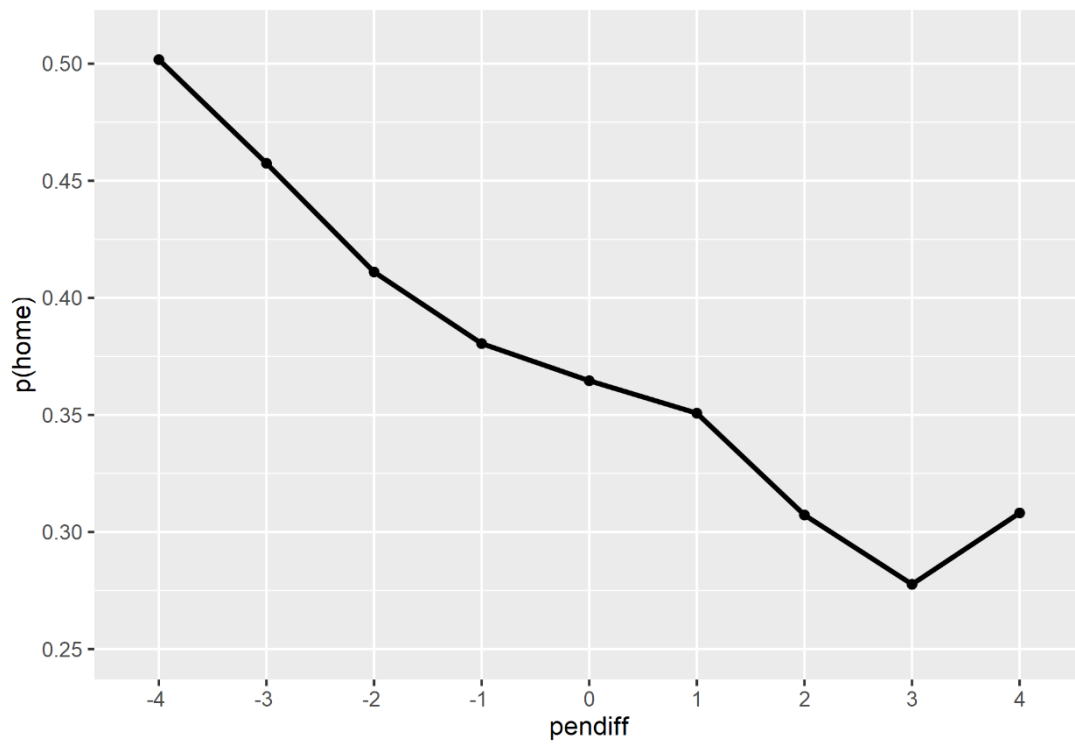
Kuva 4. Muuttujien yhteisvaikutus *goaldiff*-muuttujan eri arvoilla.

Tarkastellessa merkittävimmän yhteisvaikutuksen omaavien muuttujien *inRow*, *pendiff* ja *goaldiff* osittaisvaikutusta (kuvat 5–7) voidaan huomata, että todennäköisyys vierasjoukkueen saamaan jäähyyn kasvaa huomattavasti muiden muuttujien vaikutuksen marginalisoiduttua ja vierasjoukkueen jäähy on todennäköisempi jopa tilanteissa, joissa muuttujien yhteisvaikutus kallistui kotijäähyyn. Tämä viittaisi siihen, että neutraalissa ottelukontekstissa ja joukkueiden ollessa täysin tasapuolisia niin pelillisesti kuin jäähyjen hankkimisen osalta, on kotietu jäähyjen jakautumisessa vielä suurempi kuin mitä aikaisemmin kävi ilmi. Sama trendi on havaittavissa myös jokaisessa muussa mallissa käytetyn muuttujan osittaisvaikutuksessa.

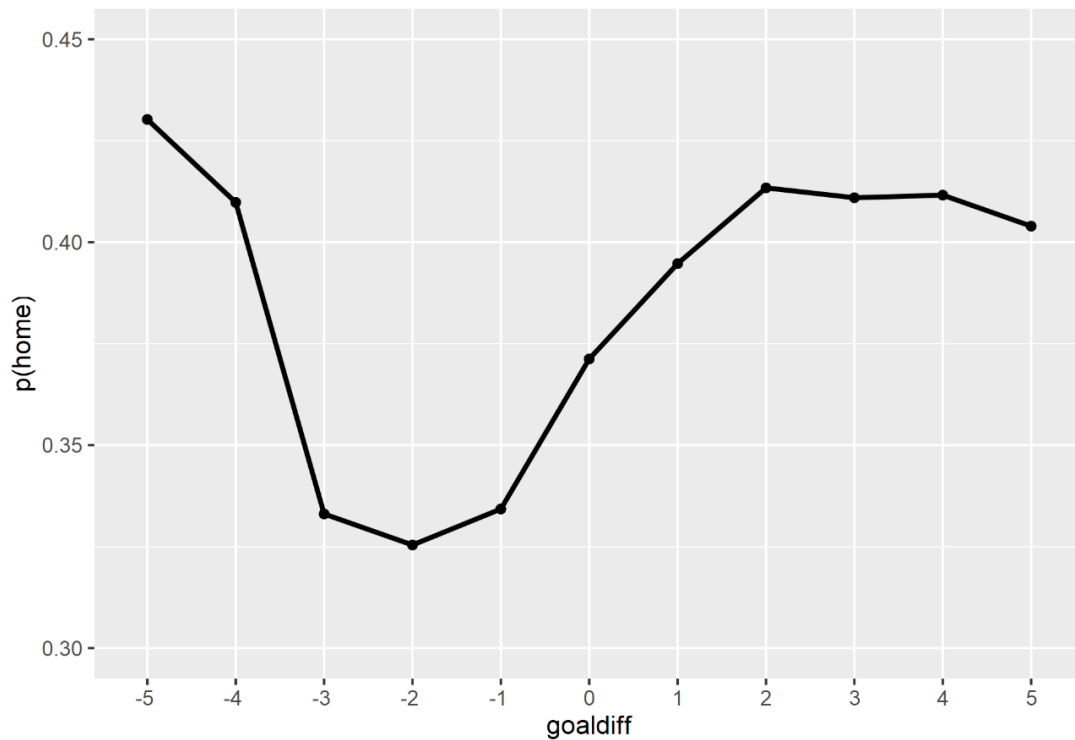
Toinen merkille pantava asia on se, että siinä missä muuttujat *pendiff* ja *goaldiff* (kuvat 6 ja 7) pitkälti säilyttävät yhteisvaikutuksessa havaittavan trendinsä, poikkeaa *inRow*-muuttujan osittaisvaikutus (kuva 5) siitä selvästi. Peräkkäisjäähyjen merkitys jäähyjakaumaan on vieläkin suuri, mutta se näyttää laskevan ensimmäisen jäähyyn jälkeen aikaisemmin havaitun kasvun sijaan. Yksi mahdollinen selitys tälle on se, että *inRow* ja *pendiff* ovat toisistaan lineaarisesti riippuvaisia, ja osittaisvaikutusta arvioiva menetelmä ei kykene eristämään tarpeeksi hyvin keskenään vahvasti riippuvaisten muuttujien vaikutusta (Friedman, 2001). Tällöin voisi olettaa myös *pendiff*-muuttujan trendin muuttuvan, mutta koska se kuvaa koko ottelun jäähytasapainoa ja vaihtelee täten vähemmän jäähyjen välillä, voidaan olettaa sen vaikutuksen olevan vähemmän riippuvainen jyrkemmin muuttuvasta, peräkkäisjäähyjä kuvaavasta *inRow*-muuttujasta. On tietenkin täysin mahdollista, että peräkkäisyys ei vaikuta tuomarien toimintaan niinkään määränsä osalta, vaan pelkästään edellisen jäähyyn saaja jäisi tuomarien mieleen. Tämä vaikuttaa kuitenkin epätodennäköiseltä *pendiff*-muuttujan säilyttäessä trendinsä, mikä kuvastaa tuomarien huomioivan ottelun jäähytasapainon tarkemmin ainakin jollain tasolla.



Kuva 5. Muuttujan *inRow* osittaisvaikutus.



Kuva 6. Muuttujan *pendiff* osittaisvaikutus.



Kuva 7. Muuttujan *goaldiff* osittaisvaikutus.

Muiden muuttujien yhteis- ja osittaisvaikutusten osalta ei voitu osoittaa selviä tai tarpeeksi merkittäviä trendejä, mikä viittaisi niiden olevan enemmän riippuvaisia toisten muuttujien arvoista. Täyttöastetta ilmaiseva muuttuja *capacity* ei onnistunut todentamaan teoriaa vahvemmassa kotiedusta otteluissa, joissa jäähalli on täynnä tai lähes täynnä, mikä antaisi olettaa yleisön onnistuvan luomaan suurempaa painetta tuomareita kohtaan. Kotiedun on kuitenkin havaittu jalkapallossa riippuvan niin yleisömäärästä (Boyko ja muut, 2007; Page & Page, 2010) kuin yleisön etäisyydestä kenttään (Buraimo ja muut, 2010), joten saman vaikutuksen voisi olettaa olevan läsnä jääkiekossakin. Tämä voi selittyä yksinkertaisesti sillä, että yksittäinen muuttuja ei ole riittävä kuvaamaan yleisön ja kotiedun tai tuomarien välistä suhdetta.

5 Keskustelu

Tämän tutkielman tulosten perusteella voidaan osoittaa, että monelle suomalaisen liigajääkiekkoilun tuomaritoimintaan yhdistetylle kliseelle on olemassa totuuspohja. Tulkitsemalla tutkielmassa käytettyä dataa ja siihen sovitettua koneoppimismallia löydetään näyttöä kotiedun, ottelun ajan, peräkkäisten jäähyjen, jäähysuhteen ja ottelun tilanteen vaikuttavan jäähyjen määrään ja jakaumaan merkittävällä tavalla. Tulokset ovat enimmäkseen yhteneväisiä aikaisemman tuomaritoiminnan kaavamaisuutta käsittelevän tutkimuskirjallisuuden kanssa. Merkittävyys muodostuu kuitenkin siitä, että tämä tutkielma on tiettävästi ensimmäinen, joka käsittelee suomalaisen joukkueurheilulajin tuomarien ennakoitavuuksia ja tuomaritoiminnasta aiheutuvaa kotietua. Edeltävien

jääkiekkosarjojen tuomaritoimintaa käsitelleiden tutkielmien kohteena on ollut yleensä NHL, jossa joukkueiden pelityyli ja yleinen toimintakulttuuri poikkeaa eurooppalaisesta, joten tämä tutkielma luo uuden näkökulman myös tältä osin.

Johdannossa esitetty ja aikaisemmissa tutkimuksissa osoitettu yleisömäärän vaikutus tuomareista aiheutuvaan kotietuun muodostumiseen oli ainoa merkittävä kaavamaisuus, jota tässä tutkielmassa ei onnistuttu osoittamaan. Boyko ja muut (2007) sekä Page ja Page (2010) löysivät näyttöä siitä, että suurempi yleisömäärä johti tuomarin suosimaan päätöksillään kotijoukkuetta keskimäärin enemmän ja havaitsivat yksittäisten tuomarien paineensietokyvyn vaihtelevan merkittävästi yksilöiden välillä. On epätodennäköistä, että jääkiekkotuomarit poikkeaisivat tältä osin, joten tulevien aiheeseen liittyvien tutkimusten menetelmien tulisi ottaa huomioon useampia muuttujia kuvaamaan yleisön vaikutusta. Tutkielmassa käytetyssä mallissa ei huomioitu tuomariparien yksilöllisiä vaikutusta ylipäättään ollenkaan, sillä niiden sisällyttäminen heikensi mallin tulosta. Beaudoin ja muut (2016) esittivät käyttämälleen mallilleen mielenkiintoisen bootstrap-menetelmän arvioimaan tuomarien yksittäisvaikutusta arvottamalla heidät sen perusteella, kuinka lähelle ”ideaalia” tasajakaumaa tuomarin viheltämät jäähyt pääsivät. Menetelmää voisi käyttää jatkotutkimuksessa yksittäisten tuomarien tarkastelemiseen, joskin tasajakauma on tuskin paras mittari tuomarien suoriutumiseksi tiettyjen joukkueiden hankkiessa enemmän jäähyjä kuin toiset.

Koska osa esitetyistä tuloksista voitaisiin tuomaritoiminnan sijasta perustella joukkueiden pelityyleillä ja niiden muutoksilla ottelun eri konteksteissa, on hyvä tarkastella pelityyliä merkitystä jäähyjakaumaan. On luonnollista, että tietyt joukkueet ja pelaajat hankkivat pelitavallaan enemmän rangaistuksia joko itselleen tai vastustajalleen. Pelkästään maineen aggressiivisesti pelaavasta joukkueesta on osoitettu vaikuttavan tuomarien päätöksiin jalkapallossa (Jones ja muut, 2002). Joukkueet pelaavat myös eri tavalla ollessaan voitolla tai tappiolla, kuten aikaisemmin maalisuhteen vaikutusta jäähyjakaumaan selitettäessä osoitettiin. Voidaanko mahdollisesti muitakin tutkielman osoittamista kaavamaisuuksista perustella siis pelkillä pelitavallisilla eroilla? Lopez ja Snyder (2013) esittävät, että tämä on epätodennäköistä, sillä pelityyliin muutokset kompensoivat toisiaan: vaikka puolustava joukkue hankkii keskimäärin enemmän jäähyjä, hankkii myös tasoitusta tavoitteleva joukkue enemmän jäähyjä pelatessaan aggressiivisemmin. Pelaajan muuttaessa tapaansa pelata kesken ottelun voidaan se nähdä myös pikemminkin pelaajakohtaisen riskin muutoksena rangaistuksen saamisen osalta tuomarista aiheutuvan vaikutuksen sijasta.

Olettaen, että tuomaritoiminnan kaavamaisuuksia ei voida perustella enemmän muiden tekijöiden aiheuttamiksi, on perusteltua argumentoida tuomarien olevan merkittävä tai jopa merkittävin syy kotiedun muodostumiseen. Tämän tutkielman tulokset ja siinä viittaamani tuomareita käsittelevä kirjallisuus osoittavat kaikki siihen, että

kotijoukkue hyötyy johdonmukaisesti enemmän näistä ennakoitavuuksista. Pohtiessa syitä tuomareista aiheutuvalla kotiedulle, ovat tässä tutkielmassa esille tulleet ottelutapahtumien konteksti ja yleisön aiheuttama paine tuomareita kohtaan todistetusti merkittäviä tekijöitä sen synnyssä, mutta ne eivät varmasti ole ainoita. Esimerkiksi Dohmenin ja Sauermannin (2015) katsaus tutkimuksiin eri urheilusarjojen tuomaritoiminnasta osoittaa, että on olemassa näyttöä tuomarien suosivan pelaajia ja joukkueita, jotka jakavat heidän etnisen ryhmän, kotipaikan tai kansalaisuuden. Lopezin ja muiden (2017) vertailu pohjoisamerikkalaisten urheilusarjojen joukkueiden voittotodennäköisyyksiin kuitenkin viittaa siihen, että sarjan joukkueiden välisen tasoeron laskiessa vähenee samalla kotiedunkin merkitys, joten sen muodostumisen vierittäminen yksiselitteisesti tuomarien vastuulle on myös kyseenalaista. Kotietu ja sen syyt ovat siis yksi mahdollinen tutkimussuunta tästä tutkielmasta eteenpäin.

Muodostaessa teoreettista pohjaa tuomarien kaavamaisuuksissa täytyy sitä pohjustaa ensin huomioimalla tuomarien tehtävän vaikeus, mikä korostuu etenkin jääkiekon kaltaisessa nopeatempoisessa joukkuelajissa. Tätä voidaan kuvata The Sports Dailyn (2018) artikkelissa esitetyllä datalla, jossa sääntökirjan tulkintoihin perustuen havainnoitiin 76 NHL-ottelun otannassa 3793 rikkeen määritelmän täyttävää tapahtumaa. Keskimäärin vain 17 prosenttia rikkeistä johtivat rangaistuksen antamiseen, joten jääkiekkotuomarit joutuvat toimimaan valtavaan tulkinnanvaraisuuteen perustuen. Jos tähän yhtälöön lisätään äänekäs yleisö, joka pyrkii vaikuttamaan päätöksentekoon, on inhimillistä odottaa tuomioiden tarkkuuden vaihtelevan parhaista yrityksistä huolimatta. Pelkästään koetilassa soitetun yleisömelun on havaittu vaikuttavan negatiivisesti tuomarisuoritukseen (Nevill ja muut, 2002), joten vaikutuksen voi olettaa olevan vielä suurempi todellisen ottelun aikana. Tuomaritoiminnan teoriaa käsittelevässä kirjallisuudessa MacMahon ja Plessner (2008) ovat kehittäneet menetelmän, jossa eri lajien tuomarit luokitellaan pelin tai ottelun vuorovaikutusasteen osalta. Jääkiekkotuomarit on asetettu vaativimpaan, eniten ottelun kanssa tekemisissä olevien vuorovaikuttajien (*interactor*) kategoriaan.

Huomioiden edellä mainitut seikat jääkiekkotuomarien tehtävän vaikeudesta ja tulkinnanvaraisuudesta, on kohtuullista olettaa tuomarien pyrkivän kompensoimaan näistä aiheutuvia ongelmia jollakin tavalla tiedostamattaankin. Yksi aikansa kokeneimmista NHL-tuomareista, Kelly Fraser, kuvailee esimerkkinä tilannetta, jossa tuomarin ensireaktio virheelliseen rangaistukseen on viheltää hyvityksenä seuraava rangaistus tästä kärsineen joukkueen hyväksi (TSN, 2011). Fraser sanoo tämän reaktion olevan ymmärrettävä ja vaikeasti ohitettava, sillä se on osa ihmisluontoa. Tämä inhimillisyys nousee esille Lopezin ja Snyderin (2013) teoriassa puolueellisesta puolueettomuudesta (*biased impartiality*), jossa he esittävät tuomarien toimivan otteluiden aikana hetkittäin puolueellisesti luodakseen vaikutelman puolueettomuudesta.

Näin toimien tuomarit kykenevät myös saavuttamaan tietynasteisen kontrollin tehtävänsä tulkinnanvaraisuuden osalta. Jääkiekkotuomari voidaan ymmärtää pikemminkin Mascarenhasin ja muiden (2002) esittämänä pelinjohtajana (*game manager*), joka pelkän sääntöjen tulkitsijan roolin sijasta vuorovaikuttaa ottelun ja pelaajien kanssa sekä mahdollistaa pelitapahtumien kulun. Jos tällaisten osittain alitajuisten metasääntöjen hyödyntäminen auttaa tuomareita johtamaan peliä mielekkäällä tavalla ja niiden olemassaolo on hyväksytty jokaisen toimijan osalta ainakin tiettyyn pisteeseen saakka, voidaan kaavamaisuuksia pitää olennaisena osana jääkiekon tuomaritoimintaa.

6 Yhteenveto

Tutkielman tavoitteena oli selvittää, pystyykö suomalaisessa liigajääkiekkoilussa vallalla olevia stereotyyppisiä käsityksiä tuomarien toiminnasta ja tämän ennakoitavuudesta todentamaan ja mittaamaan kvantifioitavalla tavalla. Koneoppimismenetelmiä hyödyntäen oli mahdollista osoittaa, että käsitykset ottelun tilanteen, ajan, jäähyjen määrän, jäähyjen peräkkäisyyden sekä kotiedun vaikutuksesta tuomareihin perustuvat ainakin jossain määrin todellisuuteen. Erityisesti kotiedun ja aikaisempien jäähyjen jakautumisen joukkueiden välillä voitiin havaita vaikuttavan merkittävällä tavalla tulevaan jäähyjakaumaan. Tulokset ovat myös pitkälti yhtenäisiä joukkueurheilun tuomaritoiminnan kaavamaisuuksia käsittelevän kirjallisuuden kanssa. Tuomaritoiminnan johdonmukainen ennakoitavuus voi mahdollisesti selittyä sillä, että etenkin kaoottisissa, vaikeasti tuomaroitavissa lajeissa metasääntöjen kaltaiset, joko tiedostetut tai tiedostamattomat kaavamaisuudet mahdollistavat rakenteellisuuden luomisen ottelun kulkuun.

Lähdeluettelo

- Abrevaya, J. & McCulloch, R. (2014). Reversal of fortune: a statistical analysis of penalty calls in the National Hockey League. *Journal of Quantitative Analysis in Sports*, 10(2), 207–224. <https://doi.org/10.1515/jqas-2013-0067>
- Alpaydin, E. (2014). *Introduction to Machine Learning*. The MIT Press.
- Beaudoin, D., Schulte, O. & Swartz, T. (2016). Biased penalty calls in the National Hockey League. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5), 365–372. <https://doi.org/10.1002/sam.11320>
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Boyko, R. H., Boyko, A. R. & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences*, 25(11), 1185–1194. <https://doi.org/10.1080/02640410601038576>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buja, A., Stuetzle, W. & Shen, Yi. (2005). *Loss functions for binary class probability*

estimation and classification: structure and applications. <https://www-stat.wharton.upenn.edu/~buja/PAPERS/paper-proper-scoring.pdf>

- Buraimo, B., Forrest, D. & Simmons, R. (2010). The 12th man?: refereeing bias in English and German soccer. *Journal of the Royal Statistical Society Series A*, 173(2), 431–449. <https://doi.org/10.1111/j.1467-985X.2009.00604.x>
- Dohmen, T. (2008). The influence of social forces: evidence from the behavior of football referees. *Economic Inquiry*, 46, 411–424. <https://doi.org/10.1111/j.1465-7295.2007.00112.x>
- Dohmen, T. & Sauermann, J. (2015). Referee bias. *Journal of Economic Surveys*, 30(4), 679–695. <https://doi.org/10.1111/joes.12106>
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- GitHub. (2020). *Scrapy, a fast high-level web crawling & scraping framework for Python*. <https://github.com/scrapy/scrapy> (Haettu 12.4.2020)
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Jones, M. V., Paull, G.C. & Erskine, J. (2002). The impact of a team’s aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences*, 20(12), 991–1000. <https://doi.org/10.1080/026404102321011751>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI’95: Proceedings of the 14th international joint conference on Artificial intelligence*, 2, 1137–1143. <https://doi.org/10.1.1.48.529>
- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Langley, P. (2011). The changing science of machine learning. *Machine Learning*, 82(3), 275–279. <https://doi.org/10.1007/s10994-011-5242-y>
- Liiga. (2020). *Etusivu*. <https://liiga.fi/fi/> (Haettu 12.4.2020)
- Lopez, M. J., Matthews, G. J. & Baumer, B. S. (2018). How often does the best team win? A unified approach to understanding randomness in North American sport. *The Annals of Statistics*, 12(4), 2483–2516. <https://doi.org/10.1214/18-AOAS1165>
- Lopez, M. J. & Snyder, K. (2013). Biased impartiality among National Hockey League referees. *International Journal of Sport Finance*, 9, 208–223. <https://doi.org/10.2139/ssrn.2259798>
- MacMahon, C. & Plessner, H. (2008). The sport official in research and practice. In D. Farrow, J. Backer & C. MacMahon (Eds.), *Developing sport expertise: Lessons from theory and practice* (172–190). Routledge.
- Mascarenhas, D. R. D, Collins, D. & Mortimer, P. (2002). The art of reason versus the exactness of science in elite refereeing: comments on Plessner and Betsch (2001). *Journal of Sport and Exercise Psychology*, 24(3), 328–333. <https://doi.org/10.1123/jsep.24.3.328>
- Mitchell, R. (2018). *Web Scraping with Python: Collecting Data from the Modern Web*. O’Reilly Media.
- Murto, F. (2020). *Liigan jäähydata*. <https://github.com/f-murto/liiga-penalties>

- Nevill, A. M., Balmer, N. J. & Williams, M. A. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), 261–272. [https://doi.org/10.1016/S1469-0292\(01\)00033-4](https://doi.org/10.1016/S1469-0292(01)00033-4)
- Page, K. & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, 31(2), 192–199. <https://doi.org/10.1016/j.joep.2009.08.007>
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Ridgeway, G. (2007). *Generalized boosted models: a guide to the gbm package*. The R Foundation. <https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf>
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210–229. <https://doi.org/10.1147/rd.33.0210>
- Schuckers, M. & Brozowski, L. (2012). Referee analytics: An analysis of penalty rates by National Hockey League officials. *MIT Sloan Sports Analytics Conference, 2012*.
- Sutter, M. & Kocher, M. G. (2004). Favoritism of agents – The case of referees' home bias. *Journal of Economic Psychology*, 25(4), 461–469. [https://doi.org/10.1016/S0167-4870\(03\)00013-8](https://doi.org/10.1016/S0167-4870(03)00013-8)
- The Sports Daily. (2018). *The NHL has an officiating problem*. <https://thesportsdaily.com/2018/03/16/the-nhl-has-an-officiating-problem/> (Haettu 12.4.2020)
- TSN. (2011). *Fraser: Refs are human, but make-up calls don't atone*. https://web.archive.org/web/20110617062838/https://www.tsn.ca/blogs/kerry_fraser/?id=368803 (Haettu 12.4.2020)
- Unkelbach, C. & Memmert, D. (2008). Game management, context effects, and calibration: the case of yellow cards in soccer. *Journal of Sport & Exercise Psychology*, 30(1), 95–109. <https://doi.org/10.1123/jsep.30.1.95>