



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

Data-analytiikkaan perustuvien ennustemallien ja ve- donlyöntikertoimien suhde jalkapallossa

Tietojärjestelmätieteen kandidaattitutkielma

Laatija:
Daniel Lakka

Ohjaaja:
Samuli Laato

19.5.2026
Turku

Opiskelijan lausunto tekoölyn käytöstä tähän tutkielmaan liittyen:

En ole käyttänyt tekoölyä hyödyntäviä työkaluja tätä tutkielmaa kirjoittaessani.

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

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

Kandidaatintutkielma

Oppiaine: Tietojärjestelmätiede

Tekijä: Daniel Lakka

Otsikko: Data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien suhde jalkapallossa

Ohjaaja: Tohtori Samuli Laato

Sivumäärä: 32 sivua + liitteet 1 sivu

Päivämäärä: 19.5.2026

Tiivistelmä

Vedonlyöntimarkkinat ovat keskeinen osa jalkapallo-otteluiden lopputulosten ennustamista, sillä vedonlyöntikertoimet heijastavat markkinoilla saatavilla olevaa informaatiota joukkueiden voimasuhteista. Samalla data-analytiikkaan perustuvat ennustemallit, kuten tilastolliset ja koneoppimismenetelmät, ovat kehittyneet merkittävästi viime vuosina. Tämän kandidaatintutkielman tavoitteena on tarkastella vedonlyöntikertoimien ja ennustemallien välistä suhdetta sekä selvittää, kuinka hyvin kertoimista johdetut implisiittiset todennäköisyydet vastaavat toteutuneita ottelutuloksia. Lisäksi tutkielmassa tarkastellaan, miten vedonlyöntikertoimet reagoivat uuteen informaatioon ennen ottelua. Tutkimus toteutetaan kirjallisuuskatsauksena, jossa hyödynnetään aiempaa tutkimusta vedonlyöntimarkkinoiden tehokkuudesta, ennustemalleista ja probabilististen ennusteiden arvioinnista

Tulosten perusteella vedonlyöntikertoimet tarjoavat keskimäärin hyvin kalibroidun ja kilpailukykyisen arvion otteluiden lopputuloksista. Vaikka data-analytiikkaan perustuvat mallit voivat täydentää markkinoiden tuottamaa informaatiota, ne eivät kykene systemaattisesti ylittämään vedonlyöntimarkkinoiden suorituskykyä. Vedonlyöntikertoimet päivittyvät jatkuvasti uuden informaation perusteella, jolloin selkeästi mitattavat tekijät, kuten kokoonpanomuutokset, heijastuvat kertoimiin nopeasti, kun taas vaikeammin tulkittavat tekijät vaikuttavat epäsuoremmin. Kokonaisuutena vedonlyöntimarkkinat kokoavat tehokkaasti informaatiota useista lähteistä, mutta markkinoilla esiintyy myös pieniä tehottomuuksia, mikä viittaa siihen, että markkina ei ole täysin tehokas.

Avainsanat: vedonlyöntikertoimet, ennustemallit, jalkapallo, data-analytiikka, kalibraatio, Poisson-malli

SISÄLLYS

1	Johdanto	7
2	Ennustemallien ja vedonlyöntikertoimien suhde kirjallisuudessa	9
	2.1 Vedonlyöntikertoimet todennäköisyyksinä ja markkinoiden tehokkuus	9
	2.2 Tilastolliset ennustemallit ja niiden vertailu vedonlyöntikertoimiin	10
	2.3 Koneoppiminen ja moderni data-analytiikka jalkapallovedonlyönnissä	12
	2.4 Ennustetarkkuus ja taloudellinen kannattavuus	13
3	Implisiittisten todennäköisyyksien vastaavuus toteutuneisiin tuloksiin	15
	3.1 Vastaavatko todennäköisyydet toteutumia?	15
	3.2 Systemaattiset vinoumat implisiittisissä todennäköisyyksissä	17
	3.3 Vedonlyöntimarkkinat tietojärjestelminä ja informaation aggregointina	17
	3.4 Miten vastaavuutta arvioidaan?	18
	3.5 Missä määrin kertoimet vastaa toteutumia?	19
4	Vedonlyöntikertoimien reagointi uuteen informaatioon	20
	4.1 Kokoonpanot ja pelaajakohtainen informaatio	20
	4.2 Joukkueen suorituskky ja viimeaikainen formi	21
	4.3 Kotietu, stadion ja sääolosuhteet	22
	4.4 Valmennus ja organisatoriset muutokset	24
	4.5 Miten eri informaatiotekijät heijastuvat kertoimiin?	25
5	Yhteenveto ja johtopäätökset	27
	Lähteet	30
	Liitteet	33
	Liite 1 Selvitys tekoälyn käytöstä	33

KUVIOT

Kuvio 1 Poisson-jakauma joukkueen maalimäärälle ($\lambda = 1,5$)	10
Kuvio 2 Ennustemallin kalibraatio	16
Kuvio 3 Vedonlyöntikertoimien muutos ennen ottelua	20

1 Johdanto

Urheiluedonlyöntimarkkinat ovat kehittyneet merkittävästi digitalisaation ja laskennallisten menetelmien myötä. Erityisesti jalkapallovedonlyönnissä vedonlyöntikertoimet toimivat keskeisenä mekanismina, jonka avulla vedonlyöntiyhtiöt hinnoittelevat lopputuloksiin liittyviä todennäköisyyksiä. Aiemmissa tutkimuksissa on osoitettu, että vedonlyöntikertoimet sisältävät merkittävää informaatiota otteluiden todennäköisistä lopputuloksista ja heijastavat markkinoilla saatavilla olevaa informaatiota (Angelini ym. 2022). Jalkapallo on maailman seuratuin ja taloudellisesti merkittävin urheilulaji, minkä vuoksi se tarjoaa erinomaisen lähtökohdan vedonlyöntikertoimien tarkastelulle.

Aihe on tärkeä useille toimijoille. Vedonlyöntiyhtiöille vedonlyöntikertoimien hinnoittelu on keskeinen osa liiketoiminnan kannattavuuden ja riskienhallinnan näkökulmasta. Vedonlyöjille kertoimet toimivat keskeisenä informaation lähteenä ottelun todennäköisestä lopputuloksesta. Akateemisesta näkökulmasta vedonlyöntikertoimet tarjoavat mahdollisuuden tutkia markkinatehokkuutta, ennustamista sekä päätöksenteon taustalla olevaa informaation hyödyntämistä (Levitt, 2004). Data-analytiikkaan perustuvat ennustemallit, kuten tilastolliset ja koneoppimiseen perustuvat mallit ovat nousseet keskeiseen rooliin urheilutulosten ennustamisessa, mutta niiden suhdetta asetettuihin vedonlyöntikertoimiin on tutkittu rajallisemmin.

Aiempi kirjallisuus on pääasiassa keskittynyt vedonlyöntimarkkinoiden tehokkuuden arviointiin sekä kertoimien ennustekyvyn tarkasteluun (Angelini ym. 2022; Forrest ym. 2008). Vaikka aiempi kirjallisuus on tarkastellut vedonlyöntimarkkinoiden tehokkuutta ja kertoimien ennustekykyä, vähemmän on kiinnitetty huomiota siihen, miten data-analytiikan kehitys on mahdollisesti muuttanut kertoimien muodostusprosessia erityisesti jalkapallossa. Tämän ilmiön tarkastelu muodostaa tutkimusaukon, johon tämä kandidaattitutkielma pyrkii vastaamaan.

Tutkielman tavoitteena on muodostaa kokonaiskuva data-analytiikan roolista jalkapallon vedonlyöntikertoimien muodostamisessa aikaisemman tutkimuksen perusteella. Tutkimuksessa tarkastellaan, miten data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien välistä suhdetta on kuvattu aiemmassa tutkimuksessa. Lisäksi käsitellään, miten uuden informaation ja tilastollisen datan, kuten joukkueiden viimeaikaisten suoritusten, loukkaantumisten ja kokoonpanotietojen vaikutusta vedonlyöntikertoimiin on tarkasteltu kirjallisuudessa. Näiden näkökulmien avulla pyritään jäsentämään ja selventämään data-analytiikan merkitystä vedonlyöntikertoimien muodostumisessa jalkapallon kontekstissa.

Tutkielma toteutettiin kirjallisuuskatsauksena, jonka tavoitteena oli muodostaa kokonaiskuva data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien suhteesta jalkapallossa aieman tutkimuksen perusteella. Kirjallisuutta haettiin hyödyntämällä Google Scholaria sekä yliopiston tarjoamia tietokantoja. Aineistoon valittiin erityisesti vertaisarvioituja tutkimuksia, jotka käsitelivät vedonlyöntimarkkinoiden tehokkuutta, jalkapallo-otteluiden ennustamista ja data-analytiikan hyödyntämistä urheilukontekstissa. Lisäksi työssä hyödynnettiin alan keskeisiä klassisia tutkimuksia. Kirjallisuutta tarkasteltiin teemoittain tutkimuskysymysten mukaisesti keskittyen ennustemallien ja vedonlyöntikertoimien suhteeseen, implisiittisten todennäköisyyksien vastaavuuteen sekä vedonlyöntikertoimien reagointiin uuteen informaatioon.

Tutkielmaa ohjaavat seuraavat tutkimuskysymykset:

1. Miten aiempi tutkimus kuvaa data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien välistä suhdetta jalkapallossa?
2. Missä määrin vedonlyöntikertoimista johdetut implisiittiset todennäköisyydet vastaavat toteutuneita ottelutuloksia?
3. Miten uusi informaatio (esim. loukkaantumiset, kokoonpanot, viimeaikaiset suoritukset) heijastuvat vedonlyöntikertoimiin ennen ottelua?

2 Ennustemallien ja vedonlyöntikertoimien suhde kirjallisuudessa

Aiempi tutkimus on tarkastellut data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien suhdetta useista näkökulmista (esim. Angelini ym. 2022; Forrest ym. 2008). Kirjallisuuden perusteella vedonlyöntikertoimet toimivat usein vahvana vertailukohtana ennustemalleille, sillä ne sisältävät merkittävän määrän markkinoilla saatavilla olevaa informaatiota. Tilastolliset ja koneoppimiseen perustuvat mallit voivat saavuttaa kilpailukykyisen ennustetarkkuuden, ja jossain tapauksissa ylittää yksinkertaisemmat markkinapohjaiset arviot. Kuitenkin systemaattisen taloudellisen ylituoton saavuttaminen markkinoihin nähden on osoittautunut vaikeaksi. Seuraavissa alaluvuissa tarkastellaan näitä tuloksia yksityiskohtaisemmin.

2.1 Vedonlyöntikertoimet todennäköisyyksinä ja markkinoiden tehokkuus

Vedonlyöntikertoimet toimivat mekanismina, jonka avulla vedonlyöntiyhtiöt hinnoittelevat urheilutapahtumien eri lopputuloksiin liittyviä todennäköisyyksiä. Desimaalikertoimista voidaan johtaa implisiittinen todennäköisyys kaavalla $p = 1/k$, missä k on vedonlyöntikerroin. Käytännössä vedonlyöntiyhtiöt sisällyttävät kertoimiin oman katteensa (ns. overround), minkä vuoksi implisiittisten todennäköisyyksien summa ylittää yhden (Forrest ym. 2005). Tämän vuoksi kertoimet eivät suoraan vastaa objektiivista todennäköisyyttä, vaan ne heijastavat sekä arvioitua lopputuloksen todennäköisyyttä että vedonlyöntiyhtiöiden riskienhallintaa.

Akateemisesta näkökulmasta vedonlyöntikertoimia on tarkasteltu erityisesti markkinatehokkuuden näkökulmasta. Tehokkaiden markkinoiden hypoteesin (Fama 1970) mukaan hinnat heijastavat kaikkia saatavilla olevaa informaatiota, eikä systemaattista ylituottoa voida saavuttaa julkisen tiedon perusteella. Vedonlyöntikertoimet tarjoavat kiinnostavan tutkimusympäristön, koska lopputulos realisoituu nopeasti ja todennäköisyysennusteet ovat eksplisiittisesti havaittavissa kertoimissa (Sauer 1998).

Jalkapallon kontekstissa vedonlyöntimarkkinoiden tehokkuutta on tarkasteltu erityisesti vertaamalla kertoimista johdettuja implisiittisiä todennäköisyyksiä toteutuneisiin ottelutuloksiin. Pope ja Peel (1989) sekä Kuypers (2000) analysoivat Englannin jalkapallomarkkinoita ja havaitsivat, että markkinat ovat pääosin informatiivisia, mutta niissä esiintyy tiettyjä systemaattisia vinoumia. Yksi tunnetuimmista on niin sanottu favourite-longshot bias, jossa otteluiden suosikkien ja altavastajien kertoimet eivät täysin vastaa toteutuneita todennäköisyyksiä (Cain ym. 2000).

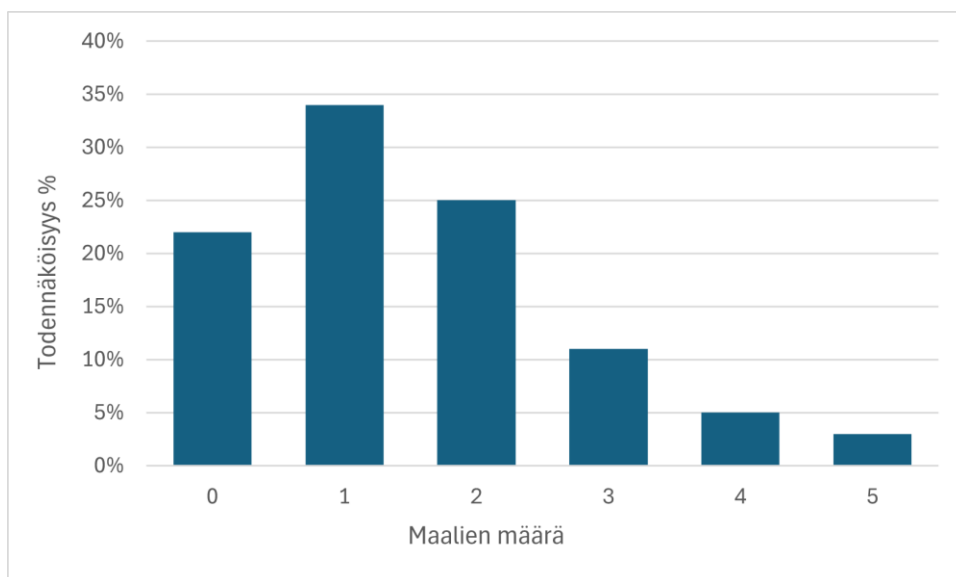
Myöhemmät tutkimukset ovat vahvistaneet, että jalkapallon vedonlyöntimarkkinat ovat suhteellisen tehokkaita, mutta eivät täysin virheettömiä. Vedonlyöntikertoimet sisältävät merkittävää informaatiota otteluiden todennäköisistä lopputuloksista ja toimivat vahvana ennustajana suhteessa yksinkertaisiin tilastollisiin malleihin (Simmons 2005; Angelini & De Angelis 2019; Angelini ym. 2022).

Tutkimuksissa on myös havaittu, että tehokkuuden taso voi vaihdella liigoittain ja ajanjaksoittain.

Jalkapallovedonlyönnin näkökulmasta vedonlyöntikertoimia voidaan näin ollen tarkastella kollektiivisina ennustajina, jotka aggregoivat vedonlyöntiyhtiöiden mallinnuksen sekä markkinaosapuolten informaation yhdeksi numeeriseksi arvoksi. Kirjallisuuden perusteella markkinat toimivat useimmiten tehokkaasti, mutta niissä esiintyvät pienet vinoumat ja ajoittaiset tehottomuudet avaavat mahdollisuuden tarkastella tarkemmin, miten data-analytiikkaan perustuvat ennustemallit suhteutuvat näihin markkinoihin. Tätä suhdetta käsitellään tarkemmin seuraavassa alaluvussa.

2.2 Tilastolliset ennustemallit ja niiden vertailu vedonlyöntikertoimiin

Jalkapallo-otteluiden lopputulosten ennustamiseen on kehitetty useita tilastollisia malleja, joita on käytetty myös vedonlyöntimarkkinoiden tehokkuuden arvioinnissa. Näitä malleja käytetään jalkapallossa myös vedonlyöntikertoimien ennustekyvyn arviointiin, minkä vuoksi ne muodostavat keskeisen vertailukohtan vedonlyöntimarkkinoille. Varhaisissa tutkimuksissa Poisson-jakaumaan perustuvat mallit ovat olleet keskeisessä roolissa (Dixon & Coles 1997). Poisson-mallin toimintaa voidaan havainnollistaa jakaumana, jossa eri maalimäärille estimoidaan todennäköisyydet (ks. Kuvio 1).



Kuvio 1 Poisson-jakauma joukkueen maalimäärälle ($\lambda = 1,5$)

Näissä malleissa joukkueiden maalimääriä mallinnetaan stokastisena prosessina, jonka perusteella voidaan johtaa todennäköisyydet eri lopputuloksille. Dixon ja Coles (1997) osoittavat, että mallipohjaiset todennäköisyydet voivat jossain tapauksissa tuottaa markkinoita parempaa ennustetarkkuutta, mutta tulokset ovat herkkiä aineiston rajaukselle.

Elo-luokitusjärjestelmä on toinen laajasti käytetty lähestymistapa joukkueiden suhteellisen vahvuuden arviointiin (Hvattum & Arntzen 2010). Elo-mallit päivittävät joukkueiden vahvuusarvioita dynaamisesti otteluiden jälkeen, mikä tekee niistä soveltuvia päivittäiseen ennustamiseen. Vertailu vedonlyöntikertoimiin osoittaa kuitenkin, että vaikka Elo-mallit tarjoavat kilpailukykyisiä ennusteita, vedonlyöntimarkkinat saavuttavat usein vähintään saman ennustetarkkuuden.

Urheilutaloustieteen kirjallisuus puolestaan osoittaa, että vedonlyöntikertoimet toimivat vahvana vertailukohtana yksittäisille tilastollisille malleille (Forrest ym. 2005; Angelini & De Angelis 2019). Useissa tutkimuksissa markkinakertoimet saavuttavat yhtä hyvän tai paremman ennustetarkkuuden kuin yksittäiset Poisson- tai regressiomallit. Tämä viittaa siihen, että vedonlyöntimarkkinat toimivat kollektiivisena ennustejärjestelmänä, joka sisällyttää enemmän informaatiota kuin yksittäinen tilastollinen malli.

Tietojärjestelmätieteen näkökulmasta vedonlyöntimarkkinoita voidaan tarkastella informaation aggregointijärjestelminä. Ennustemarkkinoita koskevissa tutkimuksissa on osoitettu, että markkinahinnat kykenevät tehokkaasti yhdistämään hajautetun informaation yhdeksi numeeriseksi ennusteeksi (Wolfers & Zitzewitz, 2004; Chen ym. 2011). Chen ym. (2011) julkaistuna *Information Systems Research* -lehdessä osoittavat, että markkinamekanismi voi ylittää yksittäisen ennustajan informaation yhdistämisessä. Tämä tukee näkemystä, jonka mukaan vedonlyöntikertoimet eivät ole pelkästään vedonlyöntiyhtiöiden arvioita, vaan heijastavat laajempaa informaatiovirtaa.

Lisäksi algoritmisen päätöksenteon ja analytiikan tutkimus tietojärjestelmätieteessä korostaa, että dataperusteiset hinnoittelumeکانismit voivat kehittyä oppiviksi järjestelmiksi (Brynjolfsson & McElheran 2016; Shmueli & Koppius 2011). Shmueli ja Koppius (2011), julkaistuna *MIS Quarterly* lehdessä, erottavat selittävä ja ennustava mallintaminen toisistaan ja korostavat ennustemallien arvioinnissa käytettävien mittareiden merkitystä. Tätä viitekehitystä voidaan soveltaa myös vedonlyöntikertoimien ja tilastollisten mallien vertailuun. Keskeistä ei ole pelkästään mallin selitysvoima, vaan sen ennustetarkkuus suhteessa markkinahintaan.

Yhteenvetona aiempi tutkimus osoittaa, että tilastolliset ennustemallit voivat saavuttaa korkean ennustetarkkuuden, mutta vedonlyöntikertoimet muodostavat vahvan vertailukohdan niiden

suorituskyvyille. Markkinakertoimet näyttävät aggregoivan laajemman informaatiokokonaisuuden kuin yksittäinen malli, mikä tukee näkemystä vedonlyöntimarkkinoista kollektiivisena ennustejärjestelmänä. Tämä asettaa korkean vertailutason yksittäisille data-analytiikkaan perustuville malleille, joiden on kyettävä tuottamaan lisäinformaatiota suhteessa markkinahintaan.

2.3 Koneoppiminen ja moderni data-analytiikka jalkapallovedonlyönnissä

Viime vuosikymmenen aikana jalkapallo-otteluiden ennustamisessa on siirrytty perinteisistä tilastollisista malleista kohti koneoppimiseen perustuvia menetelmiä. Useat tutkimukset ovat soveltaneet esimerkiksi neuroverkkoja, satunnaismetsiä (random forests) ja gradienttivahvistusmenetelmiä ottelutulosten ennustamiseen (Hubáček ym. 2019; Baboota & Kaur 2019). Näiden menetelmien etuna on niiden kyky mallintaa epälineaarisia ja monimutkaisia riippuvuussuhteita ilman vahvoja jakaumaoletuksia, jotka rajoittavat perinteisiä Poisson- ja regressiomalleja (James ym. 2021).

Koneoppimismallien kehitystä on tukenut myös urheiludatan määrän ja laadun kasvu. Tapahtuma-kohtainen pelidata (event data), joka sisältää yksityiskohtaista tietoa syötöistä, laukauksista ja pallonhallinnasta, on mahdollistanut entistä tarkempien suorituskykymittareiden kehittämisen (Pappalardo ym. 2019). Yksi keskeisimmistä uusista mittareista on odotettujen maalien (expected goals, xG) malli, joka arvioi maalinteon todennäköisyyttä laukaisutilanteen ominaisuuksien perusteella (Rathke 2017). Näiden mittareiden hyödyntäminen on parantanut ennustemallien kykyä arvioida joukkueiden todellista suorituskykyä pelän lopputuloksen sijaan.

Hubáček ym. (2019) osoittavat, että koneoppimismallit voivat parantaa ennustetarkkuutta verrattuna yksinkertaisiin tilastollisiin malleihin, mutta markkinakertoimet säilyvät usein vahvana vertailukohdana. Samansuuntaisia tuloksia esittävät Baboota ja Kaur (2019), joiden mukaan koneoppimismenetelmät tarjoavat tarkempia ennusteita kuin perinteiset menetelmät, mutta niiden kyky tuottaa johdonmukaista taloudellista tuottoa on rajallinen. Näin ollen kirjallisuus viittaa siihen, että vaikka koneoppiminen parantaa tilastollista ennustetarkkuutta, markkinoiden systemaattinen voittaminen on erityisesti informaatorikkaissa jalkapalloliigoissa haastavaa.

Tietojärjestelmätutkimuksessa digitaalisia markkinoita on tarkasteltu informaation aggregointimekanismeina, joissa hajautettu tieto yhdistyy markkinahinnaksi (Wolfers & Zitzewitz, 2004; Chen ym. 2011). Chen ym. (2011) osoittavat, että ennustemarkkinat kykenevät tehokkaasti yhdistämään useiden toimijoiden informaation yhdeksi kollektiiviseksi ennusteeksi. Tätä näkökulmaa soveltaen vedonlyöntikertoimet voidaan tulkita dynaamiseksi informaatiotuotteiksi, jotka heijastavat sekä vedonlyöntiyhtiön että markkinaosapuolten analytiikkaa. Tällaisessa ympäristössä yksittäisen

koneoppimismallin on vaikea saavuttaa pysyvää kilpailuetua, koska markkinahinnat mukautuvat nopeasti uuteen informaatioon.

Tietojärjestelmätieteen näkökulmasta koneoppiminen voidaan nähdä ennustavan analytiikan soveltuksena, jossa mallien arvioinnissa painotetaan ennustetarkkuutta selitysvoiman sijaan (Shmueli & Koppius 2011). Tämä näkökulma on keskeinen vedonlyöntimarkkinoiden tutkimuksessa, sillä vedonlyöntikertoimet tarjoavat eksplisiittisen vertailukohdan mallien tuottamille todennäköisyyksille. Ennustemallien suorituskykyä arvioidaan tyypillisesti vertaamalla niiden tuottamia todennäköisyysennusteita markkinakertoimista johdettuihin implisiittisiin todennäköisyyksiin.

Modernin data-analytiikan ja koneoppimisen kehitys on lisännyt ennustemallien kykyä hyödyntää suuria ja monimuotoisia datalähteitä. Kirjallisuus viittaa kuitenkin siihen, että vedonlyöntimarkkinat mukautuvat nopeasti uuteen informaatioon ja analytiikkaan, mikä rajoittaa yksittäisten mallien mahdollisuuksia saavuttaa pysyvää kilpailuetua. Näin ollen koneoppimismallien ja vedonlyöntikertoimien suhde voidaan tulkita dynaamisena vuorovaikutuksena, jossa markkina toimii jatkuvasti päivittyvänä metaennusteena.

2.4 Ennustetarkkuus ja taloudellinen kannattavuus

Aiemmassa kirjallisuudessa ennustemallien suorituskykyä arvioidaan tyypillisesti kahdesta näkökulmasta: ennustetarkkuutena (forecast accuracy) ja taloudellisena kannattavuutena (profitability). Nämä eivät kuitenkaan ole sama asia, eikä parantunut ennustetarkkuus automaattisesti johda taloudellisesti kannattavaan vedonlyöntistrategiaan. Vedonlyöntimarkkinoilla tuottoa syntyy vain, jos mallin tuottamat todennäköisyydet poikkeavat systemaattisesti markkinakertoimista. Lisäksi vedon tulee olla vedonlyöjän näkökulmasta positiivisen odotusarvon veto, ja tämän edun on ylitettävä vedonlyöntiyhtiöiden kate (overround) (Forrest ym. 2005).

Hvattum ja Arntzen (2010) vertailevat Elo-luokituksiin perustuvia ennusteita vedonlyöntikertoimiin ja osoittavat, että vaikka tilastolliset mallit voivat saavuttaa kilpailukykyisen ennustetarkkuuden, vedonlyöntimarkkinat toimivat usein vähintään yhtä tehokkaasti. Tämä viittaa siihen, että markkinakertoimet sisältävät jo merkittävän määrän hajautettua informaatiota, mikä vaikeuttaa systemaattisen ylituoton saavuttamista.

Todennäköisyysennusteiden tarkkuutta arvioidaan kirjallisuudessa usein ns. proper scoring rules -mittareilla, jotka mittaavat ennusteen kalibraatiota ja informatiivisuutta. Yksi varhaisimmista ja laajimmin käytetyistä mittareista on Brier score (Brier 1950), joka mittaa ennustettujen todennäköisyyksien ja toteutuneiden lopputulosten välistä neliöpoikkeamaa. Gneiting ja Raftery (2007)

osoittavat, että proper scoring rules -mittarit, kuten Brier score ja logaritminen tappio (log loss), ovat teoreettisesti perusteltuja arvioitaessa probabilistisia ennusteita, koska ne kannustavat hyvin kalibroituhiin ja totuudenmukaisiin todennäköisyyksiin. Näitä mittareita voidaan soveltaa probabilistisiin ennusteisiin, mukaan lukien ennustemallien tuottamat ja vedonlyöntikertoimista johdetut todennäköisyydet.

Taloudellinen kannattavuus puolestaan riippuu siitä, kykeneekö ennustemalli tunnistamaan markkinahinnoittelussa systemaattisia poikkeamia. Vedonlyöntimarkkinoilla vedonlyöntiyhtiöiden kate (overround) asettaa rakenteellisen esteen systemaattiselle ylituotolle. Vaikka malli tuottaisi tarkkoja todennäköisyysennusteita, taloudellinen hyöty syntyy vain, jos mallin arvioima todennäköisyys poikkeaa riittävästi markkinakertoimesta siten, että vedolla on positiivinen odotusarvo.

Wunderlich ja Memmert (2020) tarkastelevat suoraan ennustetarkkuuden ja vedonlyöntituottojen välistä suhdetta ja osoittavat, että nämä mittarit eivät välttämättä korreloi keskenään. Heidän tulosensa mukaan malli voi saavuttaa paremman tilastollisen ennustetarkkuuden ilman, että se tuottaa korkeampaa taloudellista tuottoa, ja päinvastoin. Tämä viittaa siihen, että pelkkä osumatarkkuus tai todennäköisyysennusteiden kalibraatio ei yksin riitä taloudellisen kannattavuuden arviointiin.

Taloudellinen kannattavuus riippuu myös panostusstrategiasta. Kellyn kriteeri tarjoaa teoreettisen viitekehysten optimaalisen panoksen määrittämiseksi tilanteessa, jossa vedolla on positiivinen odotusarvo. Kelly (1956) osoittaa, että toistuvissa vedonlyöntitilanteissa optimaalinen panos voidaan määrittää maksimoimalla varallisuuden odotettu logaritminen kasvunopeus. Tällöin panostusstrategia ei perustu yksittäisen vedon onnistumiseen, vaan pitkän aikavälin pääoman kasvuun.

Ennustemallien arvioinnissa vedonlyöntikontekstissa on siten erotettava tilastollinen suorituskyky ja taloudellinen suorituskyky. Probabilistiset tarkkuusmittarit (Brier 1950; Gneiting & Raftery 2007) arvioivat ennusteiden laatua, kun taas taloudellinen kannattavuus riippuu markkinahinnoittelun poikkeamista, vedonlyöntiyhtiöiden katteesta sekä panostusstrategiasta (Kelly 1956). Lisäksi empiirinen tutkimus osoittaa, että ennustetarkkuuden ja vedonlyöntituottojen välinen suhde ei ole yksiselitteinen (Wunderlich & Memmert 2020). Tämä korostaa, että data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien suhdetta arvioitaessa on huomioitava sekä tilastollinen tarkkuus että markkinaympäristön taloudellinen rakenne.

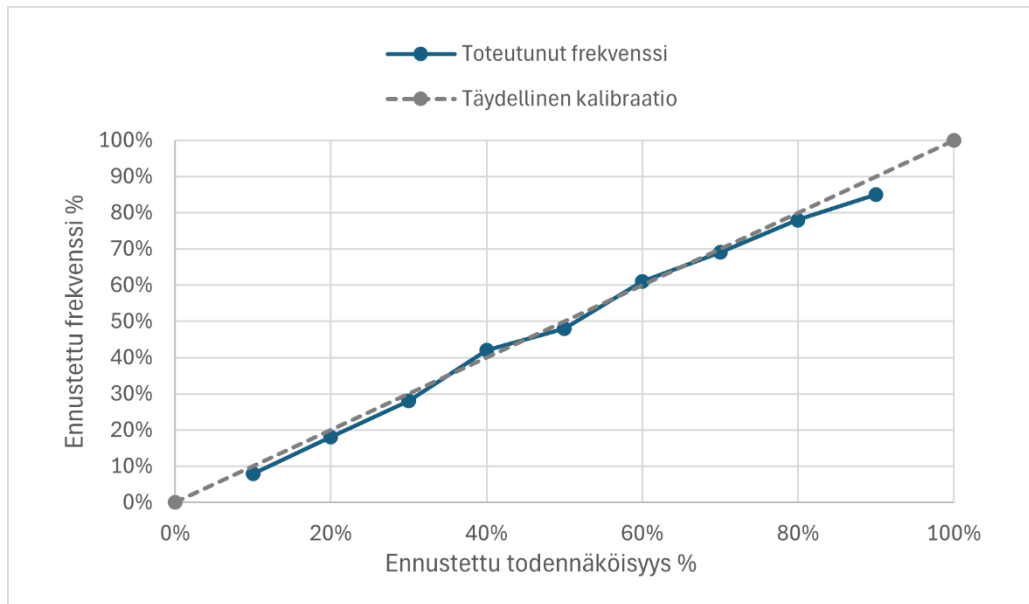
3 Implisiittisten todennäköisyyksien vastaavuus toteutuneisiin tuloksiin

Jalkapallovedonlyönnissä vedonlyöntikertoimista johdettujen implisiittisten todennäköisyyksien ja toteutuneiden ottelutulosten välinen vastaavuus muodostaa keskeisen näkökulman vedonlyöntimarkkinoiden ennustekyvyn arvioinnissa. Kirjallisuuden perusteella tätä suhdetta tarkastellaan vertaamalla kertoimista johdettuja todennäköisyyksiä toteutuneisiin lopputuloksiin, jolloin voidaan arvioida markkinoiden kalibraatiota ja tehokkuutta.

Forrest ym. (2005) sekä Wolfers ja Zitzewitz (2004) osoittavat, että vedonlyöntikertoimet voidaan tulkita informaation aggregointimekanismeiksi, joissa yhdistyvät sekä vedonlyöntiyhtiöiden analytiikka että markkinaosapuolten hajautettu tieto. Tietojärjestelmätieteen näkökulmasta tämä mahdollistaa vedonlyöntimarkkinoiden tarkastelun järjestelminä, joiden suorituskykyä voidaan arvioida niiden tuottamien todennäköisyysarvioiden ja toteutuneiden tulosten välisen vastaavuuden perusteella.

3.1 Vastaavatko todennäköisyydet toteutumia?

Keskeinen tapa arvioida vedonlyöntikertoimien laatua on tarkastella niiden kalibraatiota, eli sitä, kuinka hyvin implisiittiset todennäköisyydet vastaavat toteutuneiden lopputulosten frekvenssejä (Gneiting & Raftery 2007). Kalibraatiolla tarkoitetaan sitä, että esimerkiksi 60% todennäköisyyden saavat tapahtumat toteutuvat pitkällä aikavälillä noin 60% tapauksista (Brier 1950; Gneiting & Raftery 2007).



Kuvio 2 Ennustemallin kalibraatio

Tämä on keskeinen ominaisuus probabilistisille ennusteille, sillä hyvin kalibroitu ennustejärjestelmä tuottaa pitkällä aikavälillä oikeansuuntaisia todennäköisyysarvioita, vaikka yksittäiset ennusteet voivatkin olla virheellisiä.

Forrest ym. (2005), Hvattum ja Arntzen (2010) sekä Angelini ym. (2022) osoittavat, että vedonlyöntimarkkinat ovat keskimäärin hyvin kalibroituja ja implisiittiset todennäköisyydet vastaavat melko tarkasti toteutuneita ottelutuloksia eri jalkapalloliigoissa. Tämä tarkoittaa käytännössä sitä, että markkinat onnistuvat hinnoittelemaan otteluiden lopputulosten todennäköisyydet suhteellisen realistisesti, mikä viittaa korkeaan informaationsältöön. Lisäksi Hubáček ym. (2019) sekä Pappalardo ym. (2019) osoittavat, että vedonlyöntikertoimet säilyvät vahvana vertailukohtana myös kehittyneiden koneoppimismallien rinnalla, mikä korostaa markkinoiden kykyä aggregoida hajautettua informaatiota tehokkaasti.

Täydellistä kalibraatiota ei kuitenkaan saavuteta. Pope ja Peel (1989) sekä Kuypers (2000) havaitsevat systemaattisia poikkeamia implisiittisten todennäköisyyksien ja toteutuneiden tulosten välillä. Constantinou ja Fenton (2013) osoittavat lisäksi, että kalibraatio voi vaihdella kontekstin, kuten liigan ja käytettävissä olevan informaation mukaan. Tämä viittaa siihen, että vedonlyöntimarkkinoiden tehokkuus ei ole staattinen ominaisuus, vaan se kehittyy ajan myötä datan ja analytiikan lisääntyessä. Näin ollen vedonlyöntimarkkinoita voidaan pitää hyvin, mutta ei täydellisesti kalibroituina järjestelminä.

3.2 Systemaattiset vinoumat implisiittisissä todennäköisyyksissä

Sauer (1998) sekä Cain ym. (2000) osoittavat, että implisiittiset todennäköisyydet sisältävät systemaattisia vinoumia. Tunnetuin näistä on favourite-longshot bias, jossa altavastaaajien todennäköisyyksiä yliarvioidaan ja suosikkien aliarvioidaan. Tämä ilmiö tarkoittaa, että vedonlyöntimarkkinoilla hinnat eivät ole täysin neutraaleja todennäköisyysarvioita, vaan niihin sisältyy systemaattisia vääristymiä

Cain ym. (2000) osoittavat vinouman esiintyvän useissa vedonlyöntimarkkinoissa, mikä viittaa sen yleispätevyyteen. Angelini ym. (2022) puolestaan osoittavat, että vinouman voimakkuus voi vaihdella markkinoiden kehittyessä ja analytiikan lisääntyessä. Levitt (2004) sekä Sauer (1998) selittävät vinoumaa sekä käyttäytymistaloustieteen että markkinarakenteen näkökulmasta. Vedonlyöjät suosivat suuria mutta epätodennäköisiä voittoja, kun taas vedonlyöntiyhtiöt voivat hinnoitella kertoimia strategisesti riskien hallitsemiseksi.

Vinoumien olemassaolo siis viittaa siihen, että vedonlyöntimarkkinat eivät ole täysin tehokkaita, vaikka ne sisältävät merkittävän määrän informaatiota (Fama 1970; Pope & Peel 1989). Tämä on keskeinen havainto, sillä se selittää, miksi implisiittiset todennäköisyydet eivät täysin vastaa toteutuneita tuloksia.

3.3 Vedonlyöntimarkkinat tietojärjestelminä ja informaation aggregointina

Wolfers ja Zitzewitz (2004) osoittavat, että ennustemarkkinat toimivat tehokkaina informaation aggregointimekanismeina, joissa useiden toimijoiden tieto yhdistyy yhdeksi kollektiiviseksi ennusteeksi. Chen ym. (2011) tarkastelevat tällaisia markkinoita tietojärjestelminä, joissa käyttäjien tuottama informaatio yhdistyy algoritmisten prosessien kautta. Tämä näkökulma on keskeinen, koska se siirtää tarkastelun yksittäisistä ennusteista kokonaisvaltaiseen järjestelmään.

Vedonlyöntikertoimet voidaan tulkita tämän järjestelmän tuottamaksi ulostuloksi, jossa yhdistyvät vedonlyöntiyhtiöiden analytiikka, vedonlyöjien näkemykset ja reaaliaikainen informaatio (Chen ym. 2011; Angelini ym. 2022). Shmueli ja Koppius (2011) korostavat, että tällaiset järjestelmät toimivat päätöksenteon tukijärjestelminä, joiden arvo perustuu niiden kykyyn tuottaa käyttökelpoisia ennusteita. Tämä tarkoittaa, että vedonlyöntikertoimia voidaan arvioida samalla tavoin kuin muita tietojärjestelmiä, esimerkiksi niiden tuottaman informaation laadun ja luotettavuuden perusteella.

Pappalardo ym. (2019) ja Hubáček ym. (2019) osoittavat, että moderni data-analytiikka on lisännyt markkinoilla käsiteltävän informaation määrää merkittävästi. Tätä tukevat Forrest ym. (2005) sekä

Wolfers ja Zitzewitz (2004), jotka osoittavat, että vedonlyöntikertoimet sisältävät jo lähtökohtaisesti suuren määrän relevanttia informaatiota hajautetun tiedon aggregoinnin kautta. Yhdessä nämä havainnot viittaavat siihen, että markkinamekanismi on tehokas tapa yhdistää kasvavaa informaatiomäärää kollektiiviseksi ennusteeksi.

Kehittyneet tilastolliset menetelmät ja koneoppimismallit mahdollistavat entistä yksityiskohtaisemman otteluanalyysin, mikä on parantanut sekä vedonlyöntiyhtiöiden että vedonlyöjien saatavilla olevan informaation laatua. Tämä kehitys on osaltaan vahvistanut vedonlyöntimarkkinoiden kykyä tuottaa informatiivisia todennäköisyysarvioita.

3.4 Miten vastaavuutta arvioidaan?

Implisiittisten todennäköisyyksien ja toteutuneiden tulosten vastaavuutta tarkastellaan kirjallisuudessa tyypillisesti probabilististen ennusteiden arviointiin kehittyneiden mittarien avulla. Näiden mittareiden keskeinen rooli ei ole pelkästään mitata ennustetarkkuutta, vaan arvioida, kuinka hyvin todennäköisyysarviot heijastavat todellisia lopputuloksia.

Gneiting ja Raftery (2007) korostavat, että probabilististen ennusteiden arvioinnissa keskeistä on kalibraation ja erottelukyvyn välinen tasapaino. Tämä tarkoittaa, että ennusteen tulee sekä vastata toteutuneita frekvenssejä että erotella tehokkaasti eri lopputulosten todennäköisyyksiä. Brier score (Brier 1950) ja log loss -mittari tarjoavat tähän keskeiset työkalut, mutta niiden tulkinta vaatii kontekstin huomioimista.

Vedonlyöntimarkkinoiden kohdalla nämä mittarit osoittavat, että implisiittiset todennäköisyydet suoriutuvat johdonmukaisesti hyvin suhteessa vaihtoehtoisiin ennustemalleihin (Hvattum & Arntzen 2010; Angelini ym. 2022). Tämä viittaa siihen, että markkinat eivät ainoastaan tuota tarkkoja ennusteita, vaan myös kalibroituja todennäköisyysarvioita, jotka vastaavat todellisia lopputuloksia pitkällä aikavälillä.

Shmueli ja Koppius (2011) huomauttavat, että ennustejärjestelmien arvioinnissa tulisi huomioida myös niiden käytännöllinen merkitys. Vedonlyöntimarkkinoiden tapauksissa tämä tarkoittaa, että ennusteiden laatua ei voida arvioida pelkästään tilastollisen tarkkuuden perusteella, vaan myös sen perusteella, kuinka hyvin ne mahdollistavat päätöksenteon parantamisen.

Näin ollen mittareiden perusteella vedonlyöntikertoimista johdetut implisiittiset todennäköisyydet eivät ole ainoastaan tilastollisesti tarkkoja, vaan myös käytännössä informatiivisia. Samalla mittarit

paljastavat pieniä mutta systemaattisia poikkeamia, jotka selittävät, miksi vastaavuus toteutuneisiin tuloksiin ei ole täydellinen.

3.5 Missä määrin vedonlyöntikertoimet vastaavat toteutuneita tuloksia?

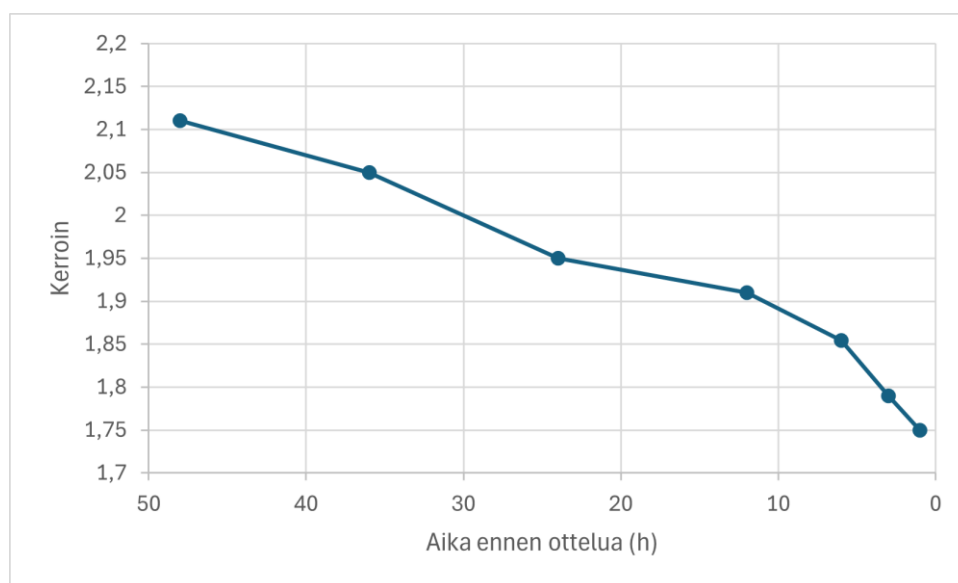
Forrest ym. (2005) ja Angelini ym. (2022) osoittavat, että vedonlyöntikertoimet ovat keskimäärin hyvin kalibroituja ja vastaavat melko tarkasti toteutuneita ottelutuloksia. Toisaalta Cain ym. (2000) ja Sauer (1998) osoittavat systemaattisten vinoumien olemassaolon, mikä estää täydellisen vastaavuuden.

Wolfers ja Zitzewitz (2004) sekä Chen ym. (2011) tarkastelevat vedonlyöntimarkkinoita tietojärjestelminä, jotka aggregoivat hajautettua informaatiota tehokkaasti. Tämä näkökulma auttaa selittämään, miksi vedonlyöntikertoimet ovat yleensä tarkkoja, mutta eivät täydellisiä. Järjestelmä toimii tehokkaasti, mutta siihen vaikuttavat rakenteelliset ja käyttäytymiseen liittyvät rajoitteet.

Näin ollen voidaan todeta, että vedonlyöntikertoimista johdetut implisiittiset todennäköisyydet vastaavat toteutuneita ottelutuloksia suurelta osin, mutta eivät täydellisesti. Ne tarjoavat luotettavan arvon otteluiden lopputuloksista, mutta sisältävät systemaattisia virheitä, jotka johtuvat markkinoiden toimintatavasta, kuten käyttäytymistekijöistä, informaatioepäsymmetrioista ja vedonlyöntiyhtiöiden hinnoittelustrategioista.

4 Vedonlyöntikertoimien reagointi uuteen informaatioon

Vedonlyöntikertoimet päivittyvät ennen ottelua jatkuvasti uuden informaation perusteella.



Kuvio 3 Vedonlyöntikertoimien muutos ennen ottelua

Tämä informaatio voi liittyä esimerkiksi kokoonpanomuutoksiin, pelaajien loukkaantumisiin, sääolosuhteisiin tai ottelun pelipaikkaan. Forrest ym. (2005) sekä Štrumbelj (2014) tarkastelevat näiden tekijöiden vaikutusta ja osoittavat niiden heijastuvan vedonlyöntikertoimiin joukkueiden arvioidun suorituskyvyn muutoksen kautta.

Käytännössä tämä tarkoittaa, että kun markkinoille tulee uutta ja relevanttia tietoa, vedonlyöntikertoimet mukautuvat vastaamaan muuttuneeseen informaatioon. Esimerkiksi tähtipelaajan poissaolo tai odottamaton kokoonpanomuutos voi muuttaa joukkueiden voimasuhteita, mikä näkyy nopeasti kertoimen päivittymisenä (Štrumbelj 2014). Vastaavasti ottelun kontekstiin liittyvät tekijät, kuten pelipaikan muutos tai poikkeavat sääolosuhteet, voivat vaikuttaa joukkueiden suorituskykyyn ja siten myös kertoimiin (Forrest ym. 2005).

Näin ollen vedonlyöntikertoimet eivät kuvaa pelkästään yksittäistä ennustehetkeä, vaan ne heijastavat jatkuvasti päivittyvää arviota ottelun lopputuloksesta, jossa uusi informaatio integroidaan markkinoiden muodostamaan kokonaisnäkemykseen.

4.1 Kokoonpanot ja pelaajakohtainen informaatio

Kokoonpanotiedot muodostavat keskeisen informaatiolähteen vedonlyöntimarkkinoilla, sillä ne tarjoavat suoraa ja ajantasaista tietoa joukkueiden todellisesta vahvuudesta ennen ottelua. Virallisten

kokoonpanojen julkaisu vähentää epävarmuutta joukkueiden suorituskyvystä, minkä seurauksena vedonlyöntikertoimet päivittyvät usein nopeasti uuden tiedon perusteella (Forrest ym. 2005).

Erityisesti avainpelaajan poissaolot, kuten joukkueen kapteenin äkillinen loukkaantuminen voivat vaikuttaa kertoimiin merkittävästi. Tällaiset muutokset ovat usein helposti tulkittavissa ja kvantifioitavissa, koska yksittäisen pelaajan kontribuutiota joukkueen suorituskykyyn voidaan arvioida tilastollisesti (Štrumbelj 2014). Tämä tekee pelaajakohtaisesta informaatiosta erityisen vaikutusvaltaista verrattuna epäsuorempiin tekijöihin.

Simon (2023) osoittaa, että juuri ennen ottelua julkaistava kokoonpanoinformaatio on yksi keskeisimmistä kertoimien muutoksia selittävistä tekijöistä, mikä korostaa informaation ajoituksen merkitystä. Lisäksi Constantinou ja Fenton (2013) osoittavat, että yksittäisten pelaajien vaikutus ottelun lopputulosten todennäköisyyksiin voidaan mallintaa eksplisiittisesti, mikä tukee ajatusta siitä, että markkinat reagoivat erityisen voimakkaasti sellaisiin muutoksiin, joiden vaikutus on selkeästi arvioidavissa.

Näin ollen kokoonpanot ja pelaajakohtainen informaatio eivät ainoastaan lisää markkinoiden käytettävissä olevan informaation määrää, vaan myös vähentävät ennusteisiin liittyvää epävarmuutta. Tämä selittää, miksi juuri tällainen informaatio johtaa usein nopeisiin ja selkeisiin muutoksiin vedonlyöntikertoimissa.

4.2 Joukkueen suorituskyky ja viimeaikainen formi

Jalkapallossa joukkueen viimeaikainen suorituskyky, kuten voittoputket, heikot tulokset tai muutokset pelillisessä tehokkuudessa, vaikuttaa vedonlyöntikertoimiin, koska se tarjoaa informaatiota joukkueen nykyisestä tasosta. Toisin kuin kokoonpanotiedot, tällainen informaatio ei kuitenkaan ole yksiselitteistä, vaan sen tulkinta perustuu useiden otteluiden aikana kertyneeseen dataan (Štrumbelj 2014).

Tämän vuoksi viimeaikaisen suorituskyvyn vaikutus kertoimiin muodostuu usein asteittain. Yksittäinen ottelu ei välttämättä muuta markkinoiden arviota merkittävästi, mutta useamman ottelun muodostama trendi voi johtaa kertoimien systemaattiseen päivittymiseen. Tämä on linjassa Dixon ja Colesin (1997) tutkimuksen kanssa, jossa joukkueiden suorituskykyä mallinnetaan pitkän aikavälin datan avulla ja havaitaan, että aiemmat tulokset ja ottelukohtaiset tekijät vaikuttavat ennusteisiin kumulatiivisesti.

Hubáček ym. (2019) sekä Pappalardo ym. (2019) osoittavat, että moderni data-analytiikka mahdollistaa joukkueiden suorituskyvyn entistä tarkemman arvioinnin hyödyntämällä tilastollisia aineistoja. Tämä on lisännyt viimeaikaiseen suorituskykyyn liittyvän informaation merkitystä vedonlyöntimarkkinoilla, koska arviot voidaan perustaa aiempaa systemaattisempaan ja yksityiskohtaisempaan dataan

Vedonlyöntikertoimien ajallista kehitystä tarkastellut Simon (2023) osoittaa, että markkinat päivittävät arvioitaan jatkuvasti ottelun lähestyessä uuden informaation myötä. Samalla tutkimus korostaa, että tämä päivittyminen ei johda systemaattisesti ennusteiden paranemiseen, mikä viittaa siihen, että markkinoiden kyky hyödyntää kaikkea saatavilla olevaa informaatiota ei välttämättä ole täysin optimaalinen.

Lisäksi Constantinou ja Fenton (2013) osoittavat, että joukkueiden suorituskykyä voidaan mallintaa probabilistisesti huomioimalla sekä pitkän aikavälin taso että lyhyen aikavälin vaihtelu. Tämä tukee käsitystä siitä, että vedonlyöntimarkkinat eivät reagoi pelkästään toteutuneisiin tuloksiin, vaan myös odotuksiin tulevasta suorituskyvystä ja niin sanottuun momentumiin.

Näin ollen viimeaikainen formi vaikuttaa vedonlyöntikertoimiin eri tavalla kuin välitön ja konkreettinen informaatio. Sen vaikutus on usein hitaampi, kumulatiivinen ja osittain tulkinnanvarainen, mutta samalla keskeinen osa markkinoiden muodostamaa kokonaisarviota joukkueiden suorituskyvystä.

4.3 Kotietu, stadion ja sääolosuhteet

Jalkapallossa ottelun kontekstuaaliset tekijät, kuten kotietu ja pelipaikka, vaikuttavat vedonlyöntikertoimiin, koska ne muokkaavat joukkueiden välistä voimasuhdetta ennen ottelun alkua. Kotiedun on havaittu olevan systemaattinen ja tilastollisesti merkittävä etu jalkapallossa ja se sisällytetään vaikiintuneesti jalkapallokertoimiin (Forrest ym. 2005). Kotietu voi ilmetä esimerkiksi joukkueen parempana suoritusasana kotikentällä, mikä liittyy muun muassa yleisön tukeen, matkustamisen puutteeseen ja tuttuun pelialustaan.

Kotiedun vaikutusta on selitetty useilla mekanismeilla. Nevill ja Holder (1999) sekä Pollard (2008) osoittavat, että kotijoukkueen kannattajien läsnäolo voi vaikuttaa sekä pelaajien suoritukseen että tuomarien päätöksentekoon, mikä lisää kotijoukkueen etua. Lisäksi matkustamiseen liittyvät tekijät voivat heikentää vierasjoukkueen suorituskykyä, kun taas kotijoukkue hyötyy tutusta ympäristöstä ja rutiineista.

Kotiedun voimakkuus ei kuitenkaan ole vakio, vaan se voi vaihdella ottelukohtaisesti. Esimerkiksi katsojamäärän on havaittu vaikuttavan kotiedun suuruuteen, sillä suurempi yleisö voi vahvistaa psykologista painetta ja lisätä kotijoukkueen etua (Nevill ym. 2002). Unkelbach & Memmert (2010) osoittavat, että erityisesti äänekäs kotiyleisö voi vaikuttaa tuomareiden päätöksentekoon ja lisätä kotijoukkueelle suotuisia ratkaisuja.

Vastaavasti tilanteissa, joissa ottelu pelataan ilman yleisöä tai neutraalilla kentällä, kotiedun on havaittu heikkenevän merkittävästi, mikä viittaa siihen, että yleisön äänekkyys ja siihen liittyvä psykologinen paine ovat keskeisiä kotiedun selittäviä tekijöitä (Reade ym. 2020; McCarrick ym. 2021). Tätä ilmiötä havainnollistavat erityisesti COVID-19-pandemian aikana pelatut yleisöttömät ottelut, joissa kotiedun systemaattinen heikkeneminen tarjoaa empiiristä näyttöä yleisön merkityksestä ottelun lopputuloksen kannalta.

Myös sääolosuhteet muodostavat keskeisen ottelun kontekstiin liittyvän tekijän, joka vaikuttaa vedonlyöntikertoimiin. Sääolosuhteet, kuten voimakas sade, tuuli tai alhainen lämpötila voivat muuttaa pelin tempoa, lisätä virheiden määrää ja vaikuttaa maalinteon todennäköisyyteen (Štrumbelj 2014; Anderson & Sally 2013). Lisäksi tutkimus viittaa siihen, että sääolosuhteiden vaikutus ei ole kaikille joukkueille samanlainen, vaan se riippuu joukkueen pelityylistä ja tottumuksesta erilaisiin olosuhteisiin (Peeters & Van Ours 2021).

Sääolosuhteiden vaikutus kytkeytyy myös laajempiin pelipaikan erityispiirteisiin, kuten ilmastoon ja pelialustaan. Joukkueet, jotka ovat tottuneet pelaamaan tietyissä olosuhteissa, kuten kylmässä ilmastossa tai tekonurmella, voivat hyötyä näistä tekijöistä kohdatessaan vastustajia, joille vastaavat olosuhteet ovat epätavallisia. Tällaiset erot voivat vaikuttaa esimerkiksi pelin tempoon, tekniseen suoritustasoon ja fyysiseen kuormitukseen, mikä puolestaan voi heijastua ottelun lopputuloksen todennäköisyyksiin. (Andersson ym. 2008; Peeters & Van Ours 2021; Anderson & Sally 2013).

Constantinou ja Fenton (2013) osoittavat, että ottelun lopputulokseen vaikuttavia tekijöitä voidaan mallintaa probabilistisesti huomioimalla myös ulkoiset olosuhteet. Tämä tukee näkemystä siitä, että vedonlyöntimarkkinat eivät huomioi pelkästään joukkueisiin liittyvää informaatiota, vaan myös ottelun ulkoiset olosuhteet osana kokonaisvaltaista ennustejärjestelmää. Lisäksi Simon (2023) osoittaa, että markkinat päivittävät arvioitaan jatkuvasti esimerkiksi tarkentuvien sääennusteiden tullessa.

Näin ollen kotietu, stadionin ominaisuudet ja sääolosuhteet muodostavat yhdessä keskeisen osan vedonlyöntimarkkinoiden hyödyntämässä informaatiosta. Vaikka näiden tekijöiden vaikutus on

usein epäsuorempi kuin kokoonpanoihin liittyvän informaation, ne vaikuttavat systemaattisesti ottelun lopputuloksen todennäköisyyksiin ja siten myös vedonlyöntikertoimien muodostumiseen.

4.4 Valmennus ja organisatoriset muutokset

Jalkapallossa myös valmennukseen ja johtamiseen liittyvä informaatio voi vaikuttaa vedonlyöntikertoimiin, koska se voi muuttaa joukkueen taktista lähestymistapaa, pelityyliä ja päätöksentekoa ottelutilanteissa. Esimerkiksi valmentajan vaihdos voi johtaa muutoksiin kokoonpanovalinnoissa, pelijärjestelmässä ja pelin intensiteetissä, mikä puolestaan vaikuttaa ottelun lopputuloksen todennäköisyyksiin (Bridgewater 2010; Peeters & Van Ours 2021).

Valmentajavaihdokset voivat ilmetä käytännössä esimerkiksi tilanteissa, joissa joukkueen päävalmentaja erotetaan ja joukkuetta johtaa väliaikaisesti varavalmentaja, tai kun uusi valmentaja aloittaa tehtävässään juuri ennen ottelua. Tällaisissa tilanteissa joukkueen pelillinen identiteetti ja taktiset valinnat voivat muuttua nopeasti, mutta muutosten vaikutus ottelun lopputulokseen on usein epävarma. Bridgewater (2010) osoittaa, että valmentajavaihdokset voivat johtaa lyhyen aikavälin suorituskyvyn vaihteluun, mutta vaikutukset eivät ole systemaattisia.

Peeters ja Van Ours (2021) tutkimus viittaa siihen, että valmentajan vaikutus joukkueen suorituskykyyn on kontekstisidonnainen ja riippuu esimerkiksi joukkueen lähtötasosta sekä muutoksen ajoituksesta. Tämän vuoksi valmennukseen liittyvä informaatio on luonteeltaan epävarmempaa kuin esimerkiksi kokoonpano- tai loukkaantumistieto, joiden vaikutus voidaan arvioida suoremmin.

Valmennukseen liittyvän informaation keskeinen haaste on sen vaikea kvantifioitavuus. Hubáček ym. (2019) korostavat, että vedonlyöntimarkkinoilla hyödynnettävän informaation arvo riippuu siitä, kuinka helposti se voidaan muuttaa kvantitatiiviseen muotoon ja sisällyttää ennustemalleihin. Tämän vuoksi selkeästi mitattavat tekijät, kuten pelaajapoissaolot, heijastuvat kertoimiin nopeasti, kun taas vaikeammin selitettävät tekijät, kuten valmennukselliset muutokset, vaikuttavat enemmän siihen, miten markkinat tulkitsevat ja arvioivat joukkueen tulevaa suorituskykyä, mikä voi johtaa hitaampaan ja epäsuorempaan reagointiin (Hubáček ym. 2019).

Vedonlyöntikertoimien ajallista kehitystä tarkastellut Simon (2023) osoittaa, että markkinat päivittävät arvioitaan jatkuvasti uuden informaation perusteella, mutta reagoinnin nopeus ja tarkkuus vaihtelevat informaation luonteen mukaan. Tämä viittaa siihen, että valmennukseen liittyvät muutokset voivat heijastua kertoimiin viiveellä, kun niiden vaikutus joukkueen suorituskykyyn konkretisoituu vasta useamman ottelun aikana.

Tätä tukee myös ennustemallien tutkimus, jossa on havaittu, että vaikeasti mitattavat tekijät, kuten taktiset muutokset ja johtamiseen liittyvät päätökset, sisältyvät ennusteisiin usein epäsuorasti historiallisten tulosten kautta (Constantinou & Fenton 2013). Näin ollen valmennuksellisten muutosten vaikutus kertoimiin ei välttämättä näy välittömästi, vaan heijastuu markkinoiden arvioihin asteittain toteutuneen suorituskyvyn kautta.

Näin ollen valmennukseen ja organisatorisiin muutoksiin liittyvä informaatio vaikuttaa vedonlyöntikertoimiin eri tavalla kuin suoraan mitattavat tekijät. Sen vaikutus on usein epäsuorempi, tulkinnanvaraisempi ja ajallisesti viivästynyt, mutta se muodostaa silti keskeisen osan vedonlyöntimarkkinoiden kokonaisarviota joukkueiden suorituskyvystä ja ottelun lopputuloksesta (Hubáček ym. 2019; Simon 2023).

4.5 Miten eri informaatiotekijät heijastuvat kertoimiin?

Kirjallisuuden perusteella vedonlyöntikertoimet reagoivat uuteen informaatioon systemaattisesti, mutta reagoinnin nopeus ja tarkkuus vaihtelevat informaation luonteen mukaan. Selkeästi mitattavat ja suoraan joukkueiden suorituskykyyn vaikuttavat tekijät, kuten kokoonpanojen muutokset ja pelaajapoissaolot, heijastuvat kertoimiin nopeasti, koska niiden vaikutus voidaan arvioida suhteellisen tarkasti (Štrumbelj 2014; Constantinou & Fenton 2013). Tämä tukee käsitystä siitä, että vedonlyöntimarkkinat kykenevät tehokkaasti hyödyntämään helposti kvantifioitavaa informaatiota.

Sen sijaan epäsuoremmat ja vaikeammin mitattavat tekijät, kuten joukkueen viimeaikainen formi, valmennukselliset muutokset tai ottelun kontekstuaaliset olosuhteet, vaikuttavat kertoimiin usein hitaammin ja tulkinnanvaraisemmin. Näiden tekijöiden vaikutus muodostuu usein asteittain ja riippuu siitä, miten markkinat tulkitsevat niiden merkityksen joukkueiden tulevan suorituskyvyn kannalta (Hubáček ym. 2019; Simon 2023). Tämä viittaa siihen, että vedonlyöntimarkkinat eivät reagoi pelkästään uuden informaation olemassaoloon, vaan myös siihen, kuinka luotettavana ja relevanttina kyseinen informaatio nähdään.

Lisäksi vedonlyöntikertoimien päivittyminen on jatkuva prosessi, jossa uutta informaatiota integroidaan markkinoiden muodostamaan kokonaisarvioon ottelun lopputuloksesta (Forrest ym. 2005; Simon 2023). Tämä prosessi muistuttaa tietojärjestelmätieteessä kuvattuja adaptiivisia järjestelmiä, joissa järjestelmä mukautuu jatkuvasti uuden tiedon perusteella ja tarkentaa ennusteitaan ajan myötä (Chen ym. 2011; Shmueli & Koppius 2011).

Näin ollen vedonlyöntikertoimet voidaan tulkita dynaamisina ja adaptiivisina ennusteina, jotka heijastavat markkinoiden kykyä yhdistää hajautettua ja monimutkaista informaatiota. Vaikka

markkiant reagoivat tehokkaasti moniin informaatiotekijöihin, reagoinnin epätasaisuus eri informaatiotyyppien välillä osoittaa, että kaikki informaatio ei siirry kertoimiin yhtä nopeasti tai täydellisesti. Tämä on keskeinen havainto arvioitaessa vedonlyöntimarkkinoiden toimintaa ja niiden tuottamien todennäköisyysarvioiden luotettavuutta.

5 Yhteenveto ja johtopäätökset

Tutkielman tavoitteena oli tarkastella data-analytiikkaan perustuvien ennustemallien ja vedonlyöntikertoimien suhdetta jalkapallossa aikaisemman kirjallisuuden perusteella. Lisäksi tavoitteena oli arvioida, missä määrin vedonlyöntikertoimista johdetut implisiittiset todennäköisyydet vastaavat toteutuneita ottelutuloksia sekä miten uusi informaatio heijastuu kertoimiin ennen ottelua.

Kirjallisuuskatsauksen perusteella voidaan todeta, että vedonlyöntikertoimet toimivat keskeisenä vertailukohtana ennustemalleille, sillä ne selittävät merkittävän määrän markkinoilla saatavilla olevaa informaatiota. Vedonlyöntimarkkinat voidaan tulkita kollektiivisina ennustejärjestelminä, jotka aggregoivat hajautettua tietoa useista eri lähteistä yhdeksi numeeriseksi arvoksi ottelun lopputuloksesta. Vaikka tilastolliset ja koneoppimiseen perustuvat mallit voivat saavuttaa korkean ennustetarkkuuden, ne eivät useimmiten kykene ylittämään vedonlyöntimarkkinoiden suorituskykyä.

Tutkimuksen keskeinen havainto on, että vedonlyöntikertoimista johdetut implisiittiset todennäköisyydet ovat keskimäärin hyvin kalibroituja ja vastaavat melko tarkasti toteutuneita ottelutuloksia. Täydellistä vastaavuutta ei kuitenkaan saavuteta, sillä markkinoilla esiintyy systemaattisia vironoumia, kuten favourite-longshot bias. Tämä viittaa siihen, että vedonlyöntimarkkinat ovat tehokkaita, mutta eivät täysin virheettömiä.

Tutkielma myös osoittaa, että vedonlyöntikertoimet reagoivat jatkuvasti uuteen informaatioon ennen ottelua. Selkeästi mitattavat tekijät, kuten kokoonpanomuutokset ja pelaajapoissaolot, heijastuvat kertoimiin nopeasti, kun taas vaikeammin tulkittavat tekijät, kuten joukkueen viimeaikainen suorituskyky tai valmennukselliset muutokset, vaikuttavat kertoimiin hitaammin ja epäsuoremmin. Tämä korostaa vedonlyöntimarkkinoiden luonnetta dynaamisina ja adaptiivisina järjestelminä.

Tulokset viittaavat siihen, että vaikka data-analytiikka ja kehittyneet ennustemallit ovat parantaneet ennustetarkkuutta, vedonlyöntimarkkinat mukautuvat nopeasti uuteen informaatioon ja analytiikkaan. Tämän vuoksi yksittäisten mallien on vaikea saavuttaa pysyvää kilpailuetua markkinoihin nähden. Ennustetarkkuuden ja taloudellisen kannattavuuden välinen ero korostuu erityisesti vedonlyöntikontekstissa, jossa markkinarakenteet ja vedonlyöntiyhtiöiden kate rajoittavat systemaattisen ylituoton mahdollisuuksia.

Tutkielman keskeinen havainto on kokonaisvaltaisen kuvan muodostaminen siitä, miten data-analytiikka, ennustemallit ja vedonlyöntikertoimet liittyvät toisiinsa jalkapallossa. Tulokset tukevat näkemystä, jonka mukaan vedonlyöntimarkkinat toimivat tehokkaina informaation yhdistäjinä, mutta niissä esiintyvät pienet tehottomuudet jättävät edelleen tilaa jatkotutkimukselle.

Yhteenvetona voidaan todeta, että vedonlyöntikertoimet tarjoavat vahvan ja käytännössä kilpailukykyisen ennusteen jalkapallo-otteluiden lopputuloksista. Vaikka data-analytiikkaan perustuvat mallit voivat täydentää markkinoiden tuottamaa informaatiota, niiden keskeinen hyöty syntyy usein markkinoiden rinnalla toimimisesta.

Lähteet

- Anderson, C. & Sally, D. 2013. *The Numbers Game: Why Everything You Know About Football is Wrong*. Penguin
- Andersson, H., Ekblom, B. & Krustrup, P. 2008. Elite football on artificial turf versus natural grass. *Journal of Sports Sciences*, 26(2), 113–122. <https://doi.org/10.1080/02640410701422076>
- Angelini, G., De Angelis, L. & Singleton, C. 2022. Forecasting football match outcomes with bookmakers' odds. *European Journal of Operational Research*, 297(1), 305–315.
- Baboota, R. & Kaur, H. 2019. Predictive analysis and modelling football results using machine learning approach. *Journal of Intelligent & Fuzzy Systems*, 36(4), 3203–3214.
- Bridgewater, S. 2010. *Football Management*. Palgrave Macmillan
- Brier, G.W. 1950 *Verification of forecasts expressed in terms of probability*. *Monthly Weather Review*, 78(1), 1–3.
- Brynjolfsson, E. & McElheran, K. 2016. The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133–139.
- Cain, M., Law, D. & Peel, D. 2000. The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy*, 47(1), 25–36. <https://doi.org/10.1111/1467-9485.00151>
- Chen, H., Chiang, R. H. L. & Storey, V. C. 2011. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 35(4), 1165–1188. <https://doi.org/10.2307/41703503>
- Chen, Y., Goes, P. & Marsden, J. R. 2011. Predicting market outcomes in prediction markets: How trading behavior affects market prices. *Information Systems Research*, 22(1), 131–148.
- Dixon, M.J. & Coles, S.G. 1997. Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2), 265–280. <https://doi.org/10.1111/1467-9876.00065>
- Fama, E.F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
- Forrest, D. & Simmons, R. 2008. Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), 119–126. <https://doi.org/10.1080/00036840701522895>
- Forrest, D., Goddard, J. & Simmons, R. 2005. Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21(3), 551–564. <https://doi.org/10.1016/j.ijforecast.2005.03.003>

- Gneiting, T. & Raftery, A.E. (2007) *Strictly proper scoring rules, prediction, and estimation*. *Journal of the American Statistical Association*, 102(477), 359–378.
<https://doi.org/10.1198/016214506000001437>
- Hubáček, O., Šourek, G. & Železný, F. 2019. Learning to predict soccer results from relational data. *Machine Learning*, 108, 29–47.
- Hvattum, L. M. & Arntzen, H. 2010. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26(3), 460–470.
<https://doi.org/10.1016/j.ijforecast.2009.10.002>
- James, G., Witten, D., Hastie, T. & Tibshirani, R. **2021**. *An Introduction to Statistical Learning* (2nd ed.). New York: Springer.
- Kelly, J. L. 1956 A new interpretation of information rate. *Bell System Technical Journal*, 35(4), 917–926. https://www.princeton.edu/~wbialek/rome/refs/kelly_56.pdf
- Kuypers, T. 2000. Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics*, 32(11), 1353–1363. <https://doi.org/10.1080/00036840050151449>
- Levitt, S.D. 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal*, 114(495), 223–246. <https://doi.org/10.1111/j.1468-0297.2004.00207.x>
- McCarrick, D., Bilalić, M., Neave, N. & Wolfson, S. 2021 Home advantage during COVID-19: Analysis of European football leagues. *Psychology of Sport and Exercise*, 56.
<https://doi.org/10.1016/j.psychsport.2021.102013>
- Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D. & Giannotti, F. 2019. A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, 6, 236.
<https://doi.org/10.1038/s41597-019-0247-7>
- Pope, P.F. & Peel, D.A. 1989. Information, prices and efficiency in a fixed-odds betting market. *Economica*, 56(223), 323–341.
- Rathke, A. 2017. An examination of expected goals and shot efficiency in soccer. *Journal of Sports Analytics*, 3(2), 103–114.
- Reade, J. J., Schreyer, D. & Singleton, C. 2020. Eliminating supportive crowds reduces home advantage in association football. <https://doi.org/10.1111/ecin.13063>
- Sauer, R.D. 1998. The economics of wagering markets. *Journal of Economic Literature*, 36(4), 2021–2064.
- Shmueli, G. & Koppius, O. 2011. Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553–572. <https://doi.org/10.2307/23042796>
- Simon, T. (2023). Inefficient forecasts at the sportsbook: An analysis of real-time betting line movement. Saatavilla: <https://www.researchgate.net/publication/372441761>

- Štrumbelj, E. 2014. On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4), 934–943. <https://doi.org/10.1016/j.ijforecast.2014.02.008>
- Unkelbach, C. & Memmert, D. 2010. Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport & Exercise Psychology*, 32(4), 483–498. <https://doi.org/10.1123/jsep.32.4.483>
- Wolfers, J. & Zitzewitz, E. 2004. Prediction markets. *Journal of Economic Perspectives*, 18(2), 107–126. <https://doi.org/10.1257/0895330041371321>
- Wunderlich, F. & Memmert, D. 2020. Are betting returns a useful measure of accuracy in (sports) forecasting? *International Journal of Forecasting*, 36(2), 713–722.

Liitteet

Liite 1 Selvitys tekoälyn käytöstä

Kandidaattitutkielmassa olen hyödyntänyt generatiivisen tekoälyn työkaluja, kuten ChatGPT:tä sekä Scopus AI:ta. Tekoälyä on käytetty erityisesti tutkielman rakenteen suunnittelun tukena, lähteiden etsimisessä, terminologian selkeyttämisessä sekä sopivien sanamuotojen muotoilussa. Lisäksi tekoälyä on hyödynnetty tekstin kielenhuollollisessa tarkastelussa.

Kaikki tutkielmassa esitetyt johtopäätökset ja tulkinnat perustuvat tekijän omaan analyysiin sekä käytettyyn tieteelliseen kirjallisuuteen. Tekoäly on toiminut kirjoitusprosessia tukevana apuvälineenä.