

Patrik Saarelainen

**TIEDONLOUHINTATEKNOLOGIAN
HYÖDYNTÄMINEN
URHEILUVEDONLYÖNNISSÄ**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2020

TIIVISTELMÄ

Saarelainen, Patrik

Tiedonlouhintateknologian hyödyntäminen urheiluedonlyönnissä

Jyväskylä: Jyväskylän yliopisto, 2020, 68 s.

Tietojärjestelmätiede, pro gradu -tutkielma

Ohjaaja: Kyppö, Jorma

Tämä tutkimus on pro gradu -tutkielma, jonka tarkoituksena on selvittää, onko tiedonlouhintateknologian avulla mahdollista laatia vedonlyöntimarkkinaa täsmällisempiä todennäköisyysarvioita tulevien urheilutapahtumien lopputuloksista ja näitä hyödyntämällä saavuttaa rahallista ylituottoa vedonlyöntiin panostamalla. Vedonlyöntimarkkina on vuosi vuodelta kasvava markkina ja vedonlyönnin rooli yhtenä kaupankäynnin muotona kasvaa jatkuvasti. Samaan aikaan datan määrä maailmassa kasvaa hurjaa vauhtia. Myös urheilu tuottaa huomattavia määriä tilastollista informaatiota sen eri muuttujista, joita voidaan käyttää hyödyksi tulevien urheilutapahtumien todennäköisyysarvioita laadittaessa. Tutkielmassa selvitetään, kuinka tätä tilastollista dataa voidaan käyttää hyödyksi urheiluedonlyönnissä. Tutkielmassa havaitaan urheiluedonlyönnin olevan lähempänä perinteistä sijoitustoimintaa kuin uhkapeliä. Lisäksi vedonlyöntimarkkinan katsotaan olevan suhteellisen tehokas, mutta ei kuitenkaan täysin tehokas. Käsitellyt tulokset tiedonlouhintateknologian hyödyntämisestä urheiluedonlyönnissä kertovat, että varmuudella ei voida sanoa tiedonlouhintateknologian tarjoavan kiistatonta hyötyä urheiluedonlyönnissä. Tutkielman kirjallisuuskatsauksessa määritellään alaan liittyvät keskeiset käsitteet sekä esitellään aikaisempaa tutkimustietoa vedonlyöntiin, tiedonlouhintateknologiaan sekä todennäköisyyksiä laskeviin algoritmeihin liittyen. Tutkielman empiirisen tutkimuksen osiossa vastataan tutkimuskysymyksiin numeerisesta havaintoaineistosta tehtyjen laskelmien perusteella sekä pohditaan tutkimustulosten merkitystä.

Asiasanat: urheiluedonlyönti, vedonlyöntimarkkina, urheiluanalytiikka, tiedonlouhinta, tiedonlouhintateknologia

ABSTRACT

Saarelainen, Patrik

Utilization of Data Mining Technologies in Sports Betting

Jyväskylä: University of Jyväskylä, 2020, 68 p.

Information Systems, Master's Thesis

Supervisor: Kyppö, Jorma

This research paper is a Master's Thesis. The aim of this paper is to find out is it possible to estimate probabilities in sports more precisely than sports betting markets do and utilize these estimations to earn extra profit by investing in sports betting. The size of sports betting markets will increase a year after year and the role of sports betting as a one of the investing strategies will increase as well. At the same time there is a huge increasing amount of data in the world. Sport also generates significant amounts of data for different factors which can be utilize when estimating probabilities in sport matches. This study reveals how this statistical data can be utilize in sports betting. This paper finds sports betting to be more investing than gambling. In addition, this paper finds out that sports betting markets are a quite efficient but not fully efficient. The studies covered by this master's thesis show that cannot be said with certainty that data mining technologies would offer undeniable benefits in sports betting. Literature review in this study provides a comprehensive description of key concepts in the field of sports betting and introduce preview studies related to sports betting, data mining technologies and algorithms that estimate probabilities to sport matches. The empirical part of this paper includes answers to research questions based on the calculations of the numerical data and discussion about research results and their significance.

Keywords: sports betting, betting markets, sport analytics, data mining, data mining technologies

KUVIOT

Kuvio 1 Tiedonlouhintaprosessi	24
Kuvio 2 Tilastodatan kerääminen ja varastointi.....	25
Kuvio 3 Tietokannan tähtirakenne	25
Kuvio 4 Monte Carlo -simulaatio	28
Kuvio 5 Kaava osumatarkkuusprosentille.....	30
Kuvio 6 Accuscoren NFL-data.....	40
Kuvio 7 Oddsportal.com -kerroinvertailusivuston näkymä	42
Kuvio 8 Excel-laskelmat	43
Kuvio 9 Esimerkki panostamiseen	45

TAULUKOT

Taulukko 1 Algoritmin tuottama datasetti	27
Taulukko 2 Selitykset taulukossa 1 käytetyille lyhenteille	28
Taulukko 3 Kotiennakkosuosikit.....	46
Taulukko 4 Vierasennakkosuosikit	47
Taulukko 5 Toteutuneet lopputulokset.....	47
Taulukko 6 Vedonlyöntikohteiden jakauma koti- ja vierasjoukkueiden välillä. 47	
Taulukko 7 Odotusarvojen suuruus.....	48
Taulukko 8 Osumatarkkuus.....	49
Taulukko 9 Palautusprosentit	49
Taulukko 10 Panostus.....	51
Taulukko 11 Pelikassan kehitys.....	51

SISÄLLYS

TIIVISTELMÄ.....	2
ABSTRACT.....	3
KUVIOT	4
TAULUKOT	4
SISÄLLYS.....	5
1 JOHDANTO	7
1.1 Tutkimuksen tavoitteet ja tutkimuskysymykset.....	8
1.2 Tutkimuksen rakenne.....	9
2 KIRJALLISUUSKATSAUS.....	10
2.1 Urheiluedonlyönnin perusteet ja keskeiset käsitteet	10
2.1.1 Mitä on urheiluedonlyönti?	10
2.1.2 Vedonlyöntimarkkina	11
2.1.3 Todennäköisyysarvio	11
2.1.4 Kerroin.....	12
2.1.5 Palautusprosentti	13
2.1.6 Odotusarvo.....	13
2.2 Urheiluedonlyönti sijoitusmuotona	14
2.2.1 Uhkapeli	14
2.2.2 Objekttiivinen todennäköisyys	15
2.2.3 Subjekttiivinen todennäköisyys.....	16
2.2.4 Panostaminen.....	17
2.2.5 Urheiluedonlyönnin suhde sijoittamiseen	18
2.2.6 Yhteenvedo.....	19
2.3 Tiedonloughintateknologia urheiluedonlyönnissä.....	20
2.3.1 Vedonlyöntimarkkinan tehokkuus	21
2.3.2 Tiedonloughinta yleisesti	23
2.3.3 Tiedonloughinta urheiluedonlyönnissä	24
2.3.4 Monte Carlo -simulaatio.....	28
2.3.5 Tiedonloughintateknologian hyödyntäminen urheiluedonlyönnissä.....	29
3 KIRJALLISUUSKATSAUKSEN YHTEENVETO JA JOHTOPÄÄTÖKSET	32
3.1 Kirjallisuuskatsauksen yhteenvedo.....	32
3.2 Kirjallisuuskatsauksen johtopäätökset.....	34

4	EMPIIRINEN TUTKIMUS	36
4.1	Tutkimuksen tavoite.....	36
4.2	Tutkimusmenetelmät ja tutkimuksen toteutus	37
4.2.1	Taustatietoja	38
4.2.2	Aineiston keruu	38
4.2.3	Empiirisen tutkimuksen vaiheet	39
4.2.4	Aineiston analysointi.....	44
5	TUTKIMUKSEN TULOKSET	46
5.1	Todennäköisyysarvioiden yhdenmukaisuus vedonlyöntimarkkinan kanssa.....	46
5.2	Odotusarvo.....	47
5.3	Kohteiden osumatarkkuus.....	48
5.4	Palautusprosentti	49
5.5	Pelikassan kehitys.....	50
6	POHDINTA.....	53
6.1	Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?	53
6.2	Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?	55
6.3	Onko urheiluedonlyöntimarkkina tehokas?	56
6.4	Tutkimuksen luotettavuus, yleistettävyyys ja rajoitukset	58
7	YHTEENVETO JA JOHTOPÄÄTÖKSET	59
7.1	Tutkimuksen yhteenveto.....	59
7.2	Tutkimuksen johtopäätökset	60
7.3	Tutkimuksen merkitys	62
7.4	Jatkotutkimusaiheet.....	63
	LÄHTEET	64

1 Johdanto

Urheiluviedonlyönti on maailmanlaajuinen teollisuuden ala, jonka suosio on kasvanut viime vuosina räjähdysmäisesti. Internet on avannut urheiluviedonlyöntiin uusia mahdollisuuksia, jotka ovat lisänneet viedonlyönnin määrää verkossa. Vuonna 2015, kansainvälisen viedonlyöntimarkkinan kooksi arvioitiin 37,91 miljardia yhdysvaltain dollaria. Ennusteen mukaan vuonna 2020 rahamäärän urheiluviedonlyöntimarkkinoilla on arvioitu kasvavan 59,79 miljardiin yhdysvaltain dollariin (Statista, 2015).

Urheiluviedonlyönti on näennäisen helppoa - muutaman urheilutapahtuman lopputuloksen oikein ennustamista. Tehtävä on kuitenkin todellisuudessa vaikea (Vuoksenmaa, 2016). Urheilutapahtumien lopputulosten ennustaminen on nykyaikana suosittua urheilunseuraajien kesken ympäri maailman, osittain viedonlyönnistä johtuen. Tänä päivänä urheilu tuottaa huomattavia määriä tilastollista informaatiota muun muassa pelaajista, joukkueista, otteluista sekä sarjakaudesta (Haghighat, Rastegari & Nourafza, 2013). Tämän tutkielman tarkoituksena on selvittää, kuinka tätä tilastollista dataa voidaan hyödyntää urheiluviedonlyönnissä.

Tiedonlouhintateknologiaa urheiluviedonlyönnin saralla ovat aiemmin tutkineet muun muassa Kahn (2003), joka kehitti amerikkalaisen jalkapallon NFL:n otteluiden lopputuloksia ennustavan algoritmin. Buursma (2010) puolestaan kehitti tutkimuksessaan jalkapallo-otteluiden todennäköisyyksiä laskevan algoritmin. Cao (2012) sen sijaan tutki neljää eri tiedonlouhintamenetelmää koripallo-otteluiden lopputulosten arviointiin. Tiedonlouhintateknologian hyödyntämistä urheilutapahtumien lopputulosten ennakoinnissa tutkivat tutkimukset ovat noudattaneet usein samaa kaavaa. Tutkimuksissa erinäisistä muutujista on luotu algoritmi, joka pyrkii laskemaan todennäköisyyssarvioita ennalta määritettyihin urheilutapahtumiin.

Aikaisemmissa tiedonlouhintateknologian hyödyntämistä urheilutapahtumien lopputulosten ennustamisessa käsittelevissä tutkimuksissa on arvioitu todennäköisyyksiä erinäisiin urheilutapahtumiin. Näissä aiemmissa tutkimuksissa ei kuitenkaan ole menty niin pitkälle, että laadittuja todennäköisyyssarvioita olisi hyödynnetty nimenomaan urheiluviedonlyönnin puitteissa. Tämä pro

gradu -tutkielma on ensinnäkin ensimmäinen suomenkielinen tutkimus tiedonlouhintateknologian hyödyntämistä nimenomaan urheiluedonlyönnissä käsittelevä tutkimus. Lisäksi tämä tutkielma on myös ensimmäinen, jossa syvennytään urheiluedonlyöntiin toteuttaen urheiluedonlyönnin prosessi oikeaoppisesti vaihe vaiheelta siten, että tutkimuksessa käytetään tiedonlouhintateknologiaa hyödyntämällä tuotettua dataa todennäköisyysarvioista. Tutkielman tarkoituksena on tarjota uutta tutkimustietoa siitä, onko dataa keräämällä ja sitä urheiluedonlyönnin tarpeisiin jalostamalla kyennyt historiallisesti tekemään ylituottoa vedonlyöntimarkkinoilla.

Tässä pro gradu -tutkielmassa perehdytään tiedonlouhintateknologian hyödyntämiseen urheiluedonlyönnissä. Tutkielmassa määritellään ensin alaan olennaisesti liittyvät käsitteet, esitellään aikaisempaa vedonlyönnistä sekä tiedonlouhintateknologiasta tehtyä tutkimusta sekä vastataan tutkimuksen tutkimuskysymyksiin ensin kirjallisuuskatsauksen ja tämän jälkeen empirisen tutkimuksen puitteissa. Pro gradu -tutkielman tarkoituksena on havainnollistaa sekä lisätä ymmärrystä aiheesta, josta aikaisempaa tutkimustietoa ei vielä ole olemassa. Tutkimus tarjoaa syvällisempää ymmärrystä tiedonlouhintateknologian hyödyntämisestä urheiluedonlyönnissä, alaan liittyvistä keskeisistä käsitteistä sekä toimintamalleista. Tutkielmassa avataan todennäköisyyslaskennan prosessia, todennäköisyysarvioita sekä niiden käyttämistä urheiluedonlyönnin tarpeisiin.

1.1 Tutkimuksen tavoitteet ja tutkimuskysymykset

Tutkielman tavoitteena on selvittää, onko tiedonlouhintateknologian avulla mahdollista laatia vedonlyöntimarkkinaa täsmällisempiä todennäköisyysarvioita tulevien urheilutapahtumien lopputuloksista ja näitä hyödyntämällä saavuttaa rahallista ylituottoa vedonlyöntiin panostamalla. Tutkimus toteutetaan kvantitatiivisena eli määrällisenä tutkimuksena, jonka tarkoituksena on tuottaa yleistettävissä olevia tutkimustuloksia. Tutkimuksen päätutkimuskysymys on:

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?

Päätutkimuskysymykseen vastaamalla on tarkoituksena tarjota uutta tutkimustietoa siitä, onko dataa keräämällä ja sitä urheiluedonlyönnin tarpeisiin jalostamalla kyennyt historiallisesti tekemään ylituottoa vedonlyöntimarkkinoilla. lähdetään etsimään dataan perustuen ja siitä laskelmia tehden.

Tutkimuksen kohteena olevan aiheen syvällisemmän ymmärryksen hankkimiseksi sekä aineiston analysoimisen helpottamiseksi päätutkimuskysymyksen rinnalle tutkimukseen määritellään kaksi apututkimuskysymystä:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?

- Onko vedonlyöntimarkkina tehokas?

Tutkimuksen päättökysymykseen sekä kahteen apututkimuskysymykseen lähdetään ensin hakemaan vastauksia kirjallisuuskatsauksen puitteissa, jossa lähdemateriaalina käytetään aikaisempia vedonlyöntialaa käsitteleviä tutkimuksia. Kirjallisuuskatsauksen tarkoituksena on antaa hyvä teoriapohja tutkielman empiirisen tutkimuksen osioon, jossa tutkimuskysymyksiin etsitään vastauksia numeerisen havaintoaineiston perusteella erinäisiä laskelmia suorittaen.

1.2 Tutkimuksen rakenne

Tutkimus koostuu kahdesta osiosta: Kirjallisuuskatsauksesta sekä empiirisen tutkimuksen osiosta. Tutkielman ensimmäinen osio toteutetaan systemaattisena kirjallisuuskatsauksena, jossa esitetyt määritelmät sekä havaitut tulokset perustuvat lähdekirjallisuuteen. Kirjallisuuskatsauksessa määritellään ensin urheiluvedonlyönti sekä aiheeseen olennaisesti liittyvät keskeiset käsitteet. Tämän jälkeen kuvataan urheiluvedonlyönnin teoriaa sekä konkreettisten esimerkkien kautta avataan muuta alaan liittyvää metodologiaa. Lisäksi kirjallisuuskatsauksessa perehdytään tiedonlouhintateknologiaan käsitteenä sekä esitellään, kuinka tätä voidaan käyttää hyödyksi urheiluvedonlyönnissä.

Kirjallisuuskatsauksen tavoitteena on tarjota lukijalle hyvä perusymmärrys urheiluvedonlyönnistä sekä siihen liittyvistä keskeisistä käsitteistä ja antaa täten hyvä teoriapohja tutkielman empiirisen tutkimuksen osioon. Tutkielman kolmannessa luvussa käydään läpi kirjallisuuskatsauksen yhteenvetoa sekä esitellään kirjallisuuskatsauksen johtopäätökset.

Tämän pro gradu -tutkielman toinen osio on empiirisen tutkimuksen osuus, joka suoritetaan tutkielman neljännessä ja viidennessä luvussa. Empiirisen tutkimuksen osiossa esitellään ensin tutkimuksen tavoitteet, tutkimusmenetelmät sekä tutkimuksen toteutustavat. Tutkielman viidennessä luvussa esitellään empiirisen tutkimuksen tulokset, joiden perusteella vastataan tutkielman päättökysymykseen sekä kahteen apututkimuskysymykseen.

Tutkielman kuudennessa, pohdintaluvussa vastataan tutkielmassa esitettyihin tutkimuskysymyksiin tutkimustulosten perusteella. Kuudes luku sisältää myös pohdintaa tutkimustuloksiin liittyen. Lisäksi luvussa pohditaan tutkimuksen luotettavuutta, yleistettävyyttä sekä rajoituksia.

Pro gradu -tutkielman seitsemännessä ja viimeisessä luvussa esitellään tutkimuksen yhteenveto sekä johtopäätökset. Lisäksi luvussa keskustellaan tutkimuksen merkityksestä sekä esitellään mahdollisia jatkotutkimusaiheita.

2 Kirjallisuuskatsaus

Tutkielman toinen pääluku on toteutettu systemaattisena kirjallisuuskatsauksena. Luvussa määritellään urheiluedonlyönti, kuvataan sen teoriaa sekä avataan tutkielmassa käytettäviä vedonlyöntialaan olennaisesti liittyviä käsitteitä. Luvussa käytetään konkreettisia esimerkkejä helpottamaan urheiluedonlyönnin lainalaisuuksien ymmärtämistä. Kirjallisuuskatsauksessa käydään läpi aikaisempaa tutkimustietoa urheiluedonlyönnin saralta ja näiden kautta pyritään vastaamaan tutkielman tutkimuskysymyksiin kirjallisuuskatsauksen puitteissa.

2.1 Urheiluedonlyönnin perusteet ja keskeiset käsitteet

2.1.1 Mitä on urheiluedonlyönti?

Kun puhutaan urheiluedonlyönnistä, tarkoitetaan sillä prosessia, jossa ennustetaan ottelun tulos ja asetetaan panokset ennakko-odotusten mukaan. Vedonlyönnin tapauksessa on kaksi osapuolta: vedonvälittäjät, jotka määrittävät hinnat ja pelaajat, jotka panostavat erinäisiä urheilutapahtumia annetuilla kertoimilla. Riippuen urheilutapahtuman tuloksesta, toinen kahdesta – vedonvälittäjä tai pelaaja – voittaa ja toinen häviää (Ioulianou ym., 2011.)

Malaskan ja Virtasen (2007) mukaan urheiluedonlyönti on kertaluonteinen todennäköisyyspeli, jossa oletetaan, että tietty tapahtuma, joka on vedonlyönnin kohteena joko tapahtuu tai jää tapahtumatta. Esimerkkinä he käyttävät hevoskilpailua, jossa pelaaja lyö vetoa tietyn hevosen voitosta tietyllä voittokerroimella.

2.1.2 Vedonlyöntimarkkina

Urheiluvedonlyöntimarkkina muodostuu lukuisista eri vedonvälittäjistä, jotka määrittävät kertoimet sekä vedonlyöjistä, jotka lyövät vetoa annetuilla kertoimilla.

Vedonvälittäjillä on yksi ainoa tavoite: tehdä tuottoa tarjoamallaan kertoimilla (Xu, 2011). Tämä ajaa heidät asettamaan kertoimet riittävän korkeiksi säilyttääkseen kilpailukykyensä markkinoilla, mutta riittävän mataliksi, jotta vedonlyöntitoiminta olisi tuottavaa niille (Šikonja, Štrumbelj, 2010).

Makropoulou ja Markellos (2011) esittävät, että informaation hallinnan suhteen on olemassa kolmea erilaista vedonlyöjätyyppiä. Heidän mukaansa on olemassa heikosti informoituja vedonlyöjiä, hyvin informoituja vedonlyöjiä sekä vedonlyöjiä, jotka omaavat jonkinlaista sisäpiirin tietoa. Heidän mukaansa heikosti informoidut pelaajat ovat niin kutsuttuja viihdevedonlyöjiä, jotka eivät lyö vetoa saatavilla olevaan informaatioon perustuen, vaan pelaavat pikemminkin tunteella. Makropoulou ja Markellosen mukaan hyvin informoitujen vedonlyöjien eli järkivedonlyöjien ja viihdevedonlyöjien välinen ero syntyy hyvin informoitujen vedonlyöjien kyvystä prosessoida tietoa. Heidän mukaansa niin kutsutut sisäpiiriläiset perustavat vetonsa heillä olevaan sisäpiiritietoon, jota puolestaan järkivedonlyöjillä tai vedonvälittäjillä ei ole. Makropoulou ja Markellos esittävät, että vedonvälittäjät sekä järkivedonlyöjät kuuluvat samalle tasolle informaation hallinnan suhteen, sillä heidän tietonsa perustuu julkisesti saatavilla olevaan tietoon.

Niin pitkään kuin vedonlyöjän mieltymykset ja näkemykset ovat puolueettomia, vedonvälittäjät tekevät parhaansa asettaakseen informaationallisesti tehokkaat kertoimet, jotka heijastavat kyseisen urheilutapahtuman todellista todennäköisyysarviota (Flepp, Nuesch, Franck, 2016). Muussa tapauksessa vedonvälittäjät saattavat kärsiä suuria tappioita, mikäli vedonlyöjät havaitsevat ja käyttävät hyväkseen vääristyneistä todennäköisyysarvioista johdettuja kertoimia (Levitt, 2004).

2.1.3 Todennäköisyysarvio

Todennäköisyystarkastelussa pelitilanne formalisoidaan kertailmiöiden subjektiivisen todennäköisyyden lähestymistavan eli ns. Dutch Book -säännösten muotoon (Malaska & Virtanen, 2007).

Tarkastellaan vedonlyöntitilannetta, missä välittäjä tarjoaa vetoa tapahtuman E esiintymisestä (tai ei-esiintymisestä) tietyssä hyvin määritellyssä tilanteessa voittosuhteella q . Mikäli E esiintyy, välittäjä maksaa vedonlyöjälle sovitun voittosumman x . Hyväksyessään vedon vedonlyöjä maksaa välittäjälle etukäteen pelipanoksen, joka on p :s osa x :stä eli määrän px , missä $p < 1$. Mikäli E ei esiinny, välittäjä ei maksa mitään vedonlyöjälle, vaan pitää pelipanoksen. Luku p kuvastaa vedonlyöjän subjektiivista näkemystä voittotapahtuman E esiintymisen uskottavuudesta. Se on E:n toteutumista koskevaa tiedontilaa (state of the knowledge) kuvaava luku, vedonlyöjän uskon voimakkuuden ilmaus. Lukua p kutsutaan tapahtuman E todennäköisyydeksi, ja vedonlyöjän

sanotaan tällöin lyöväan vetoa voittomahdollisuudella $(1 - p)/p : 1$ (Malaska, Virtanen, 2007.)

Hajekin (2008) mukaan Dutch Book säännöstö olettaa, että vedonlyöjän laatima todennäköisyysarvio kulkee samassa linjassa panoksen kanssa. Tämä tarkoittaa sitä, että mikäli vedonlyöjä uskoo väitteen E toteutuvan varmuudella, hän lyö vetoa väitteen E puolesta millä tahansa voittokertoimella (Bradley, Leitgeb, 2006).

Jos vedonlyöjä on arvioinut joukkueen A voittotodennäköisyydeksi 50%, hän panostaa kohteeseen siinä tapauksessa, mikäli saa kertoimeksi vähintään 2,00. Jos vedonlyöjä puolestaan on arvioinut joukkueen A voittavan ottelun keran kolmesta eli noin 33% todennäköisyydellä, hän panostaa kohteeseen vain silloin, kun saa kertoimeksi vähintään 3,00, ja niin edelleen.

2.1.4 Kerroin

Urheiluviedonlyönnissä pelikohteen tuotto määräytyy kertoimen mukaan. Kun ymmärretään miten kertoimet muodostuvat, voidaan määrittää voittava taktiikka. Kertoimet muodostuvat tapahtuman todennäköisyyden käänteislukuna. Esimerkiksi kun jonkin tapahtuman todennäköisyys on 50%, silloin kertoimeksi saadaan 2 (Annala, 2014.)

Kertoimella tarkoitetaan vedonvälittäjän tarjoamaa lukua, jonka perusteella vedonlyöjä tekee oman pelipäätöksensä. Jos vedonlyöjä esimerkiksi saa kohteelle kertoimeksi 2,2, ja hän panostaa kohteeseen 10 euroa, hän saa kohteen mennessä oikein 22 euroa. Tällöin hänen voittonsa on 12 euroa (22 - 10) (Virtanen, Vänni, 2005.)

Seuraavat esimerkissä käyttämäni kertoimet ovat otettu kotimaisen Veikkauksen sivuilta 28.1.2020 (Veikkaus, 2020). Veikkaus tarjoaa amerikkalaisen jalkapalloilun NFL:n Super Bowl-otteluun Kansas City Chiefs - San Francisco 49ers kertoimia 1,80 - 2,00. Tämä tarkoittaa sitä, että Veikkaus yrittää myydä asiakkailleen edellä mainittuja todennäköisyysarvioita, jotka on johdettu kertoimen muotoon. Koska todennäköisyysarvio on kertoimen käänteisluku, saadaan kyseisen pelikohteen todennäköisyysarviot laskettua seuraavasti:

- 1 = Kansas City Chiefs (1/1,80) -> 0,556
- 2 = San Francisco 49ers (1/2,00) -> 0,500

Mikäli vedonlyöjän laatima todennäköisyysarvio ylittää yllämainitussa esimerkissä käytetyn Veikkauksen todennäköisyysarvion, kannattaa hänen sijoittaa rahaa kyseiselle merkille. Tällöin kyseessä on niin kutsuttu ylikerroin eli kerroin, joka on suurempi kuin sitä vastaavan todennäköisyyden käänteisluku (Virtanen, Vänni, 2005). Päinvastoin alikertoimeksi puolestaan kutsutaan kerrointa, joka on pienempi kuin sitä vastaavan todennäköisyyden käänteisluku.

Edellisessä esimerkissä täytyy kuitenkin ottaa huomioon, että Veikkauksen kohteeseen tarjoamien todennäköisyysarvioiden summa ylittää luvun 1. Tämä johtuu siitä, että Veikkaus ei palauta vedonlyöjille jokaista kohteeseen

pelattua euroa, vaan ottaa pelivaihdosta ensin itselleen tietynsuuruisen osuuden. Kyse on vedonvälittäjän palautusprosentista, jonka turvin he pyörittävät liiketoimintaansa.

2.1.5 Palautusprosentti

Vedonvälittäjän palautusprosentti kertoo vedonvälittäjän teoreettisen tuotto-odottaman suhteessa vedonlyöntitapahtuman liikevaihtoon. Vedonvälittäjän palautusprosentti saadaan selville laskemalla vedonlyöntitapahtuman lopputuloksille asetettujen kertoimien käänteislukujen summa (Lahdensuo, 2015.)

Edelleen edellisessä alaluvussa laadittua esimerkkiä käyttäen, Veikkauksen teoreettinen tuotto-odottama Kansas City Chiefsin ja San Francisco 49ersin väliseen Super Bowl -otteluun saadaan selville laskemalla ottelun eri lopputuloksille asetettujen kertoimien käänteislukujen summa. Veikkauksen teoreettinen palautusprosentti kyseisessä ottelussa on $0,556 + 0,500 = 1,0556$. Tässä tapauksessa Veikkaus tekee kyseisessä pelikohteessa siis keskimäärin noin viiden ja puolen prosentin tuoton.

Vedonvälittäjän palautusprosentista voidaan johtaa myös vedonvälittäjän tarjoama palautusprosentti, joka kertoo, kuinka paljon se keskimäärin maksaa voittoja suhteessa asetettuihin panoksiin. Toisin sanoen se kertoo, miten paljon vedonlyöjät keskimäärin saavat voittoina takaisin pelaamistaan panoksista. Vedonvälittäjän tarjoama palautusprosentti on vedonvälittäjän teoreettisen palautusprosentin käänteisluku (Lahdensuo, 2015.) Käyttäen edelleen edellistä esimerkkiä alaluvusta 2.1.4, Veikkauksen tarjoama palautusprosentti esimerkki-kohteessa on $1/1,0556 = 0,947$.

Mikäli oletetaan, että Veikkauksen laatima todennäköisyysarvio kyseisessä pelikohteessa on täysin oikea, vedonlyöjät saavat keskimäärin voittoina takaisin pelaamistaan panoksista noin 94,7 prosenttia.

Jorma Vuoksenmaa havainnollistaa Pelaajalta pelaajille (2016) kirjassaan palautusprosentin merkitystä rulettiesimerkkiä hyödyntäen: "Ruletissa pelaaja saa kertoimen 36 pelatessaan yhtä numeroa. Jos pelaaja saisikin kertoimen 38, ei kasinoteollisuutta olisi olemassakaan. Syy tähän on osumistodennäköisyydessä, joka on ruletissa $1/37$. Kertoimella 36 pelaaja saa takaisin $36/37$ (97,3 prosenttia) panoksistaan. Kertoimella 38 pelaaja saisi takaisin $38/37$ (noin 103 prosenttia) sijoituksistaan. Tämä pienen pieni eri palautusprosentteissa on luonut Las Vegasin kaltaisia kaupunkeja ja ajanut lukemattomia ihmisiä itsemurhan partaalle – ja sen yli."

2.1.6 Odotusarvo

Odotusarvo kertoo, mikä olisi tiettyä matemaattista mallia noudattavan kokeen keskiarvo, jos koetta toistettaisiin äärettömän monta kertaa (Karvinen, 2015). Vedonlyönnissä odotusarvo saadaan laskemalla vedonvälittäjän tarjoaman kertoimen ja vedonlyöjän oman todennäköisyysarvion tulo. Rationaalinen vedon-

lyöjä pelaa ainoastaan pelejä, joiden odotusarvo ylittää yhden (Virtanen, Vänni, 2005).

Jos kahdella osapuolella on eriävät todennäköisyysarviot tietyn epävarman tapahtuman toteutumisesta, he voivat lyödä asiasta vedon, joka tarjoaa heille molemmille positiivisen odotusarvon (Sebenius, Geanakoplos, 1983). Esimerkissä toisena osapuolena toimii vedonvälittäjä ja toisena osapuolena puolestaan vedonlyöjä.

Käytetään edelleen esimerkkinä Veikkauksen tarjoamaa Super Bowl -ottelua Kansas City Chiefs – San Francisco 49ers kertoimilla 1,80 – 2,00. Mikäli vedonlyöjä on arvioinut San Franciscon voittavan ottelun 45% todennäköisyydellä ja päättäisi sijoittaa joukkueen voittoon Veikkauksen tarjoamalla kertoimilla, saisi hän vedolleen odotusarvon $(0,45 \times 2,00)$ 0,9. Tällöin, odotusarvon ollessa alle 1, hän tulisi panostaneeksi alikertoimeen, eikä vedon lyöminen olisi järkevää. Mikäli vedonlyöjä sen sijaan olisi arvioinut San Franciscon voiton todennäköisyydeksi 55%, hän saisi vedolleen odotusarvon $(0,55 \times 2,00)$ 1,10. Tässä tapauksessa kyseessä olisi ylikerroin, johon rationaalisen vedonlyöjän kannattaisi sijoittaa rahaa.

2.2 Urheiluedonlyönti sijoitusmuotona

Tutkielman kirjallisuuskatsauksen toisessa alaluvussa käydään läpi uhkapelaamisen ja rationaalisen urheiluedonlyönnin välisiä keskeisiä eroja. Tässä luvussa pyritään tekemään selvä ero uhkapelaamisen ja urheiluedonlyönnin välille ja tutkitaan vedonlyöntiä nimenomaan sijoittamisen näkökulmasta.

Luvussa määritellään ensin lähdekirjallisuuden avulla uhkapelin käsite. Tämän jälkeen pohditaan uhkapelaamiselle ominaisia piirteitä ja verrataan urheiluedonlyönnin eroavaisuuksia esitettyihin määritelmiin.

Käsitteiden määrittelyn jälkeen esitellään joitakin sijoittamisen lainalaisuuksia ja mietitään niiden soveltuvuutta urheiluedonlyönnin saralle, pyrkien rinnastamaan rationaalinen urheiluedonlyönti yhdeksi sijoittamisen muodoksi. Luvun lopussa vastataan toiseen apututkimuskysymykseen: ”Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?”

2.2.1 Uhkapeli

Helenius (2016) kirjoittaa, että Suomen arpajaislaki (23.11.2001/1047,2 §) määrittelee uhkapelin käsitteen seuraavasti: ”Uhkapelillä tarkoitetaan veikkaus-, bingo-, toto- ja vedonlyöntipeliä, raha- tai tavara-arpajaisia ja pelikasinotoimintaa sekä muuta vastaavaa peliä tai toimintaa, jossa voiton saaminen perustuu kokonaan tai osittain sattumaan taikka peliin tai toimintaan osallistuvista riippumattomiin tapahtumiin ja jossa mahdollinen häviö on ilmeisessä epäsuhteessa ainakin jonkun osallistujan maksukykyyn.”

Silvennoisen (2012) mukaan uhkapeliksi luokitellaan peli, jossa panoksena on yleensä rahaa tai jotain muuta arvokasta materiaalia ilman tietoa siitä, miten pelissä tulee lopulta käymään.

Buchdahl (2003) esittää, että uhkapelaamisen ja sijoittamisen välillä on kaksi keskeistä eroavaisuutta. Toinen liittyy riskille altistumiseen. Uhkapelissä pelaaja menettää usein koko panoksensa, mikäli on väärässä. Sijoittamisessa sen sijaan on erittäin epätodennäköistä menettää kaikki sijoituksiin panostetut varat. Toinen ilmeinen eroavaisuus uhkapelaamisen ja sijoittamisen välillä on Buchdahlin mukaan spekulatiota ajan suhteen. Perinteisen sijoitustoiminnan tapauksessa puhutaan usein viikkojen, kuukausien tai vuosien aikajänteestä. Uhkapelaamisessa sen sijaan lopputulema on selvillä maksimissaan muutaman tunnin kuluttua. Buchdahl on havainnut, että uhkapeli on korkeariskistä, lyhytkestoista spekulatiota, kun puolestaan sijoittamisen hän näkee matalariskisenä sekä pitkäkestoisena toimintana.

Taitopeli on peli, jossa pelaaja voi vaikuttaa lopputulokseen erilaisilla fyysisillä ja henkisillä tekniikoilla. Nämä tekniikat kehittyvät usein pelatessa, eikä taitopelin pelaaminen vaadi muuta kuin sääntöjen tuntemisen. Taitopeleissä taitavimmilla pelaajilla on todennäköisempi mahdollisuus voittaa kokemattomampia pelaajiin verrattuna. Urheiluvedonlyöntiä voi pitää taitopelinä, sillä tilastoja tarkastellen, laskelmia tehden ja korkeita kertoimia etsimällä voi tehdä elannon (Helenius, 2016.)

Oikeaoppisesti toteutettua urheiluvedonlyöntiä voidaankin pitää taitopelinä, ei niinkään uhkapelinä. Myös urheiluvedonlyönnissä sattumalla on oma osuutensa, mutta taitava vedonlyöjä pystyy minimoimaan sen merkitystä. Näitä keinoja esitellään lisää tutkielman myöhemmässä vaiheessa.

2.2.2 Objektiivinen todennäköisyys

Tietyn joukkueen objektiivinen voiton todennäköisyys määritellään, kun tapahtuma toistetaan äärettömän monta kertaa (Ali, 1977). Tapahtuman objektiivinen todennäköisyys voidaan määrittää esimerkiksi kolikonheitossa tai ruletissa, mutta urheiluvedonlyönnissä tämä on luonnollisesti mahdottomuus. Vedonlyönti on olemukseltaan aivan erilainen kuin onnenpelit, kuten ruletti. Ruletti on "luonnonvastainen peli", jossa todennäköisyydet tiedetään etukäteen (Figlewski, 1979.)

Kun informaatio kahden epävarman tapahtuman välillä on symmetrinen, todennäköisyydet niiden toteutumiselle asetetaan yhtä suuriksi. Mikäli tapahtumat ovat toisiaan täydentävät, todennäköisyys näille molemmille tapahtumille on $1/2$, riippumatta siitä onko informaatiota niukasti vai runsaasti (Schmeidler, 1989.)

Esimerkkinä voidaan ajatella kolikonheittoa. Saatavilla olevaan informaatioon nähden klaavan ja kruunan todennäköisyyttä voidaan pitää yhtä suurena. Klaava ja kruuna myös täydentävät toisiaan mahdollisesti toteutuvina tapahtumina, joten sekä klaavan että kruunan todennäköisyys toteutua on $1/2$.

Vedonlyöjän kannattaisi sijoittaa rahaa joko klaavan tai kruunan toteutumisen puolesta, mikäli todennäköisyysarvion käänteisluku eli kerroin tapahtuman toteutumiselle olisi vähintään 2,00. Vedonvälittäjät eivät kuitenkaan objektiiviseen todennäköisyyteen perustuvasta kolikonheitosta tällaista kerrointa tarjoaisi, sillä se olisi heille pitkällä aikavälillä tappiollista toimintaa.

Jos ajatellaan, että vedonvälittäjän pelaajille tarjoama palautusprosentti on 95%, se asettaa molemmille, kruunalle sekä klaavalle kertoimen 1,90. Kun kolikonheitto toistetaan äärettömän monta kertaa, vedonvälittäjä kerää pelivaihdosta itselleen 5% tuoton. Vedonvälittäjän saama 5% tuotto puolestaan kerätään vedonlyöjien, tai pikemminkin uhkapelaajien taskuista. Tässä kolikonheitoesimerkin tapauksessa klaavaan tai kruunaan panostaminen 1,90 olisi alikertoimeen sijoittamista, joka puolestaan tarkoittaa sitä, että pelaaja jää pitkällä aikavälillä pelissä tappiolle.

Objektiiviseen todennäköisyyteen perustuvat pelit, mukaan lukien muun muassa kaikki kasinojen tarjoamat pelit ruletista peliautomaatteihin ovat vedonvälittäjälle voitollisia, sillä tapahtumat voidaan toistaa äärettömän monta kertaa. Tämä mahdollistaa todennäköisyyksien määrittämisen täysin tarkaksi ja palautusprosentin säätämisen halutulle tasolle, jolloin pelistä tulee pelaajalle pitkällä aikavälillä tappiollista.

Objektiiviseen todennäköisyyteen perustuvat pelit täyttävät myös uhkapelin merkityksen. Kolikonheiton kaltaisissa peleissä ei ole tietoa siitä, miten pelissä tulee lopulta käymään. Esimerkissä käytetty palautusprosentti 95 on kolikonheitossa pelaajalle ylitsepääsemätön, eikä tämä voi omilla tiedoillaan tai taidoillaan vähentää sattuman merkitystä.

2.2.3 Subjektiiivinen todennäköisyys

Subjektiiivinen todennäköisyys tarkoittaa päättelyä suorittavan subjektin tiedontilaa eikä mitään, mitä voitaisiin objektiivisesti mitata (Malaska, Virtanen, 2007). Malaskan ja Virtasen (2007) mukaan satunnaisuutta sisältävän ilmiön uskottavuuden asteen määrittämisessä on viime kädessä kysymys aina yhdestä kertatapahtumasta ja sen uskottavuutta koskevasta päätöksentekijän subjektiivisesta tietämyksestä, mikä ei tarvitse olla yhdenmukainen muiden henkilöiden tietämyksen kanssa. He mainitsevat tutkimuksessaan, että subjektiivisen todennäköisyyden piirissä eriävä subjektiivinen tietämys ei merkitse epärationaalisuutta, vaan erilaista evidenssiä päättelyssä.

Subjektiiiviset todennäköisyydet johonkin tapahtumaan määritellään todennäköisyyksinä, jotka johtavat vedonlyöjän tekemään tiettyjä valintoja lopputuloksesta, joka riippuu tästä kyseisestä tapahtumasta (Andersen ym., 2010).

Galavotti (2001) korostaa, että todennäköisyydet ovat subjektiivisia, eikä ole olemassa yhtä ainoaa tapaa arvioida todennäköisyyksiä. Galavottin mukaan todennäköisyyden arvioiminen on monimutkainen proseduuri, joka koostuu useista samanaikaisista muuttujista.

Koska yhtä oikeaa keinoa subjektiivisten todennäköisyyksien arvioimiseen ei ole, on hyvin todennäköistä, että vedonlyöjän ja vedonvälittäjän arviot

eroavat toisistaan. Kuten Malaska ja Virtanen esittivät, eriävä subjektiivinen tietämys ei merkitse epärationaalisuutta, vaan erilaista evidenssiä päättelyssä. Vedonlyöjä ja vedonvälittäjä saattavat esimerkiksi painottaa samaa informaatiota eri tavalla todennäköisyysarvioissaan.

Subjektiiviseen todennäköisyyteen perustuvassa urheiluviedonlyönnissä taitavalla vedonlyöjällä on mahdollisuus arvioida todennäköisyyksiä vedonvälittäjää paremmin ja täten sijoittaa rahojaan voitollisesti pitkällä aikavälillä.

Jos ajatellaan, että vedonvälittäjä tarjoaa tiettyyn urheilutapahtumaan esimerkiksi kotijoukkueen voiton kertoimeksi 2,00. Tämä tarkoittaa sitä, että vedonlyöjän tulee panostaa kyseisen joukkueen voittoa siinä tapauksessa, mikäli hän on arvioinut sen todennäköisyydeksi yli 50%. Kun vedonlyöjä arvioi kotijoukkueen voiton todennäköisyydeksi yli 50%, on kerroin 2,00 tällöin suurempi kuin sitä vastaavan todennäköisyyden käänteisluku vedonlyöjän määrittämällä todennäköisyysarviolla. Kuten ensimmäisessä luvussa määriteltiin, tässä tapauksessa kyseessä on ylikerroin, joihin panostamalla vedonlyöjän on mahdollista jäädä voitolle pitkällä aikavälillä. Ensimmäisessä luvussa määriteltiin kertoimen olevan todennäköisyysarvion käänteisluku. Jos oletetaan, että vedonlyöjä on arvioinut kotijoukkueen voiton todennäköisyydeksi 55%, tulisi kertoimen olla siis minimissään 1,82, jotta se kattaisi pelaamisen riskin. Muussa tapauksessa vedonlyönti on pitkällä aikavälillä tappiollista toimintaa, mikäli siis oletetaan vedonlyöjän arvion olevan täysin tarkka.

2.2.4 Panostaminen

Kuten Suomen arpajaislaissa määritellään, uhkapelillä viitataan peliin, jossa mahdollinen häviö on ilmeisessä epäsuhteessa ainakin jonkun osallistujan maksukykyyn. Yksi voitollisen vedonlyönnin tärkeimmistä elementeistä onkin panoksen määrittäminen oikealle tasolle suhteessa riskiin.

Oletus on, että rationaalinen vedonlyöjä lyö vetoa vain sellaisiin kohteisiin, jotka ovat odotusarvoltaan positiivisia. Mikäli vedonlyöjä haluaa maksimoida odotetun voiton tilanteessa, jossa vedon odotusarvo on positiivinen, hänen pitäisi panostaa vetoon kaikki käytettävissä olevat resurssinsa (Rotando, Thorp, 1992). Tämä taktiikka ei kuitenkaan ole pitkällä aikavälillä järkevää, sillä jossain vaiheessa vedonlyöjä menettää kaikki käytettävissä olevat resurssinsa. Mikäli vedonlyöjä saisi esimerkiksi kolikonheitossa klaavalle kertoimen 2,5, tulisi hänen panostaa voiton maksimoidakseen vetoon kaikki käytettävissä olevat resurssinsa. Veto olisi odotusarvoltaan todella hyvä, mutta 50% varmuudella se menee kuitenkin väärin ja tällöin vedonlyöjä häviää kaiken. Vedonlyönnissä optimaalinen panoskoko onkin todella olennainen osa voitollisuuden saavuttamista pitkällä aikavälillä.

Urheiluviedonlyönnissä panoskoon määrittämiseen käytetään yleensä niin kutsuttua Kellyn kaavaa. Kellyn kaavan idea on yksinkertainen: tarkoituksena on lyödä vetoa maksimoiden pääoman odotettu tuotto eli toisin sanoen maksimoida vedon odotusarvo (Thorp, 2010).

Kellyn kaava voidaan esittää muodossa: $P = (AB - 1) / (B - 1)$
Kaavassa:

- P = Optimaalinen panos
- A = Todennäköisyysarvio
- B = Kohteen kerroin

Jos ajatellaan esimerkkinä kolikonheittoa, jossa vedonvälittäjä tarjoaa klaavalle kerrointa 2,50. Klaavan todennäköisyys on luonnollisesti 50%. Tällöin optimaalinen panoskoko klaavalle olisi: $P = (0,5 \times 2,5 - 1) / (2,5 - 1) = 0,1667$. Optimaalinen panoskoko olisi siis 16,67% pelaajan käytössä olevista resursseista.

Pitkän aikavälin vedonlyöjien tulisi välttää ylipanostamista. Siksi tilanteissa, joissa tulevaisuuden todennäköisyydet ovat epävarmoja, pitkän aikavälin vedonlyöjien tulisi rajoittaa heidän panostuksiaan riittävästi välttääkseen ylipanostuksen riskin (Thorp, 2006.)

Kellyn kaavan mukaan panostettaessa vedonlyöjä ei altistu ylipanostukselle ja tällöin panoskoko on jatkuvasti optimaalinen. Vedonlyöjä ei tässä tapauksessa myöskään altistu tilanteeseen, jossa mahdollinen häviö on ilmeisessä epäsuhteessa hänen maksukykyynsä.

2.2.5 Urheiluedonlyönnin suhde sijoittamiseen

Verohallinnon internet-sivuilla (2017) sijoittaminen määritellään seuraavasti: "Sijoittamisella tarkoitetaan esimerkiksi osakkeiden hankkimista ja kaupankäyntiä, jonka tarkoituksena on tehdä voittoa."

Kuten kaikki asiat, joissa raha vaihtaa omistajaa, on myös urheiluedonlyönti yksi kaupankäynnin muoto. Urheiluedonlyönnissä kaupankäynnin kohteena ovat todennäköisyysarviot (Vuoksenmaa, 2016, 64.)

Sijoittamisessa hankitaan siis erinäisiä rahoitusinstrumentteja, kuten osakkeita. Sijoittajan tarkoituksena on tehdä voittoa käymällä kauppaa hankituilla osakkeilla. Vuoksenmaa näkee samaan tapaan urheiluedonlyönnin yhtenä kaupankäynnin muotona. Siinä missä osakemarkkinoilla käydään kauppaa pörssikursseilla, urheiluedonlyönnissä kaupankäynnin kohteena ovat todennäköisyysarviot.

Tiitun (2016) tutkimuksen mukaan vedonlyöntimarkkina ei ole vielä yhtä pitkälle kehittynyt kuin monet muut finanssimarkkinat ja erityisesti tästä syystä vedonlyöntimarkkina saattaa tarjota houkuttelevia sijoitusmahdollisuuksia osaaville vedonlyöjille.

Pickens (2006) on tutkinut urheiluedonlyönnin ja perinteisen sijoitustoiminnan välisiä eroja, tavoitteenaan selvittää miksi urheiluedonlyönti on niin paheksuttua toimintaa verrattuna sijoittamiseen. Pickens ei ymmärrä minkä takia urheiluedonlyönti on joissain valtioissa jopa kokonaan kiellettyä, kun taas sijoittaminen ainoastaan säänneltyä toimintaa. Hän tuli tutkimuksessaan siihen tulokseen, että rajanveto sijoittamisen ja urheiluedonlyönnin välillä on vaikeaa.

Pickensin mukaan urheiluedonlyönti ja sijoittaminen ovat niin samankaltaisia aktiviteetteja, että rajanveto näiden välillä on käytännössä mahdotonta. Hän perustelee väitettään sillä, että vedonlyönnin sekä sijoittamisen haittapuolellet ovat yhteneväiset. Vedonlyönnin kieltämistä on perusteltu muun muassa sillä, että se kasvattaa esimerkiksi riippuvaisuutta, rikollisuutta sekä vararikkojen lukumäärää. Pickensin mukaan sijoittamisen negatiiviset vaikutukset voivat olla täysin samat.

Molemmissa aktiviteeteissa on mahdollisuus hävitä suuria rahasummia, pahimmassa tapauksessa jopa kaikki. Mikäli kuitenkin sijoittaja toimii rationaalisesti, joko urheiluedonlyönnissä tai perinteisessä sijoitustoiminnassa, on kaikkien toimintaan varattujen resurssien menettäminen äärimmäisen epätoiminnallista.

Moderni portfolioteoria opettaa, että hajauttamalla sijoituksiaan, sijoittaja voi aina pienentää sijoitustensa varianssia (Malkiel, 1999). Markkinoiden tehokkuuteen uskovon Malkielin mukaan sijoittaminen on kuin uhkapeliä, jossa menestyminen riippuu kyvystä ennustaa tulevaisuutta. Markowitz (1991) puolestaan kirjoittaa, että sijoittaja, joka tietää tulevaisuudessa tapahtuvan asian varmuudella, voi sijoittaa kaiken vain yhteen arvopaperiin.

Todellisuudessa yhteen arvopaperiin sijoittaminen on uhkapelaamista, sillä kukaan ei voi tietää tulevaisuuden tapahtumia täysin varmasti, vaan sijoittaminen sisältää aina riskejä. Suurin uhkapelaamisen ja sijoittamisen välinen ero liittyy juuri panosten kokoon suhteessa omiin varoihin. Urheiluedonlyönnissä riskiä hajautetaan panostamalla vetoon vain optimaalinen summa suhteessa riskiin ja käytettävissä oleviin varoihin nähden. Lisäksi urheiluedonlyönnissä ylisuurilla panoksilla pelaaminen suhteessa omiin käytettävissä oleviin resursseihin nähden täyttää uhkapelaamisen kriteerit.

2.2.6 Yhteenveto

Kirjallisuuskatsauksen toisessa alaluvussa avattiin muun muassa uhkapelaamisen käsite ja pyritty tekemään eroa uhkapelaamisen ja urheiluedonlyönnin välille. Luvussa tarkoituksena oli vastata tutkielman toiseen apitutkimuskysymykseeni: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?"

Sekä sijoittaminen että urheiluedonlyönti perustuvat aina myös osittain sattumaan. Urheiluedonlyönnissä ei ole koskaan ennalta selvää, miten tapahtumassa tulee käymään. Hyvin informoidun, rationaalisesti toimivan urheiluedonlyönnin toiminta eroaa kuitenkin oleellisesti uhkapelaamisesta, jossa mahdollinen häviö on ilmeisessä epäsuhteessa pelaajan maksukykyyn. Väärin harjoitettuna urheiluedonlyöntikin on lähempänä uhkapelaamista kuin sijoitustoimintaa.

Riittävän pitkällä aikavälillä sattuman merkitys kuitenkin vähenee, mikäli urheiluedonlyönnin kykenee laskemaan todennäköisyyksiä vedonlyöntimarkkinana paremmin, panostaa ainoastaan kertoimiin, joissa odotusarvo on yli yhden ja välttää ylipanostamista pitäen panoskoon aina optimaalisena riskiin nähden. Kuten tämän luvun ensimmäisessä alaluvussa viitattu Buchdahl kirjoitti, on

sijoittaminen pitkäkestoista toimintaa. Täten myös urheiluvendonlyöntiin sijoittaminen täytyy nähdä pitkäkestoisena toimintana, ei yksittäisinä uhkapeleinä.

Lyhyellä aikavälillä sattuman merkitys on puolestaan suurempi. Tämän vuoksi optimaalinen panoskoko onkin todella olennainen osa urheiluvendonlyönnin voitollisuuden suhteen. Esimerkiksi kolikonheitossa klaavaan tai kruunaan kannattaa panostaa vain siinä tapauksessa, mikäli pelaaja saa tapahtuman toteutumiselle kertoimen, joka on yli 2,00. Vaikka pelaaja saisi jommankumman tapahtuman toteutumiselle kuinka suuren kertoimen hyvänsä, on klaavaan tai kruunan esiintymisen todennäköisyys kuitenkin silti ainoastaan 50%. Mikäli vedonlyöjä panostaa kruunaan, klaava saattaa siitä huolimatta toteutua vaikkapa seitsemän kertaa peräjälkeen. Jos vedonlyöjä kuitenkin jatkaa kruunaan panostamista optimaalisella panoskoolla kertoimen ollessa yli 2,00, voi hän riittävän pitkällä aikavälillä olla suhteellisen varma saamastaan tuotosta.

Sijoittamisessa sen sijaan yhteen arvopaperiin kaiken laittaminen on samaan tapaan uhkapelaamista. Sijoittamisen riskiä voidaan kuitenkin vähentää hajauttamisella. Oikein hajautettu osakesalkku tuottaa riittävän pitkällä aikavälillä hyvin todennäköisesti, aivan kuin oikein harjoitettu urheiluvendonlyöntikin. Lyhyellä aikavälillä sen sijaan myös osakesijoittamisen riskit ovat suurempia, sillä osakekurssit saattavat laskea.

Rajanveto uhkapelaamisen ja sijoittamisen välillä on haastavaa, sillä näiden kahden aktiviteetin välillä on runsaasti yhteneväisyyksiä. Olennaisin ero uhkapelaamisen ja sijoittamisen välillä liittyy panosten kokoon omiin varoihin nähden. Tässä luvussa esiteltyjen määritelmien kautta, oikein harjoitettu urheiluvendonlyönti on kuitenkin lähempänä perinteistä sijoitustoimintaa kuin uhkapelaamista. Täten urheiluvendonlyöntiä voidaan pitää yhtenä kaupankäynnin muotona, jota on mahdollista harjoittaa sijoitustoimintana, jossa on tarkoitus tehdä voittoa riittävän pitkällä aikavälillä.

2.3 Tiedonlouhintateknologia urheiluvendonlyönnissä

Tutkielman kirjallisuuskatsauksen kolmannessa alaluvussa tutkitaan ensin vedonlyöntimarkkinoiden tehokkuutta. Ensimmäisessä alaluvussa vastataan tutkielman toiseen apututkimuskysymykseen: ”Onko vedonlyöntimarkkina tehokas?”

Tämän jälkeen selvitetään millä tavoin tiedonlouhintateknologiaa voidaan käyttää hyödyksi urheiluvendonlyönnissä. Tiedonlouhintaprosessi on pääpiirteittäin samankaltainen jokaisessa urheilutapahtumien lopputuloksia ennustavia simulaatiomalleja käsittelevässä tutkimuksessa, joten havainnollistamisen apuna tässä tutkielmassa käytetään Caon (2012) kyseisestä aiheesta tekemää tutkimusta.

Neljännessä alaluvussa vastataan kirjallisuuskatsauksen puitteissa tutkielman päätutkimuskysymykseen: ”Voidaanko tiedonlouhintateknologiaa hyödyntämällä saavuttaa tuottoja urheiluvendonlyönnissä?” Vastausta päätutkimuskysymykseen haetaan saaduista tutkimustuloksista, joissa on testattu ur-

heilutapahtumien lopputuloksia ennustavia simulaatiomalleja. Vastaus ja pohdinnat kirjallisuuskatsauksen päätutkimuskysymykseen aikaisempiin aiheesta tehtyihin tutkimuksiin viitaten esitetään seuraavassa luvussa.

2.3.1 Vedonlyöntimarkkinan tehokkuus

Vedonlyöntimarkkinoilla ja osakemarkkinoilla on keskenään paljon yhtäläisyyksiä. Linda ja Bill Woodlandin (1994) mukaan osakemarkkinoita ja vedonlyöntimarkkinoita yhdistää ennen kaikkea markkinoilla toimivien henkilöiden runsas lukumäärä, tulevien tuottojen epävarmuus sekä informaatio, jota menneistä suorituksista on saatavilla helposti ja nopeasti. Avery ja Chevalier (1999) havaitsivat, että sekä osake- että vedonlyöntimarkkinoilla toimii lukuisia yksittäisiä sijoittajia ja myös ammattimaisia sijoittajia, joiden tavoitteena on löytää markkinoilta mahdollisia hinnoitteluvirheitä. Osakemarkkinoiden tehokkuudesta on käyty keskustelua pitkään, niin puolesta kuin vastaan.

Faman (1970) kehittämä tehokkaiden markkinoiden hypoteesi olettaa, että tehokkailla markkinoilla kaikki saatavilla oleva informaatio heijastuu välittömästi arvopapereiden hintoihin, jolloin arvopapereiden hinnat vastaisivat niiden todellisia arvoja. Mikäli markkinat olisivat täysin tehokkaat, kenenkään ei olisi mahdollista saavuttaa keskimääräistä suurempia tuottoja, sillä mitään sellaista informaatiota ei olisi enää saatavilla, jota ei olisi hinnoiteltu osakkeiden pörssikursseihin.

Markkinoiden tehokkuuden puolesta puhuvan Faman (1970) mukaan markkinoiden tehokkuudella on olemassa kolme eritasoista kategoriaa: Heikot ehdot, keskivahvat ehdot sekä vahvat ehdot. Faman mukaan markkinoiden tehokkuuden heikossa muodossa hinnat muodostuvat lähinnä historiallisesta informaatiosta. Keskivahvassa muodossa Faman tutkimuksen mukaan hinnat puolestaan sisältävät jo kaiken julkisesti saatavilla olevan informaation, kun taas vahvassa muodossa arvopapereiden hinnat sisältävät lisäksi myös sisäpiiritiedon, eikä ylituottojen saavuttaminen pitkällä aikavälillä täten olisi mahdollista.

Myös vedonlyöntimarkkinoiden tehokkuudesta on tehty runsaasti tutkimusta. Lahdensuo (2015) kirjoittaa, että urheiluedonlyöntimarkkina on toiminnaltaan ja perusoletuksiltaan hyvin samankaltainen muiden spekulatiivisten markkinoiden kanssa ja ympäristönä soveltuu erinomaisesti tehokkaiden markkinoiden hypoteesin testaamiseen. Hänen mukaansa tämä johtuu siitä, että urheiluedonlyöntimarkkinoilla sijoitushyödykkeen todellinen arvo paljastuu vedonlyöntitapahtuman päättyessä välittömästi, tarkasti ja ilman tulkinnanvara.

Useat tutkimukset ovat osoittaneet, että vedonlyöntimarkkinoilta löytyy epäjohdonmukaisuutta, joka avaa tuottomahdollisuuksia. Epäjohdonmukaisuutta ei olisi havaittavissa, mikäli markkinat olisivat täysin tehokkaat. Tutkimuksissa on esitelty erinäisiä vedonlyöntistrategioita, joilla on tutkimuksessa saavutettu vedonlyöntimarkkinan tehokkuutta vastaan oleva tulos.

Pankoff (1968) oli ensimmäinen, joka tutki markkinatehokkuutta urheiluvedonlyönnissä. Hän tuli tutkimuksessaan siihen tulokseen, että amerikkalaisen jalkapallon vedonlyöntimarkkina täyttää markkinatehokkuuden heikot ehdot. Pankoff löysi markkinasta epäjohdonmukaisuutta, mutta ero hänen ennusteidensa ja amerikkalaisen jalkapallon NFL:n toteutuneiden lopputulosten välillä ei kuitenkaan ollut riittävän suuri voitollisen vedonlyönnin mahdollistamiseen, sillä vedonvälittäjä otti komissiona osansa jokaisesta vedosta.

Snyder (1978) tutki vedonlyöntimarkkinan tehokkuutta hevosurheilussa. Hän vertasi vedonlyöntikertoimista johdettuja todennäköisyysarvioita hevoskilpailuiden toteutuneisiin lopputuloksiin. Snyder löysi tutkimuksessaan hevosvedonlyöntimarkkinasta selkeitä epäjohdonmukaisuuksia, mutta hänen mukaansa subjektiivisten todennäköisyysarvioiden ja lähtöjen toteutuneiden tulosten välinen ero ei kuitenkaan ollut riittävän suuri, jotta pelaaminen olisi pitkällä aikavälillä riittävän voitollista ylittämään hevosratojen vedoista perimät komissiot.

Myös Asch, Malkiel ja Quandt (1984) tutkivat markkinatehokkuutta hevosvedonlyönnissä. He löysivät markkinoilta väärinhinnoittelua ja havaitsivat, että hevosvedonlyönnissä voisi saavuttaa ylisuurta tuottoa panostamalla lähdön ennakkosuosikkiin eli hevoseen, jonka voiton subjektiivinen todennäköisyys on lähdön suurin. Asch ym. (1984) tulivat siihen tulokseen, että varmistukseen strategian toimivuuden, he tarvitsevat aiheesta vielä lisätutkimusta suuremmalla otoskoolla.

Kaksi vuotta myöhemmin, vuonna 1986 Asch, Malkiel ja Quandt tekivät uuden samanlaisen tutkimuksen uudemmalla aineistolla hevosvedonlyöntimarkkinan tehokkuudesta. Uudemman tutkimuksen jälkeen he kuitenkin suhtautuivat skeptisemmin hevosvedonlyöntimarkkinoiden mahdolliseen tehottomuuteen. Vuonna 1986 tehdyssä tutkimuksessa Asch ym. pääsivät tulokseen, jossa vetojen lyöminen heidän strategiallaan olisi niin kutsuttua nollasummapeiliä eli pitkällä aikavälillä vedoista ei seuraisi voittoja, eikä tappioita. He kuitenkin huomauttavat, että kun ratojen perimät komissiot otetaan huomioon, ei panostaminen heidän strategiallaan ole enää kannattavaa. Asch, Malkiel ja Quandt kuitenkin antavat edelleen mahdollisuuden hevosvedonlyöntimarkkinoiden tehottomuudelle, vaikkakin suhtautuvat siihen aiempaa skeptisemmin (Asch, Malkiel, Quandt, 1986.).

Golec ja Tamarkin (1991) puolestaan löysivät tehottomuutta tutkiessaan amerikkalaisen jalkapallon vedonlyöntimarkkinoita. He havaitsivat, että vedonlyöjät aliarvoivat kotiedun merkitystä NFL:ssä. Lisäksi heidän tutkimuksensa osoitti, että vedon lyöminen kunkin ottelun altavastajien puolesta tuotti keskimääräistä paremmin. Myös Golec ja Tamarkin totesivat kuitenkin lopuksi, että vedonvälittäjien vedoista ottama komissio saattaa tasoittaa markkinoiden lievän tehottomuuden, eikä täten välttämättä mahdollistaisi voitollista vedonlyöntiä.

Gray ja Gray (1997) jatkoivat Golecin ja Tamarkinin aikaisempaa tutkimusta NFL-markkinan tehottomuudesta. Heidän havaitsivat tutkimuksessaan, että vedon lyöminen kotijoukkueena toimivan altavastajan puolesta on palaut-

tanut heidän aineistonsa perusteella panokset takaisin yli neljän prosentin tuotolla, joka on heidän mukaansa tilastollisesti merkittävä havainto ja on riittävä kertomaan markkinoiden tehottomuudesta.

Andersonin tutkimuksen (2019) mukaan NFL:n vedonlyöntimarkkina on pitkässä juoksussa kokonaisuudessaan tehokas, mutta tietyt strategiat voivat antaa mahdollisuuden tavoitella ylituottoa lyhyemmällä aikavälillä, esimerkiksi yhden kauden aikana. Anderson (2019) kirjoittaa, että vedonlyöntimarkkina muuttuu koko ajan tehokkaammaksi, koska vedonlyöjät ovat paremmin informoituja, sillä informaatiota on saatavilla koko ajan enemmän ja enemmän.

Katić, Malarić ja Sabolić (2008) loivat yksinkertaiseen strategiaan perustuvan tietokoneohjelman, jolla he testasivat vedonlyöntimarkkinoiden tehokkuutta. Heidän saavuttamansa tulokset indikoivat markkinoiden tehottomuutta. He havaitsivat markkinoilla olevan väärinhinnoittelua ja saavuttivat tietokoneohjelmansa avulla positiivisen tuoton. Heidän mukaansa ohjelmaa kehittämällä, voisi vedonlyöntimarkkinoilta saavuttaa vielä korkeampaa tuottoa.

Monet tutkimukset ovat osoittaneet jo 1960-luvulta saakka, että vedonlyöntimarkkinat eivät ole täysin tehokkaat. Mikäli markkinoilla on tehottomuutta, avaisi se tuottomahdollisuuksia urheiluedonlyönnissä ja esimerkiksi Katić ym. saavuttivat luomansa tietokoneohjelman avulla positiivista tuottoa vedonlyöntimarkkinoilta. Lisäksi myös muut tutkimukset ovat osoittaneet vedonlyöntimarkkinoilla olevan ainakin lievää tehottomuutta, joka on ilmennyt tiettyinä epäjohtonmukaisuuksina. Seuraavaksi pyritään selvittämään, onko tiedonloughintaan perustuvien tietokonesimulaatioennusteiden avulla mahdollisuus saavuttaa tuottoa vedonlyöntimarkkinoilta.

2.3.2 Tiedonloughinta yleisesti

Maaailmaan tulvii nykyaikana päivittäin valtava määrä uutta informaatiota. Nopea tekninen kehitys tietotekniikassa on mahdollistanut suurten tietomäärien keräämisen sekä varastoimisen erilaisiin massamuisteihin (Alaharjula, 2005).

Informaation määrän räjähdysmäinen kasvu tarjoaa tutkijoille suunnattomasti uusia resursseja, joita voidaan analysoida ja tehdä niistä täten tieteellisiä löydöksiä tai havaita taloudellisesti arvokkaita menetelmiä (Hand, Mannila, Smyth, 2001, 2). Suurista datamääristä on löydettävä tutkimukselle olennainen tieto, jolloin puhutaan datan- tai tiedonloughinnasta (Turunen, 2012). Tiedonloughinta on eräänlainen prosessi, jossa etsitään tietokannoista toistuvia malleja tai kaavoja, jotka ovat hyödyllisiä päätöksenteossa (Bose, Mahapatra, 2001).

Hand, Mannila sekä Smyth (2001, 6) määrittelevät tiedonloughinnan olevan usein suurten datamäärien analysointia, jossa tavoitteena on havaita odottamattomia suhteita ja tehdä yhteenvetoa datasta uudenlaisin tavoin, jotka ovat datan haltijalle ymmärrettäviä sekä hyödyllisiä.

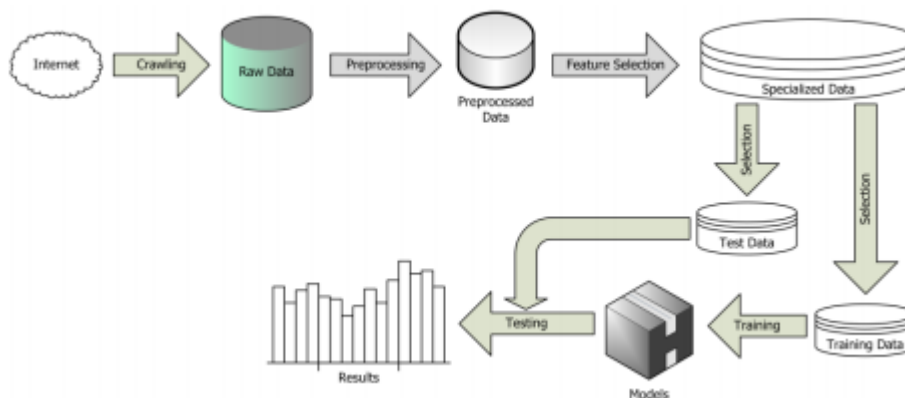
Monimutkaisista ilmiöistä yhä helpommin kerätty runsas data on johtanut datavetoiseen tutkimusasetelmaan, jossa ilmiön monimutkainen malli pyritään oppimaan automaattisilla menetelmillä suoraan datasta (Turunen, 2012).

Seuraavissa alaluvuissa selvitän, kuinka tiedonlouhintaa voidaan käyttää hyväksi urheiluedonlyönnin saralla ja onko sitä hyödyntämällä mahdollisuus kyetä voitolliseen tulokseen urheiluedonlyöntimarkkinoilla.

2.3.3 Tiedonlouhinta urheiluedonlyönnissä

Tänä päivänä internetistä löytyy valtavia datamääriä sisältäviä urheilutietokantoja. Urheilu tuottaa huomattavan määrän tilastollista informaatiota muun muassa kustakin pelaajasta, joukkueesta, peleistä sekä sarjakaudesta (Haghighat, Rastegari & Nourafza, 2013). Vuoksenmaan (2016, 288) mukaan tietotekniikka on muuttanut urheiluedonlyönnin luonnetta. Hänen mukaansa vedonlyöjät eivät kilpaile enää niinkään informaation hankinnalla, vaan kyvyllä käsitellä informaatiota. Vuoksenmaa korostaa, että menestys vedonlyönnissä on enimmäkseen tutkimusta ja osaksi urheilutietoutta.

Vedonlyönnissä tuottavien strategioiden etsimisessä käytetään yleensä historiallista dataa, jota on saatavilla laajojen tietokantojen ansiosta varsin runsaasti lähes kaikesta mitä urheilutapahtumista voidaan tilastoida. Leungin ja Josephin (2014) mukaan tiedonlouhinnan tavoitteena urheilutapahtumien ennustusten laatimisessa on löytää urheiludatasta aikaisemmin havaitsematta jääneitä, potentiaalisesti hyödyllistä informaatiota. Heidän tutkimuksessaan todetaan, että tiedonlouhinta urheilun tietokannoista auttaa luomaan ennusteita tulevaisuuden otteluista käyttämällä aikaisemmista otteluista kerättyä dataa.



Kuvio 1 Tiedonlouhintaprosessi

Jokaisessa tiedonlouhintatutkimuksessa alkuperäisen datan täytyy kulkea muutaman eri prosessointivaiheen läpi, jotta siitä voitaisiin saada irti hyödyllistä informaatiota (Zdravevski & Kulakov, 2010). Yllä olevassa kuviossa (Kuvio 1) Zdravevski ja Kulakov (2010) ovat avanneet tiedonlouhintaprosessia urheiluedonlyönnissä. Prosessi käydään läpi hyödyntäen Caon (2012) tekemää tutkimusta tiedonlouhinnasta urheiluedonlyönnissä.

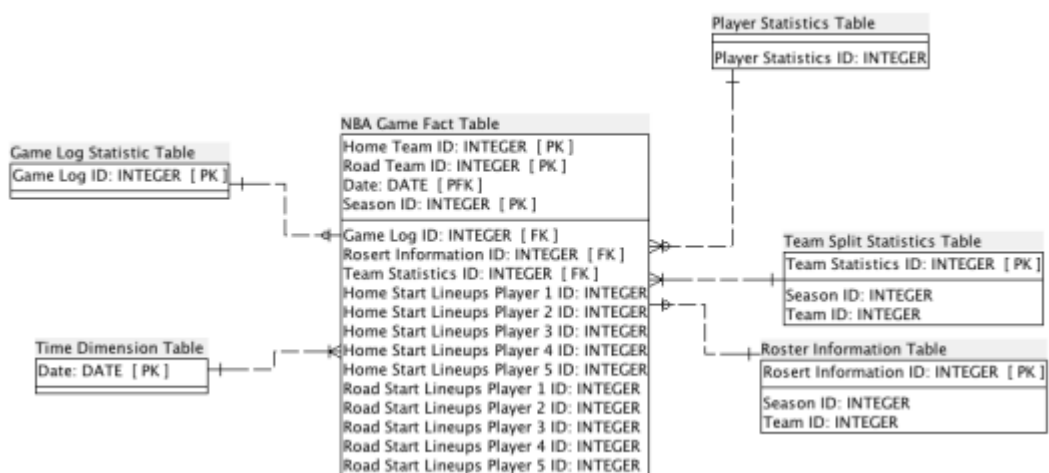
Tietokannoista louhitulla datalla kyetään rakentamaan erilaisia simulaatiomalleja, joilla voidaan ennustaa urheilutapahtumien lopputuloksia. Cao (2012) rakensi simulaatiomallin ennustamaan koripallo-otteluiden lopputulok-

sia (Kuvio 2). Mallin tarkoituksena oli hakea edellisten pelien tilastot kahden toisiaan vastaan pelaavan koripallojoukkueen väliltä ja käyttää näitä tilastoja ennustamaan tulevan ottelun lopputulosta. Ideana on yhdistellä joukkueen tilastoja, vastustajan tilastoja, otteluohjelmaa, pelaajien henkilökohtaisia tilastoja sekä joukkueiden esityksiä edellisissä otteluissa ja luoda näiden pohjalta arvio tulevan ottelun lopputuloksesta.



Kuvio 2 Tilastodatan kerääminen ja varastointi

Cao keräsi dataa NBA-aiheisilta (National Basketball Association) internet-sivuilta, jotka julkaisivat tilastodataa kyseisestä koripalloliigasta. Hän keräsi tilastollisen informaationsa internetistä automaattisin menetelmin hyödyntäen prosessissa Ruby-kieltä. Kun data on kerätty paikalliselle tietokoneelle pääasiallisesti CSV-tiedostomuodossa, Cao latsi kerätyn datan MySQL-tietokantaan Ruby- sekä MySQL -ohjelmointikieliä hyödyntäen. Cao käytti MySQL-tietokantaa datan hallinnassa johtuen sen ominaisuuksista järjestää dataa tehokkaasti. Caon NBA-joukkueista ja pelaajista keräämän tilastollisen datan hallinta onnistui MySQL-tietokannan avulla hyvin, mutta ongelmia hänelle aiheutti tietokannan erillisiin tauluihin varastoima data, joka puolestaan teki tiedonlouhinnasta erittäin raskaan toteuttaa. Ongelman ratkaistakseen Cao loi datalle paikallisvaraston, joka yhdisti erilliset taulut. Täten paikallisvarastoa kyettiin käyttämään paremmin tiedonlouhintatarkoitukseen.



Kuvio 3 Tietokannan tähtirakenne

Oheisessa kuviossa (Kuvio 3, Cao, 2012) on nähtävillä Caon luoma tähtirakenteinen paikallisvarasto. Tietokantarakenteen keskiössä on "game fact" -taulu, joka sisältää kaiken oleellisen informaation jokaisesta ottelusta. Keskiössä oleva "game fact" -taulu on yhteydessä muihin tauluihin. Caon mukaan "time dimension" -taulu on hyvin oleellinen vedonlyöntitutkimuksen kannalta, sillä se mahdollistaa analyysien tekemisen tiettyyn ajanjaksoon rajaten. Paikallisvaraston luomisen jälkeen tiedonlouhintaa pystytään hyödyntämään relevantin datan ulossaamiseksi varastosta.

Tämän jälkeen tietokannan haltija voi luoda lopputuloksia ennustavan algoritmin haluamistaan muuttujista. Zdravevski ja Kulakov (2010) valitsivat omaan algoritmiinsa kymmenen eri muuttujaa, joilla oli suurin vaikutus koripallo-otteluiden lopputuloksiin. Trawinski (2010) tutki yhteensä kahdeksaa erilaista algoritmia. Lopulta hän päätyi rakentamaan lopullisen algoritminsa viidestä sellaisesta muuttujasta, jotka toistuivat näissä kahdeksassa tutkimuksessaan algoritmissa.

Buursma (2010) tutki, kuinka jalkapallo-otteluiden lopputuloksia voitaisiin ennustaa käyttämällä dataa aikaisemmista otteluista ja mitkä muuttujat olisivat tärkeitä ennusteen laatimisessa. Hänen alustava algoritminsa koostui kymmenestä eri muuttujasta, joita hän alkoi tutkimaan tarkemmin. Buursma poisti algoritmista yhden muuttujan kerrallaan ja tutki kuinka se vaikuttaisi ennusteen tarkkuuteen. Hän poisti muuttujan, mikäli sen eliminoiminen paransi algoritmin tarkkuutta. Samaan tapaan tämän jälkeen Buursma alkoi lisäämään algoritmiin yhden muuttujan kerrallaan testaten parantaako kyseisen muuttujan lisäys algoritmin laatiman ennusteen tarkkuutta. Lopulta Buursma päätyi tutkimuksessaan yhdeksän eri muuttujaa sisältävään, ennusteita jalkapallo-otteluiden lopputuloksista laativaan algoritmiin.

Cao (2012) puolestaan tutki peräti 60:n eri muuttujan vaikutusta koripallo-otteluiden lopputulokseen. Hän päätyi käyttämään algoritmissaan lopulta 46:ta eri muuttujaa.

Taulukko 1 Algoritmin tuottama datasetti

Attribute	Sacramento Kings	Houston Rockets
Won	25	42
Lost	55	39
Pct	0.313	0.519
Homewon	18	23
Homelost	22	18
Roadwon	7	18
Roadlost	33	22
Divwon	5	9
Divlost	10	7
Confwon	16	27
Conflost	34	24
Streak	-1	-1
L10won	1	5
L10lost	9	5
FGM	38.3	37.7
FGA	84.1	84.4
FGP	45.6	44.7
3M	5.9	7.9
3A	16.9	22.4
3P	34.9	35.2
FTM	17.4	19.0
FTA	24.0	24.7
FTP	72.6	77.2
OR	11.9	11.8
DR	30.7	30.1
TR	42.6	42.0
AS	20.5	21.8
TO	14.2	13.8
ST	6.9	7.1
BL	4.5	3.9
F	22.3	20.9
P	100.0	102.4
Winning probabilities	42.8	57.2
Actual result	107	117

Yllä olevassa taulukossa on esimerkki Miljkovićin ym. (2010) tutkimuksen algoritmin tuottamasta datasetistä taulukon muodossa. Taulukkoon on laskettu historiallisen datan perusteella saadut keskiarvot kullekin eri tilastolle, joiden perusteella algoritmi on laskenut arvion ottelun lopputuloksesta sekä joukkueiden voiton todennäköisyyksistä koripallon NBA-liigan otteluun Sacramento Kings-Houston Rockets (Taulukko 1, Milković, Gajić, Kovačević & Konjović, 2010). Taulukosta 2 on puolestaan nähtävissä selitykset taulukossa 1 käytetyille lyhenteille:

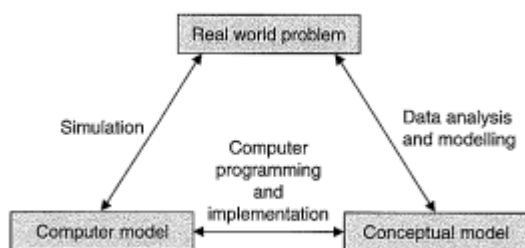
Taulukko 2 Selitykset taulukossa 1 käytetyille lyhenteille

Attribute	Description
FGM	Field goal made per game
FGA	Field goal attempted per game
FGP	Field goal percentage per game
3M	3-pointers made per game
3A	3-pointers attempted per game
3P	3-pointers percentage per game
FTM	Free throws made per game
FTA	Free throws attempted per game
FTP	Free throws percentage per game
OR	Offensive rebounds per game
DR	Defensive rebounds per game
TR	Total rebounds per game
AS	Assists per game
TO	Turnovers per game
ST	Steals per game
BL	Blocks per game
F	Fouls per game
P	Points per game

2.3.4 Monte Carlo -simulaatio

Monte Carlo -simulaatio on simulaatio, joka perustuu toistuvaan satunnaiseen näytteenottoon ja tilastolliseen analyysiin tulosten laskemiseksi (Raychaudhuri, 2008). Harrison (2010) havainnollistaa Monte Carlo -simulaatiota yksinkertaisen kolikonheittoesimerkin avulla. Harrisonin mukaan Monte Carlo -simulaatioissa on pohjimmiltaan kyse kuin sarjasta kolikonheittoja, simulaatioissa kolikkoa heitetään miljoonia, miljardeja tai vaikka triljoonia kertoja. Täten simulaatiot voivat antaa tutkijoille uusia tietoja tutkittavasta kohteesta. Harrison (2010) toteaa, että idea satunnaisuuden käyttämisestä tulevaisuuden määrittäjänä oli mullistava.

Monte Carlo -simulaatioissa ideana on määrittellä simulaatioissa käytettävät parametrit etukäteen. Tämän jälkeen Monte Carlo -simulaatio simuloi mallia toistuvasti, simuloiden joka kerta erilaisen satunnaisarvojoukon annetuista malliparametreista. Lopputuloksena simulaatio antaa lukumäärän tietyn tapahtuman esiintymiskerroista simulaatioissa (Bonate, 2001).



Kuvio 4 Monte Carlo -simulaatio

Yllä olevassa kuviossa (Kuvio 4, Bonate, 2001) on havainnollistettu Monte Carlo -simulaatiota. Ensin määritellään ongelma, johon simulaatiolla haetaan ratkaisua. Sen jälkeen dataa analysoimalla ja mallintamalla luodaan käsitteellinen malli. Bonaten (2001) mukaan tässä vaiheessa suoritettava simulaation parametrien määrittäminen on todennäköisesti koko simulaatioprosessin haasta-

vin vaihe, johtuen tietämyksen puutteesta tutkittavaa ilmiötä kohtaan. Kun malli on saatu määritettyä, täytyy siitä vielä luoda tietokonealgoritmi, jonka avulla suoritetaan simulaatiot. Ennen simulaatiota tulee vielä määrittää, kuinka monta kertaa simulaatio toistetaan. Bonate (2001) kertoo, että mikäli tarkoituksena on tutkia esimerkiksi tietyn lopputuleman esiintymistä, täytyy suorittaa useita, yleensä jopa tuhansia simulaatioita, jotta myös epätodennäköisesti ilmenivät tapahtumat saavat simulaatioissa sopivan painoarvon. Mitä enemmän simulaatioita tehdään, sitä tarkempi on myös Monte Carlo -simulaation antama lopputulos. Olennaisen tärkeää on kuitenkin ennalta määritettyjen parametrien tarkkuus.

Monte Carlo -simulaatiota voidaan hyödyntää myös urheiluedonlyönnin saralla. Simulointiprosessi toimii vastaavalla, yllä kuvatulla tavalla. Simulaatioon syötetään ennalta määritellyt parametrit, joista simuloimalla voidaan johtaa todennäköisyysarviot erinäisille urheilutapahtumille.

Hvattum (2013) tutki urheiluedonlyöntimarkkinan tehokkuutta hyödyntäen Monte Carlo -simulaatiota. Tutkimuksessaan hän päätyi siihen tulokseen, että Monte Carlo -simulaatio tarjosi paremman kokonaistuoton kahden ottelukauden ajanjaksolla kuin mitä vedonlyöntimarkkinan määrittämät kertoimet olisivat implikoineet. Hän onnistui Monte Carlo -simulaatiota hyödyntäen tekemään siis niin sanottua ylituottoa kahden kauden ajanjaksolla. Hvattumin tutkimuksen mukaan Monte Carlo -simulaatiota hyödyntämällä kykenee tekemään vedonlyöntimarkkinaa tarkempia arvioita urheilutapahtumien todennäköisyyksistä.

2.3.5 Tiedonlouhintateknologian hyödyntäminen urheiluedonlyönnissä

Tässä alaluvussa käydään läpi tuloksia muutamista tutkimuksista, jotka käsittelevät urheilutapahtumien lopputuloksien arviointia tietokonesimulaatioiden avulla. Näiden tutkimustulosten perusteella vastaan päätutkimuskysymykseeni: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?"

Kuten ensimmäisessä luvussa kirjoitin, urheiluedonlyönnissä kerroin on todennäköisyysarvion käänteisluku, joten arvioimalla vedonvälittäjää tarkemmat todennäköisyysarviot urheilutapahtumaan, pitäisi voitollisen pelaamisen olla mahdollista pitkällä aikavälillä.

Mikäli jonkin varmuudella toteutuvan tapahtuman todennäköisyys on 0,2, oikean kertoimen tulisi olla $1/0,2=5$. Vedonvälittäjä haluaa kuitenkin totta kai haluaa harjoittaa voitollista liiketoimintaa, joten mikäli he uskovat tietyn tapahtuman toteutuvan todennäköisyydellä 0,2, he eivät tarjoa kohteeseen kerrointa 5, vaan antavat kertoimeksi esimerkiksi 4,50. Tämän "epäreiluuden" johdosta tuoton tekeminen urheiluedonlyönnissä ei ole aivan niin helppoa (Buursma, 2010.). Buursma kehitti tutkimuksessaan algoritmia, joka arvioi jalkapallootteluiden todennäköisyyksiä.

Tutkimuksissa on käytetty simulaatiomallin toteutunutta tarkkuutta arviotaessa tarkkuusprosenttia, joka saadaan jakamalla historiallisesta datasta saa-

dut oikein menneet ennusteet jaettuna kaikkien suoritettujen ennusteiden lukumäärällä (Kuvio 5, Milković, Gajić, Kovačević & Konjović, 2010). Saadun tarkkuusprosentin mukaan eri simulaatiomallien menestystä pystytään vertailemaan toisiinsa.

$$\textit{accuracy} = \frac{\textit{number of correct predictions}}{\textit{total number of predictions}}$$

Kuvio 5 Kaava osumatarkkuusprosentille

Kahn (2003) tutki amerikkalaisen jalkapalloliiga NFL:n otteluiden lopputulosten ennustettavuutta. Hän laati ennusteet lopputuloksista kauden kahdelle viimeiselle ottelukierrokselle. Molemmilla kierroksilla hän arvioi tulokset kahdella eri tavalla, ensin koko kauden ajalta kerättyä dataa hyödyntämällä ja sen jälkeen edellisen kolmen viikon aikana kertynyttä dataa käyttämällä. Ensimmäisellä arvioimallaan viikolla Kahnin algoritmi saavutti osumatarkkuuden 75%, kun hän käytti dataa koko kauden ajalta. Edellisen kolmen viikon dataa käyttämällä Kahn arvioi otteluiden voittajat hieman heikommalla, 62,5% osumatarkkuudella. Seuraavan ottelukierroksen peleissä Kahn sai jälleen saman 75% osumatarkkuuden käytettyään koko kauden aikana kertynyttä dataa. Sen sijaan tällä kertaa osumatarkkuus oli ainoastaan 37,5%, kun Kahn käytti dataa kolmelta edelliseltä viikolta. Kahn tuli siihen tulokseen, että koko kauden aikana kertynyt data on huomattavasti tehokkaampi tekijä kuin otteluiden lopputulosten ennustamisessa kuin vain edellisten viikkojen ajalta kertynyt data. Kahnin mukaan NFL-joukkueet käyttäytyvät johdonmukaisesti pitkällä aikavälillä ja täten on arvioiden tarkkuuden kannalta parempi, että ennustettaessa lopputuloksia, myös tutkimuksissa käytettävää dataa kerätään pitkältä aikaväliltä.

Buursma (2010) käytti tutkimuksessaan algoritmia, joka koostui yhdeksästä tarkasti testatusta muuttujasta, joilla hän uskoi olevan suurin vaikutus jalkapallo-ottelun lopputulokseen. Tutkimuksessa Buursman algoritmi saavutti parhaimmillaan 55% osumatarkkuuden. Buursma osoitti tutkimuksessa, että vetoa tulisi lyödä vain silloin, kun oma todennäköisyysarvio on parempi kuin mitä vedonvälittäjän sitä vastaava kerroin, eikä esimerkiksi jokaiseen otteluun vedon asettaminen kotijoukkueen puolesta ole voittava strategia. Buursma havainnollisti kuinka, että mitä pidemmällä aikavälillä vetoa lyödään, sitä suuremmaksi todennäköisyys voitolle jäämiseen kasvaa.

Zdravevski ja Kulakov (2010) testasivat yhteensä 36:ta algoritmia ennustamaan koripallo-otteluiden voittajan kaikkiaan 930:een otteluun. Heidän paras tuloksensa oli algoritmi, joka onnistui ennustamaan voittajan 72,8% tarkkuudella. Mikäli jokaiseen peliin olisi veikkannut ennakkosuosikin voittoa, olisi onnistunut ennustamaan voittajan 67,8% tarkkuudella. Zdravevskin ja Kulakovin testaamista algoritmeista 21 sai ennusteistaan tarkemman tuloksen kuin 67,8%.

Miljković ym. (2010) kehittivät algoritmin, joka laski todennäköisyysarvioita koripallo-otteluihin. Algoritmi onnistui ennustamaan ottelun voittajan noin 67% tarkkuudella eli yli kaksi kertaa kolmesta. Täsmälleen oikean piste-eron

voittajan ja häviäjän välillä algoritmi onnistui laskemaan noin kymmenen prosentin tarkkuudella, joka oli Miljkovićin ym. mukaan varsin odotettu tulos, sillä täsmälleen oikean piste-eron ennustaminen koripallo-otteluun on todella vaikeaa ja tutkimuksen tarkoituksena olikin vain antaa suuntaa antava arvio lopputuloksesta.

Cao (2012) keräsi dataa koripallon NBA-liigasta vuosien 2006-2010 ajalta. Hän tutki neljää eri tiedonlouhintamenetelmää koripallo-otteluiden lopputuloksien arviointiin. Ennusteiden osumatarkkuudet näillä neljällä menetelmällä olivat 65,82%, 66,67%, 67,22% sekä 67,82%. Tämän jälkeen Cao lisäsi laskelmiinsa vielä datan kaudelta 2010-2011 ja teki uudet arviot. Jokaisen neljän menetelmän osumatarkkuus parani melko huomattavasti, kun tietokantaan syötettiin lisää dataa. Uudet osumatarkkuudet datan lisäyksen jälkeen olivat 66,25%, 68,01%, 67,70% sekä 69,67%.

McCabe ja Travathan (2008) keräsivät dataa kolmen kauden ajalta neljästä eri urheiluliigasta, jotka olivat jalkapallon Valioliiga (54,6%), rugby Australian liiga (63,2%), australialaisen jalkapallon AFL (65,1%) sekä rugby Superliiga (67,5%). Toteutuneet osumatarkkuudet on merkitty sulkeisiin. McCabe ja Travathan summasivat, että urheilussa lopputulokseen vaikuttavia tekijöitä on todella paljon ja juuri tämä tekee ennustamisesta vaikeaa.

Käsitellyt tulokset tiedonlouhintateknologiaa hyödyntävistä tutkimuksista osoittavat, että yhdelläkään algoritmilla ei saavutettu sellaisia tuloksia, että voisi varmuudella sanoa algoritmista olevan kiistatonta hyötyä urheiluedonlyönissä. Tulokset osoittavat, että mitä enemmän dataa tietokantaan kerätään louhittavaksi, sitä tarkempia ennusteita saadaan. Tutkimustuloksissa käsitellyistä algoritmeista tarkinkaan ei välttämättä kuitenkaan tulevaisuudessa samankoisella otannalla pysty yhtä hyvään osumatarkkuuteen. Tarvittaisiin siis lisätutkimusta näille algoritmeille, jotka menestyivät parhaiten tässä alaluvussa esitellyistä tutkimuksista.

Lisäksi on muistettava, ettei vedonlyönissä vielä pelkästään riitä, vaikka pystyisi laskemaan todennäköisyydet täysin tarkasti. Kuten tässä tutkielmassa on aikaisemmin havainnollistettu; vedonlyönissä voi jäädä voitolle pitkällä aikavälillä ainoastaan ylikertoimia pelaamalla ja oikeassa suhteessa panostamalla. Sellaista tutkimusta, jossa lyödään vetoa jonkin algoritmin laskemien todennäköisyyksien perusteella esimerkiksi yhden kauden ajan, ei vielä tähän mennessä ennen tätä tutkielmaa olla toteutettu.

3 Kirjallisuuskatsauksen yhteenveto ja johtopäätökset

Tutkielman kolmannessa luvussa esitellään kirjallisuuskatsauksen yhteenveto sekä johtopäätökset. Kolmannen pääluvun ensimmäisessä alaluvussa esitellään yhteenveto, jossa käydään läpi tutkimuksen sisältö sekä kirjallisuuskatsauksen kulku. Toisessa alaluvussa esitellään aikaisempien aiheesta tehtyjen tutkimusten pohjalta rakennetun kirjallisuuskatsauksen tärkeimmät havainnot sekä niistä tehdyt johtopäätökset.

3.1 Kirjallisuuskatsauksen yhteenveto

Tässä tutkielmassa tutkitaan tiedonlouhintateknologian hyödyntämistä urheiluedonlyönnissä. Tutkielman ensimmäinen osuus toteutettiin systemaattisena kirjallisuuskatsauksena. Tämän tutkimuksen kirjallisuuskatsausosion tarkoituksena oli määritellä urheiluedonlyöntiä käsitteenä, kuvata sen teoriaa sekä avata tutkielmassa käytettäviä vedonlyöntialaan olennaisesti liittyviä käsitteitä.

Kirjallisuuskatsauksen ensimmäisessä luvussa käytiin läpi urheiluedonlyönnin perusteita. Lisäksi ensimmäisessä luvussa on kuvattu urheiluedonlyönnin teoriaa sekä avattu tutkielmassa käytettäviä, vedonlyöntialaan olennaisesti liittyviä käsitteitä, kuten vedonlyöntimarkkina, todennäköisyysarvio, kerroin, palautusprosentti sekä odotusarvo. Käsitteiden määrittelyn sekä konkreettisten esimerkkien avulla on pyritty avaamaan urheiluedonlyönnin lainalaisuuksia helpommin ymmärrettäviksi.

Kirjallisuuskatsauksen toisessa luvussa perehdyttiin urheiluedonlyöntiin sijoitusmuotona. Luvussa vastattiin aikaisemmin aiheesta tehtyjen tutkimusten pohjalta, kirjallisuuskatsauksen puitteissa toiseen apututkimuskysymykseen, joka oli: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?" Apututkimuskysymystä lähestytään määrittelemällä lähdekirjallisuuden avulla uhkapeli, minkä jälkeen määritelmää verrataan rationaalisesti toteutetun urheiluedonlyönnin lainalaisuuksiin. Tämän jälkeen pohdittiin sijoitustoiminnan ominaispiirteiden soveltuvuutta urheiluedonlyönnin saralle.

Kirjallisuuskatsauksen toisessa alaluvussa havaittiin, että keskeisimmät uhkapelin ja sijoitustoiminnan väliset erot liittyvät panoksen kokoon suhteessa omiin käytettävissä oleviin varoihin. Luvussa esiteltyjen määritelmien kautta, oikein harjoitettu urheiluviedonlyönti on lähempänä perinteistä sijoitustoimintaa kuin uhkapelaamista. Täten urheiluviedonlyöntiä voidaan pitää yhtenä kaupankäynnin muotona, jota on mahdollista harjoittaa sijoitustoimintana, jossa on tarkoitus tehdä voittoa riittävän pitkällä aikavälillä.

Tutkielman kirjallisuuskatsauksen kolmannessa alaluvussa käytiin läpi tiedonlouhintateknologiaa urheiluviedonlyönnissä. Alaluvussa selvitettiin ensin viedonlyöntimarkkinoiden tehokkuutta ja pyrittiin vastaamaan kirjallisuuskatsauksen toiseen apuutkimuskysymykseen: ”Onko viedonlyöntimarkkina tehokas?”

Apuutkimuskysymykseen vastatessa selvitettiin viedonlyöntimarkkinoiden tehokkuutta käymällä läpi aiheesta tehtyjä tutkimuksia. Muun muassa Golec ja Tamarkin (1991) sekä Gray ja Gray (1997) olivat havainneet tutkimuksissaan tehottomuutta amerikkalaisen jalkapallon NFL:n viedonlyöntikertoimissa. Lisäksi Katic ym. (2008) osoittivat tutkimuksessaan, että yksinkertaisella tietokoneohjelmalla pystyttiin hankkimaan tuottoa viedonlyöntimarkkinoilta. Edellä mainittujen lisäksi myös Hvattum (2013) kykeni saamaan suhteellisen lyhyellä, kahden vuoden ajanjaksolla ylituottoa viedonlyöntimarkkinoilta Monte Carlo -simulaatiota hyödyntäen.

Kirjallisuuskatsauksessa saatiin selville, että viedonlyöntimarkkina on tehokas, mutta ei kuitenkaan täysin tehokas. Oleellista olisikin tutkia, että onko viedonlyöntimarkkina niin tehokas, ettei tiedonlouhintateknologialla ole mahdollisuutta saavuttaa ylituottoa markkinoilta. Tehtävää vaikeuttaa vielä entisestään vedonvälittäjien palautusprosentti, joka toimii vedonvälittäjille ikään kuin tietynlaisena virhemarginaalina. Tämä aiheuttaa sen, että edes tarkintaan algoritmi ei välttämättä pysty saavuttamaan ylituottoa viedonlyöntimarkkinoilta pitkällä aikavälillä.

Tutkimukset ovat kuitenkin osoittaneet, että tietoa keräämällä ja sitä tietokannoista louhimalla kykenee rakentamaan algoritmeja, jotka saavuttavat paremman tuloksen kuin puhdas arvaaminen tai vetojen asettaminen esimerkiksi pelkkien kotijoukkueiden voiton puolesta. Mutta tieteellisiä todisteita tutkimustulosten muodossa ei ole sen puolesta, että tiedonlouhintateknologiaa hyödyntämällä voisi varmuudella saavuttaa voittoa pitkällä aikavälillä viedonlyöntimarkkinoilta.

Kun kirjallisuuskatsauksen apuutkimuskysymyksiin oli aiempaa tutkimustietoa hyödyntäen vastattu, tutkielmassa alettiin etsimään vastausta kirjallisuuskatsauksen puitteissa tutkimuksen päätutkimuskysymykseen, joka on: ”Onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuutta saavuttaa ylituottoa urheiluviedonlyönnissä?”

Päätutkimuskysymystä lähestyttiin esittelemällä tiedonlouhintaa yleisellä tasolla. Hand, Mannila sekä Smyth (2001, 6) määrittivät tiedonlouhinnan olevan usein suurten datamäärien analysointia, jossa tavoitteena on havaita odottamattomia suhteita ja tehdä yhteenvetoa datasta uudenlaisin tavoin, jotka ovat

datan haltijalle ymmärrettäviä sekä hyödyllisiä. Urheiluedonlyönnissä tiedonlouhintaa hyödyntäen voidaan rakentaa algoritmeja ennustamaan urheilutapahtumien lopputuloksia. Tutkielmassa käytiin läpi Zdravevskin ja Kulakovin (2010) sekä Caon (2012) suorittamia tutkimuksia apuna käyttäen keinoja hyödyntää tiedonlouhintaa urheiluedonlyönnissä.

Kirjallisuuskatsauksen luvussa 2.3.5 esiteltiin havaittuja tutkimustuloksia tutkimuksista, jotka käsittelivät tiedonlouhintateknologiaa hyödyntäen rakenneltuja urheilutapahtumien lopputuloksia ennustavia simulaatiomalleja. Yhdessäkin alaluvussa 2.3.5 esitellyistä tutkimuksista ei saavutettu sellaisia tuloksia, että voisi varmuudella sanoa simulaatiomallista olevan kiistatonta hyötyä urheiluedonlyönnissä. Tulokset kuitenkin osoittivat, että mitä enemmän dataa tietokantaan kerätään louhittavaksi, sitä tarkempia ennusteita saadaan.

3.2 Kirjallisuuskatsauksen johtopäätökset

Tutkielman johdannossa esiteltiin tutkimuksen päätutkimuskysymys sekä kaksi apututkimuskysymystä. Tutkimuskysymyksiin pyrittiin ensin vastaamaan kirjallisuuskatsauksen puitteissa hyödyntämällä aikaisempaa aiheesta olevaa tutkimustietoa. Tutkimuksen päätutkimuskysymys oli seuraava:

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?

Apututkimuskysymykseni olivat seuraavanlaiset:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?
- Onko vedonlyöntimarkkina tehokas?

Tutkimustulokset vedonlyöntimarkkinoiden tehokkuuden suhteen osoittivat, että markkina ei ole täysin tehokas, täten siis olisi teoriassa mahdollista saavuttaa voittoa vedonlyöntimarkkinoilta. Tutkimuksissa kuitenkin muistutettiin vedonvälittäjän palautusprosentista, joka vaikeuttaa voiton saavuttamista huomattavasti.

Käsiteltyt tulokset tiedonlouhintateknologiaa hyödyntävistä tutkimuksista osoittivat, että yhdelläkään algoritmilla ei saavutettu sellaisia tuloksia, että voisi varmuudella sanoa algoritmista olevan kiistatonta hyötyä urheiluedonlyönnissä.

Tutkimukset ovat kuitenkin osoittaneet, että tietoa keräämällä ja sitä tietokannoista louhimalla kykenee rakentamaan algoritmeja, jotka saavuttavat paremman tuloksen kuin puhdas arvaaminen tai vetojen asettaminen esimerkiksi pelkkien kotijoukkueiden voiton puolesta. Tieteellisiä todisteita tutkimustulosten muodossa ei kuitenkaan ole sen puolesta, että tiedonlouhintateknologiaa hyödyntämällä voisi saavuttaa voittoa pitkällä aikavälillä vedonlyöntimarkkinoilta.

Tutkimuksia tiedonlouhintateknologian hyödyntämisestä urheiluedonlyönnissä on tehty melko vähän. Sen sijaan tutkimuksia vedonlyöntimarkkinoiden tehokkuudesta löytyy suhteellisen paljon, kuin myös tutkimuksia tiedonlouhintateknologiaa hyödyntävistä simulaatiomalleista.

Seuraavaksi tutkielman empiirisen tutkimuksen osuudessa tutkitaan tiedonlouhintateknologian hyödyntämistä nimenomaan urheiluedonlyönnin näkökulmasta, sillä läpikäydyistä aiheita koskevista tutkimuksista puuttui kokonaan vedonlyöntinäkökulma. Tutkimuksen empiirisessä osuudessa pyritään vastaamaan kerätyn datan avulla tutkielman päätutkimuskysymykseen: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?" sekä keräämään lisää tietoa vastaamalla tutkielman apu-tutkimuskysymyksiin: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?" ja "Onko vedonlyöntimarkkina tehokas?".

4 Empiirinen tutkimus

Tutkielman neljännessä pääluvussa esitellään tutkielman empiirisen tutkimuksen osio. Ensin esitellään tutkimuksen tavoite, jonka jälkeen käydään läpi tutkimusmenetelmät sekä tutkimuksen toteutus. Tutkimuksen toteutus pitää sisällään taustatietoja tutkimuksessa käytetyn datan alkuperästä, aineiston keruun sekä analysoinnin. Lisäksi luvussa esitellään empiirisen tutkimuksen vaiheet.

4.1 Tutkimuksen tavoite

Tässä tutkielmassa perehdytään tiedonlouhintateknologian hyödyntämiseen urheiluedonlyönnissä. Tutkielman tarkoituksena on selvittää, onko tiedonlouhintateknologian avulla mahdollista laatia vedonlyöntimarkkinaa täsmällisempiä todennäköisyysarvioita tulevien urheilutapahtumien lopputuloksista ja näitä hyödyntämällä saavuttaa rahallista ylituottoa vedonlyöntiin panostamalla. Tutkimuksen päätutkimuskysymykseksi määriteltiin:

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?

Päätutkimuskysymykseen vastaamalla on tarkoituksena tarjota uutta tutkimustietoa siitä, onko dataa keräämällä ja sitä urheiluedonlyönnin tarpeisiin jalostamalla kyennyt historiallisesti tekemään ylituottoa vedonlyöntimarkkinoilla. Oletuksena on, että todennäköisyyslaskennan prosessi yrityksestä ja toimijasta riippumatta on suhteellisen samankaltainen, joten tutkielman tarkoituksena on havainnollistaa ja lisätä ymmärrystä aiheesta, josta aikaisempaa tutkimustietoa ei vielä ole olemassa. Tutkielman kirjallisuuskatsauksen osiossa haettiin tietoa päätutkimuskysymykseen vastaamiseksi aiempiin aiheesta tehtyihin tutkimukseen pohjautuen. Tutkimuksen empiirisen tutkimuksen osiossa vastausta päätutkimuskysymykseen lähdetään etsimään dataan perustuen ja siitä laskelmia tehden.

Tutkimuksen kohteena olevan aiheen syvällisemmän ymmärryksen hankkimiseksi sekä aineiston analysoimisen helpottamiseksi päätutkimuskysymyksen rinnalle määriteltiin kaksi apututkimuskysymystä:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?
- Onko vedonlyöntimarkkina tehokas?

Yllä mainittuihin apututkimuskysymyksiin saatiin hankittua vastauksia tutkielman kirjallisuuskatsausosiossa, jonka pohjalta näitä oletuksia käyttäen lähdetään etenemään tutkielman empiirisen tutkimuksen osioon. Tutkielman kirjallisuuskatsauksessa selvitettiin, että urheiluedonlyöntiä voidaan harjoittaa sijoitustoimintana. Lisäksi vedonlyöntimarkkinan havaittiin olevan suhteellisen tehokas, mutta ei kuitenkaan niin tehokas, etteikö se avaisi mahdollisuuksia ylituoton saavuttamiseksi.

4.2 Tutkimusmenetelmät ja tutkimuksen toteutus

Tutkimus toteutetaan kvantitatiivisena eli määrällisenä tutkimuksena. Tutkimuksessa on käytetty hyödyksi suomalaisen, urheilutapahtumien todennäköisyyslaskentaan erikoistuneen yhtiön nimeltään Accuscore, tuottamaa historiallista dataa amerikkalaisen jalkapallon NFL-liigasta vuosilta 2011-2018.

Kvantitatiivista eli määrällistä menetelmää käyttävä tutkimus kuvaa ja tulkitsee ilmiöitä mittaamenetelmillä, jotka keräävät numeerisia tutkimusaineistoja (Vilpas, 2018). Määrällinen tutkimus on menetelmä, joka antaa yleisen kuvan muuttujien (mitattavat ominaisuudet) välisistä suhteista ja eroista. Se vastaa kysymykseen, kuinka paljon tai miten usein (Vilkka, 2007). Vilpas (2018) selvittää, että määrällinen tutkimus perustuu mittaamiseen, jonka tuloksena syntyy lukuarvoja sisältävä havaintoaineisto, jota analysoidaan tilastollisin analyysimenetelmin. Kvantitatiivisen tutkimuksen tavoitteena on tuottaa yleistettävissä olevia tutkimustuloksia (Thomas, 2003). Thomasin (2003) mukaan määrällisessä tutkimuksessa tutkijan rooli on tutkia ja mitata, objektiivisuuteen pyrkien. Vilkan (2007) mukaan tutkimustulos on objektiivinen silloin, kun se on tutkijasta riippumaton.

Tämä tutkimus perustuu Accuscorelta saatuun dataan, jota on täydennetty muista lähteistä kerätyllä datalla. Tästä lukuarvoja sisältävästä havaintoaineistosta on tutkielman puitteissa tehty erinäisiä laskelmia tutkimuskysymyksiin vastaamiseksi. Tutkimus perustuu ainoastaan kerättyyn dataan ja on täysin riippumaton tutkijasta itsestään.

Vilkka (2007) kirjoittaa, että kvantitatiivisessa tutkimuksessa tutkija tulkitsee ja selittää olennaisen numerotiedon sanallisesti. Vilkan (2007) mukaan tarkoituksena on kuvata, millä tavalla eri asiat liittyvät toisiinsa tai eroavat toisensa suhteen. Tutkimuksessa kerätty data analysoidaan ja analyysin pohjalta

tehdään johtopäätökset, joiden perusteella vastataan tutkimuksen tutkimuskysymyksiin.

4.2.1 Taustatietoja

Empiirisen tutkimuksen pohjana käytettiin suomalaisen, urheilutapahtumien todennäköisyyslaskentaan erikoistuneen Accuscore -nimisen yrityksen tuottamaa historiallista dataa amerikkalaisen jalkapallon NFL-liigan otteluista.

”Accuscore on vuodesta 2004 toiminut urheiluanalytiikkaan erikoistunut yritys. Päämarkkinamme ovat perinteisesti olleet USA:ssa. Accuscore on koko toimintansa ajan aktiivisesti kehittänyt simulointimalleja eri urheilulajeihin ja -liigoihin. Kehittämämme algoritmit pohjautuvat suuressa osassa Monte Carlo -metodiin ja laajaan kvantitatiiviseen tietämykseen simuloitavasta tapahtumasta. Käytämme historiallista dataa ja laajan määrän tärkeäksi havaittuja muuttujia varmistaaksemme korkean osumatarkkuuden simuloitavissa kohteissa. Yleensä yksittäinen ottelu tai muu tapahtuma simuloidaan vähintään 10000 kertaa.” (Accuscore, 2020).

Tutkimuksessa päädyttiin käyttämään nimenomaan amerikkalaisen jalkapallon NFL:sta kerättyä dataa siitä syystä, että kyseistä urheiluliigaa pidetään kilpailuimpana maailman vedonlyöntimarkkinoilla. Esimerkiksi Mani (2018) kertoo tutkimuksessaan, että NFL:n vedonlyöntimarkkina kerää suurimman osuuden koko maailman vedonlyöntimarkkinasta ja soveltuu täten hyvin vedonlyöntimarkkinan tehokkuutta mittaavien tutkimusten kohteeksi. Tästä johtuen kertoimiin on todennäköisesti hinnoiteltu tehokkaammin informaatiota kuin pienemmän pelivaihdon liigojen kertoimiin. Koska NFL:n vedonlyöntimarkkina on niin suuri, myös vedonvälittäjät tarjoavat sarjan otteluihin pelikohteita suuremmalla palautusprosentilla kuin pienemmän pelivaihdon liigojen otteluihin. NFL:sta on myös saatavilla todella paljon dataa analysoimisen helpottamiseksi ja todennäköisyysarvioiden laatimiseksi, joten Accuscoren todennäköisyysarvioiden voi olettaa sisältävän kaiken olennaisen saatavilla olevan informaation.

Lisäksi NFL soveltuu urheiluliigana tämän tutkimuksen tarpeisiin sen vuoksi, että liigassa pelataan yhden kauden aikana 267 ottelua. Suhteellisen pieni kausittainen ottelumäärä mahdollistaa sen, että tämän tutkielman puitteissa dataa kyetään analysoimaan useammalta kaudelta, eikä esimerkiksi ainoastaan yhdeltä kaudelta, kuten jonkin muun urheiluliigan tapauksessa, jossa kausittaiset ottelumäärät ovat suurempia. Tutkielmassa dataa on kerätty yhteensä kahdeksalta eri kaudelta, tämä tekee tutkimustuloksista luotettavampia ja vähentää sattuman merkitystä tutkimustuloksissa, kun tulokset eivät perustu ainoastaan esimerkiksi yhden kauden aikana pelattuihin otteluihin.

4.2.2 Aineiston keruu

Aineiston keruu on toteutettu käyttämällä Accuscoren tuottamaa historiallista dataa amerikkalaisen jalkapallon NFL-liigasta kahdeksan pelikauden ajalta,

vuosilta 2011-2018. Kyseisellä ajanjaksolla Accuscore on tuottanut todennäköisyysarviot kaikkiaan 2136 otteluun.

Tämän jälkeen aineistoon on haettu vedonlyöntimarkkinan kertoimet vastaaviin otteluihin. Kerroindata on haettu Oddsportal-kerroinvertailusivustolta. Tutkimuksessa kerroindatana on käytetty kunkin ottelun keskiarvokerrointa, joka muodostuu kymmenien eri vedonvälittäjien kertoimista kuhunkin otteluun. Keskiarvokerrointen käyttäminen kuvastaa parhaiten koko vedonlyöntimarkkinaa.

Tutkimuksessa Accuscoren todennäköisyysarvioita on verrattu kunkin ottelun alkamisajankohdan mukaisiin keskiarvokertoimiin ja näiden pohjalta on tehty laskelmia. Tutkimuksessa on otettu huomioon ainoastaan sellaiset kertoimet, joissa odotusarvo Accuscoren todennäköisyysarvion mukaan on yli 1,00, eli kertoimet, jotka ylittävät pelikelpoisuuden rajan. Mikäli Accuscoren arviot ovat oikeassa, tällaisissa pelikohteissa tuoton tulisi pitkällä aikavälillä olla positiivinen.

4.2.3 Empiirisen tutkimuksen vaiheet

Tutkielman empiirisen tutkimuksen osuus koostui kaikkiaan kolmesta eri vaiheesta. Ensin tutkimuksessa käytettävä data hankittiin Accuscorelta, sen jälkeen Accuscoren dataa on täydennetty hakemalla Excel-tiedostoon otteluita vastaavat keskiarvokertoimet Oddsportal-kerroinvertailusivustolta ja tämän jälkeen tehty asianmukaiset laskelmat Excelissä, joiden pohjalta vastataan tutkielman tutkimuskysymyksiin. Seuraavaksi esitellään tarkemmin empiirisen tutkimuksen vaiheet vaihe vaiheelta.

Ylläolevassa kuviossa (Kuvio 6) näkyy Accuscorelta saatu Excel-tiedosto, jossa on laskettu todennäköisyysarviot sekä koti- että vierasjoukkueen voitolle jokaiseen NFL-kauden aikana pelattuun otteluun vuosilta 2011-2018. Kuviossa 5 on nähtävillä kotijoukkueen nimi sarakkeessa F ja vierasjoukkueen nimi puolestaan sarakkeessa C.

Sarakkeessa G näkyy Accuscoren todennäköisyysarvio kotijoukkueen voitolle ja sarakkeessa D todennäköisyysarvio vierasjoukkueen voitolle. Sarakkeissa I ja F näkyy koti- sekä vierasjoukkueen pistemäärät kyseisessä ottelussa ja sarakkeessa J puolestaan se, olisiko Accuscoren todennäköisyysarvion mukainen ottelun ennakkosuosikki voittanut kyseisen ottelun.

Sarakkeet A ja B kertovat kyseisen ottelun järjestysnumeron kunkin kauden aikana (A) sekä päivämäärän, jolloin ottelu on pelattu (B). Sarakkeessa I on ilmoitettu, olisiko Accuscoren todennäköisyysarvion mukainen ottelun ennakkosuosikki voittanut vai hävinnyt kyseisen ottelun.

Tämän tutkimuksen kannalta relevantteja ovat ennen kaikkea sarakkeet G ja D, joissa ilmoitettuja Accuscoren todennäköisyysarvioita koti- ja vierasjoukkueiden voitolle on verrattu vedonlyöntimarkkinan määrittämiin kertoimiin.

Los Angeles Rams - New England Patriots

Monday, 04 Feb 2019, 01:30

Final result 3:13 (0:0, 0:3, 3:0, 0:10)

1X2		Home/Away	AH	O/U	HT/FT	O/E
FT including OT		1st Half	1Q	2Q	3Q	4Q
Bookmakers		1	2	Payout		
	10Bet ^o	↑ 2.10	↓ 1.74	95.2%		
	188BET ^o	↑ 2.03	↓ 1.75	94.0%		
	1xBet ^o	↑ 2.14	↓ 1.78	97.2%		
	888sport ^o	↓ 1.80	↑ 1.95	93.6%		
	bet-at-home ^o	↑ 2.15	↓ 1.74	96.2%		
	bet365 ^o	↑ 2.10	↓ 1.76	95.8%		
	Betclix ^o	↑ 2.15	↓ 1.80	98.0%		
	Betfair ^o	↑ 2.10	↓ 1.80	96.9%		
	Betfred ^o	↑ 2.20	↓ 1.73	-		
	Bethard ^o	↑ 2.10	↓ 1.87	98.9%		
	BetJOE ^o	↑ 2.13	↓ 1.82	98.1%		
	BetVictor ^o	↑ 2.05	↓ 1.70	92.9%		
	Betway ^o	↑ 2.10	↓ 1.77	96.0%		
	BoyleSports ^o	2.10	1.75	-		
	bwin ^o	↑ 2.15	↓ 1.72	95.6%		
	ComeOn ^o	↑ 2.15	↓ 1.77	97.1%		
	Coolbet ^o	↑ 2.16	↑ 1.86	99.9%		
	Expekt ^o	↑ 2.15	↓ 1.80	98.0%		
	Intertops ^o	↑ 2.15	↓ 1.74	96.2%		
	Interwetten ^o	↑ 2.15	↓ 1.70	94.9%		
	Jetbull ^o	↑ 2.14	↓ 1.77	-		
	Leonbets ^o	↑ 2.17	↓ 1.76	97.2%		
	Oddsring ^o	↑ 2.17	↓ 1.76	97.2%		
	Pinnacle ^o	↑ 2.12	↓ 1.81	97.6%		
	SBOBET ^o	↑ 2.16	↓ 1.74	96.4%		
	Sportingbet ^o	↑ 2.15	↓ 1.72	95.6%		
	Titanbet ^o	↑ 2.15	↓ 1.73	95.9%		
	Unibet ^o	↓ 1.80	↑ 1.95	93.6%		
	William Hill ^o	↑ 2.15	↓ 1.74	96.2%		
	youwin ^o	↑ 2.03	↓ 1.73	93.4%		
= Average		2.10	1.78	96.3%		
↑ Highest		2.17	1.95	102.7%		

Kuvio 7 Oddsportal.com -kerroinvertailusivuston näkymä

Ylläolevassa kuvassa (Kuvio 7) näkyy kuvankaappaus www.oddsportal.com -verkkosivulta. Oddsportal on verkossa toimiva kerroinvertailus sivusto, josta on haettavissa historiallista kerroindataa haluttuihin otteisiin.

Tässä tutkimuksessa on hyödynnetty vedonlyöntimarkkinan keskiarvokertoimia, jotka ovat merkitty Kuviossa 7 kuvan alaosassa näkyvällä punaisella laatikolla. Keskiarvokertoimia päädyttiin käyttämään sen vuoksi, sillä ne kuvaavat parhaiten koko vedonlyöntimarkkinaa sen sijaan, että tutkimuksessa olisi hyödynnetty esimerkiksi vain yhden vedonvälittäjän tarjoamia kertoimia.

Kuviossa 7 Oddsportalin sivustolta otetussa kuvankaappauksessa kaikkiaan 30 eri vedonvälittäjää tarjoavat kertoimia Los Angeles Rams – New England Patriots -otteluun. Keskiarvokerroin on kaikkien kuvassa näkyvien 30 eri vedonvälittäjän tarjoaman kertoimen keskiarvo. Kyseisessä ottelussa kotijoukkueen keskiarvokerroin on 2,10 ja vierasjoukkueen keskiarvokerroin 1,78. Kyseisiä keskiarvokertoimia on hyödynnetty tässä tutkimuksessa ja sama on tehty kaikille muillekin tutkimuksessa analysoiduille 2136 ottelulle.

Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
KOTIKERROIN	VIERASKERROIN			pa%	kotiOA	vierasOA				kotivoittotasapeli	koti_AC	koti_mark	kotikelly	vieraskell		osuma	osum.yht	kotipanos	vieraspan	Panos	ToWin	
						81	128										204,23					1000
1,92	1,92			0,9600	0,952302	1,040	0,872	kotisuosit_AC	173	TRUE	FALSE	TRUE	FALSE	0,006191	-0,01981	TRUE	1,92	6,191304	0	6,191304	11,8873	1005,696
2,51	1,55			0,5583	0,965732	0,777	1,064	vierassuoskit_AC	94	FALSE	TRUE	FALSE	FALSE	-0,02109	0,016603	FALSE		0	16,69717	16,69717	25,88061	988,998
1,93	1,91			0,9600		1,147	0,768			FALSE	FALSE	TRUE	FALSE	0,02261	-0,03645	FALSE		22,3614	0	22,3614	43,1575	966,6374
1,93	1,91			0,9600		1,019	0,892	kotisuosit_mark	175	TRUE	FALSE	TRUE	FALSE	0,002925	-0,01702	TRUE	1,93	2,827155	0	2,827155	5,456408	969,2667
1,37	3,19			0,5584		0,826	1,254	vierassuoskit_mar	92	TRUE	FALSE	TRUE	TRUE	-0,06703	0,016589	FALSE		0	16,07908	16,07908	51,29226	953,1876
1,39	3,07			0,5568		0,996	0,855			TRUE	FALSE	TRUE	TRUE	-0,00154	-0,01003	FALSE		0	0	0	0	953,1876
1,20	4,77			0,5588		0,939	1,025	kotivoitot	159	FALSE	FALSE	TRUE	TRUE	-0,04331	0,009932	TRUE	4,77	0	0,88839	0,88839	4,237619	956,5368
2,23	1,69			0,9614		1,132	0,825	vierasvoitot	106	FALSE	FALSE	TRUE	FALSE	0,015325	-0,03615	FALSE		14,6589	0	14,6589	32,68935	941,8779
1,27	3,93			0,5598		0,976	0,896	tasapelit	2	TRUE	FALSE	TRUE	TRUE	-0,0125	-0,00507	FALSE		0	0	0	0	941,8779
1,54	2,56			0,9616		1,054	0,799			FALSE	FALSE	TRUE	TRUE	0,014279	-0,01843	FALSE	13,44942	0	13,44942	20,71211	928,4285	
1,57	2,48			0,9614		0,981	0,923	osumat	78	TRUE	FALSE	TRUE	TRUE	-0,00478	-0,00747	FALSE		0	0	0	0	928,4285
1,77	2,10			0,9605		0,967	0,943	kaikki	209	FALSE	FALSE	TRUE	TRUE	-0,00661	-0,00736	FALSE		0	0	0	0	928,4285
1,64	2,32			0,9608		0,933	0,994			TRUE	FALSE	TRUE	TRUE	-0,01499	-0,00066	FALSE		0	0	0	0	928,4285
1,33	3,62			0,5576		0,904	1,075	palautusprosentti	97,7%	TRUE	FALSE	TRUE	TRUE	-0,04139	0,004442	FALSE		0	4,124108	4,124108	14,10445	924,3044
1,30	3,65			0,5586		0,946	0,990			FALSE	FALSE	TRUE	TRUE	-0,02552	-0,00055	FALSE		0	0	0	0	924,3044
3,19	1,37			0,5584		1,120	0,884			FALSE	FALSE	FALSE	FALSE	0,007849	-0,04461	FALSE	7,255036	0	7,255036	23,14356	917,0494	
1,94	1,89			0,9573		0,859	1,043			TRUE	FALSE	FALSE	FALSE	-0,02136	0,006947	FALSE		0	6,37077	6,37077	12,04076	910,6786
1,42	2,95			0,5586		1,007	0,835			TRUE	FALSE	TRUE	TRUE	0,002548	-0,0121	TRUE	1,42	2,320062	0	2,320062	3,294488	911,653
3,69	1,30			0,9613		0,946	0,966			FALSE	FALSE	FALSE	FALSE	-0,00284	-0,01624	FALSE		0	0	0	0	911,653
2,06	1,80			0,9606		1,043	0,877			FALSE	TRUE	TRUE	TRUE	0,005848	-0,02204	FALSE	5,331082	0	5,331082	10,98203	906,322	
2,55	1,54			0,9601		1,321	0,738			TRUE	FALSE	TRUE	FALSE	0,029576	-0,0694	TRUE	2,55	26,80541	0	26,80541	68,3538	947,8703
1,21	4,56			0,5563		0,938	1,015			TRUE	FALSE	TRUE	FALSE	-0,04235	0,000586	FALSE	0	0,555333	0,555333	2,53232	947,315	
1,61	2,38			0,9604		0,634	1,438			FALSE	FALSE	FALSE	TRUE	-0,08563	0,045292	TRUE	2,38	0	42,90572	42,90572	102,1156	1006,525
1,49	2,69			0,5589		0,988	0,902			FALSE	FALSE	TRUE	TRUE	-0,00354	-0,00824	FALSE		0	0	0	0	1006,525
2,29	1,65			0,5590		0,739	1,115			TRUE	FALSE	FALSE	FALSE	-0,02896	0,025181	FALSE		0	25,34562	25,34562	41,82028	981,1793
1,40	3,03			0,5576		0,850	1,180			FALSE	FALSE	TRUE	TRUE	-0,05364	0,01268	TRUE	3,03	0	12,44151	12,44151	37,69776	1006,436
1,12	6,57			0,5569		0,748	2,158			TRUE	FALSE	TRUE	TRUE	-0,30048	0,029706	FALSE		0	29,89738	29,89738	196,4258	976,5381
1,39	3,07			0,5568		0,817	1,249			TRUE	FALSE	TRUE	TRUE	-0,06717	0,017218	FALSE		0	16,81411	16,81411	51,61933	959,724
1,40	3,03			0,5576		0,850	1,176			TRUE	FALSE	TRUE	TRUE	-0,05364	0,01236	FALSE		0	11,86249	11,86249	35,94333	947,8615
2,09	1,77			0,5584		0,993	0,921			TRUE	FALSE	FALSE	FALSE	-0,00095	-0,0146	FALSE		0	0	0	0	947,8615
1,65	2,28			0,5573		0,843	1,098			TRUE	FALSE	TRUE	TRUE	-0,03447	0,010917	FALSE	0	10,34819	10,34819	23,59388	937,5134	
1,45	2,83			0,5588		0,909	1,051			TRUE	FALSE	TRUE	TRUE	-0,02884	0,004008	FALSE	0	3,757738	3,757738	10,6344	933,7556	
1,57	2,49			0,9629		0,729	1,325			TRUE	FALSE	FALSE	TRUE	-0,06785	0,031129	FALSE		0	29,06728	29,06728	72,37754	904,6883
1,79	2,07			0,5599		0,820	1,110			FALSE	FALSE	FALSE	TRUE	-0,03258	0,014622	TRUE	2,07	0	13,2285	13,2285	27,383	918,8428
1,39	3,12			0,9616		0,986	0,900			TRUE	FALSE	TRUE	TRUE	-0,00531	-0,00673	FALSE		0	0	0	0	918,8428
1,68	2,24			0,9600		0,993	0,902			TRUE	FALSE	TRUE	TRUE	-0,00015	-0,01134	FALSE		0	0	0	0	918,8428

Kuvio 8 Excel-laskelmat

Kuviossa 8 Accuscorelta hankittua dataa on täydennetty ensin Oddsportal-kerroinvertilusivuston kerroindatalla (sarakkeet R ja S). Tämän jälkeen datasta on tehty erinäisiä laskelmia, joiden pohjalta pyritään vastaamaan tutkielmassa esitettyihin tutkimuskysymyksiin. Datasta tehtyjä laskelmia avataan enemmän seuraavassa aluvuossa 4.2.4 Aineiston analysointi.

4.2.4 Aineiston analysointi

Tutkimuksen aineistoa analysoitiin kvantitatiivisten eli määrällisten analysointimenetelmien avulla. Laadulliset analysointimenetelmät eivät soveltuneet hyvin tutkimuksen kohteena olevan aiheen tutkimiseen, sillä tutkimusaineisto sisältää paljon numeerista dataa ja tarkoituksena on antaa yleinen kuva muuttujien välisistä suhteista ja eroista.

Todennäköisyysarvio- ja kerroindatasta kerätyn aineiston analysointi toteutettiin tekemällä Exceliä käyttäen erinäisiä laskelmia, joiden avulla pystyttiin lähestymään tutkielman tutkimuskysymyksiin vastaamista. Ensin kaikki tutkimuksen kannalta relevantti data kerättiin ensin yhteen Excel-tiedostoon.

Tutkimuksen päätutkimuskysymystä lähdettiin lähestymään luomalla Excel-tiedostoon kuvitteellinen 1000 yksikön niin sanottu pelikassa. Tarkoituksena oli kullakin kahdeksalla tutkimuksen kohteena olleella kaudella testata, olisiko tämä kassa kasvanut, jos kyseisen pelikauden aikana pelattuihin otteluihin olisi panostanut Kellyn kaavan (alaluku 2.2.4) mukaisesti Accuscoren todennäköisyysarvioilla ja annetuilla kertoimilla.

Kun Excel-tiedostossa oli kerättyä sekä Accuscoren tuottama data todennäköisyysarvioista sekä Oddsportalilta haettu data historiallisista kertoimista vastaaviin otteluihin, laskettiin aluksi kunkin ottelun odotusarvot annetuilla kertoimilla sekä laadituilla todennäköisyysarvioilla. Kuten tutkielman alaluvussa 2.1.6 havainnollistettiin, vedonlyönnissä odotusarvo saadaan laskemalla vedonvälittäjän tarjoaman kertoimen ja vedonlyöjän oman todennäköisyysarvion tulo. Samassa alaluvussa myös huomioitiin, että rationaalinen vedonlyöjä pelaa ainoastaan pelejä, joiden odotusarvo ylittää yhden. Tutkimuksessa huomioon otettiin vain ne ottelut, joissa odotusarvo on yli 1,00.

Kun datasta saatiin eroteltua kohteet, joissa odotusarvo on yli tai alle yksi, täytyi seuraavaksi alkaa laskemaan optimaalista panoskokoa arvioidulla todennäköisyysarviolla ja annetulla vedonlyöntikertoimella vastaavaan otteluun. Tutkielmassa optimaalinen panoskoko saatiin Kellyn kaavaa noudattaen. Kuten alaluvussa 2.2.4 käytiin läpi, Kellyn kaava voidaan esittää muodossa: $P = (AB - 1) / (B - 1)$. Optimaalinen panoskoko saadaan määriteltä, kun tiedossa on kohteen todennäköisyysarvio ja kerroin. Kaavassa A on todennäköisyysarvio ja B kohteen kerroin. Joten panoskoko on sitä suurempi, mitä enemmän Accuscoren todennäköisyysarvio eroaa kohteeseen tarjotusta kertoimesta.

AI	AJ	AK	AL	AM
kotipanos	vieraspan	Panos	ToWin	
				1000
6,191304	0	6,191304	11,8873	1005,696
0	16,69717	16,69717	25,88061	988,9988
22,3614	0	22,3614	43,1575	966,6374
2,827155	0	2,827155	5,456408	969,2667
0	16,07908	16,07908	51,29226	953,1876
0	0	0	0	953,1876
0	0,88839	0,88839	4,237619	956,5368
14,6589	0	14,6589	32,68935	941,8779
0	0	0	0	941,8779
13,44942	0	13,44942	20,71211	928,4285
0	0	0	0	928,4285
0	0	0	0	928,4285
0	0	0	0	928,4285

Kuvio 9 Esimerkki panostamiseen

Yllä olevassa kuviossa (Kuvio 9) havainnollistetaan tutkielmassa suoritettuja laskelmia panoskoon optimoimiseksi. Sarakkeessa AI näkyy optimaalinen panoskoko kotijoukkueelle annetuilla todennäköisyysarvioilla sekä kertoimella. Sarakkeessa AJ puolestaan sama vierasjoukkueen osalta. Jos sarakkeessa on 0, se merkitsee sitä, että kyseisen vedonlyöntikohteen odotusarvo on ollut alle yhden, eikä tässä tapauksessa sellaisiin kohteisiin panosteta lainkaan. Sarakkeeseen AL on laskettu potentiaalinen voitto, mikäli kohteeseen kuvitteellisesti lyöty veto menee oikein. Sarake AM puolestaan kertoo pelikassan koon lyödyn vedon jälkeen. Pelikauden alussa pelikassa on 1000 yksikköä, josta kassa lähtee sitten lyötyjen vetojen puitteissa joko kasvamaan tai laskemaan.

Kaikkiaan tutkimuksessa huomioidun ajanjakson 2011-2018 välisenä aikana pelattiin 2136 ottelua. Näistä otteluista kaikkiaan 1633 ottelua ylitti ylikertoimen rajan Accuscoren todennäköisyysarvioin mukaan ja täten otettiin analyysin kohteeksi tässä tutkimuksessa.

Kun edellä mainitulla tavalla oli käyty läpi jokaisen tutkimuksen kohteena olleen pelikauden jokainen ottelu ja tehty niistä asianmukaiset laskelmat, saatiin laskelmista havaittua tuloksia, joiden avulla vastataan tutkielman tutkimuskysymyksiin. Tutkielman tuloksia esitellään tarkemmin seuraavassa luvussa.

5 Tutkimuksen tulokset

Tässä luvussa esitellään ja käydään läpi tutkielman empiirisestä tutkimuksesta havaitut tutkimustulokset. Tutkimustulokset on saatu Excelissä tehdyillä laskelelmilla Accuscoren todennäköisyysdataa sekä Oddsportal.com -sivustolta kerättyä kerroindataa käyttämällä (ks. alaluku 4.2.3).

5.1 Todennäköisyysarvioiden yhdenmukaisuus vedonlyöntimarkkinan kanssa

Tutkimustulosten ensimmäisessä alaluvussa tarkastellaan Accuscoren todennäköisyysarvioiden yhdenmukaisuutta vedonlyöntimarkkinan kertoimien kanssa sekä lopuksi verrataan arvioita toteutuneisiin lopputuloksiin kullakin tutkimuksen kohteena olleella pelikaudella.

Taulukko 3 Kotiennakkosuosikit

	2011	2012	2013	2014	2015	2016	2017	2018
Kotisuosikit Accuscore	165	164	161	173	181	162	155	173
Kotisuosikit Vedonlyöntimarkkina	175	172	173	187	175	182	163	175

Yllä olevassa taulukossa (Taulukko 3) ylemmällä rivillä on nähtävillä Accuscoren todennäköisyysarvioiden mukaiset kotijoukkueena pelanneet ennakkosuosikit kausittain. Alemmalla rivillä on puolestaan esitetty vedonlyöntimarkkinan määrittämät kotiennakkosuosikit kausittain. Taulukosta on havaittavissa, että ainoastaan kaudella 2015 Accuscoren todennäköisyysarvioiden mukaisia kotijoukkueena pelanneita ennakkosuosikkeja oli enemmän kuin vedonlyöntimarkkinan määrittämiä kotiennakkosuosikkeja. Kaikilla muilla tutkimuksen kohteena olleella pelikaudella vedonlyöntimarkkinan mukaisia kotijoukkueena pelanneita ennakkosuosikkeja oli enemmän.

Taulukko 4 Vierasennakkosuosikit

	2011	2012	2013	2014	2015	2016	2017	2018
Vierassuosikit Accuscore	102	103	106	94	86	105	112	94
Vierassuosikit Vedonlyöntimarkkina	92	95	94	80	92	85	104	92

Yläpuolella näkyvässä taulukossa (Taulukko 4) on esitetty vielä sama vierasjoukkueille kuin mikä ensimmäisessä taulukossa kotijoukkueille. Taulukossa ylemmällä rivillä näkyy Accuscoren todennäköisyysarvioiden mukaiset vierasjoukkueena pelanneet ennakkosuosikit kausittain. Alemmalla rivillä puolestaan vedonlyöntimarkkinan määrittämät vierasennakkosuosikit kausittain.

Taulukko 5 Toteutuneet lopputulokset

	2011	2012	2013	2014	2015	2016	2017	2018
Kotivoitot	153	152	159	154	143	154	153	159
Vierasvoitot	114	114	107	112	124	111	114	106
Tasapelit	0	1	1	1	0	2	0	2

Taulukossa 5 on koottuna toteutuneet lopputulokset jokaiselta tutkimuksen kohteena olleelta pelikaudelta. Koti- sekä vierasvoittojen suhde pysyi suhteellisen samana jokaisella tutkimuksen kohteena olleella kaudella. Huomioitavaa on myös se, että jokaisella kahdeksalla tutkimuksen kohteena olleella kaudella vierasvoittoa tuli lopulta enemmän kuin mitä Accuscore sekä vedonlyöntimarkkina odottivat.

5.2 Odotusarvo

Tutkimustulosten esittelyn toisessa alaluvussa perehdytään Accuscoren todennäköisyysdatasta johdettuihin odotusarvoihin.

Taulukko 6 Vedonlyöntikohteiden jakauma koti- ja vierasjoukkueiden välillä

	2011	2012	2013	2014	2015	2016	2017	2018	Yht.
Vedonlyöntikohteet Kotijoukkueet	63	65	77	84	100	72	79	81	621
Vedonlyöntikohteet Vierasjoukkueet	118	125	131	124	119	135	132	128	1012
Vedonlyöntikohteet Yhteensä	181	190	208	208	219	207	211	209	1633

Taulukossa 6 on eriteltyä vedonlyönnin kohteeksi päätyneiden otteluiden määrät koti- ja vierasjoukkueiden välillä sekä vedonlyöntikohteiden yhteismäärät. Eli sellaiset kohteet, joissa odotusarvo Accuscoren todennäköisyysarvion

mukaan oli yli 1,00 suhteessa vedonlyöntimarkkinan määrittämiin kertoimiin vastaaviin otteluihin. Vedonlyönnin kohteeksi päätyi kaikkiaan 1633 ottelua tutkimuksen kohteena olleiden kahdeksan pelikauden ajanjaksolla. Yhteensä tällä ajanjaksolla pelattiin 2136 ottelua, joten odotusarvo ylitti yhden rajan 76,5% kyseisellä ajanjaksolla pelatuista otteluista. Pyöreästi sanottuna siis kolme neljästä 2011-2018 välillä pelatusta NFL-ottelusta ylitti pelikelpoisuuden rajan Accuscoren määrittämän todennäköisyysarvion mukaan (Taulukko 6).

Huomioitavaa on myös se, että jokaisella tutkimuksen kohteena olleella pelikaudella vierasjoukkueita nousi huomattavasti enemmän pelikohteiksi kotijoukkueisiin verrattuna. Yhteensä tutkimuksen kohteina olleilla kahdeksalla pelikaudella Accuscoren arvioiden mukaan vierasjoukkueiden voittoon olisi päädytty panostamaan yhteensä 1012 kertaa, kun taas kotijoukkueiden voittoon ainoastaan 621 kertaa (Taulukko 6).

Taulukko 7 Odotusarvojen suuruus

	1 - 1,1	1,1 - 1,2	1,2 - 1,3	1,3 - 1,4	1,4 - 1,5	>1,5	YHT.
2011	86	38	36	11	6	4	181
2012	95	45	24	16	7	3	190
2013	73	46	37	28	13	11	208
2014	75	40	34	23	16	20	208
2015	78	50	27	25	17	22	219
2016	86	54	34	19	6	8	207
2017	95	47	27	19	11	12	211
2018	77	45	29	18	13	27	209
YHT.	665	365	248	159	89	107	1633

Yllä olevassa taulukossa (Taulukko 7) on jaoteltu pelikohteet kausittain Accuscoren todennäköisyysarvioista johdettujen odotusarvojen mukaisesti. Taulukosta voidaan nähdä, että eniten vedonlyönnin kohteiksi päätyneistä otteluista selvästi suurimmassa osassa kohteen odotusarvo on ollut välillä 1-1,1. Sitä mukaa, kun odotusarvo on kasvanut, on pelikohteiden määrä puolestaan vähentynyt. Sama trendi on havaittavissa jokaisella tutkimuksen kohteena olleella kaudella. Melko useassa kohteessa kausittain on odotusarvo Accuscoren todennäköisyysarvion mukaan ollut jopa yli 1,5 (Taulukko 7).

5.3 Kohteiden osumatarkkuus

Tässä alaluvussa tarkastellaan pelikohteiksi valikoitujen vedonlyöntikohteiden osumatarkkuutta tutkimuksen kohteena olleiden kahdeksan pelikauden ajalta. Tutkimuksen kohteena olleen aikavälin aikana pelikohteiksi valikoitui kaikkiaan 1633 kohdetta, joissa odotusarvo ylitti rajan 1,00.

Taulukko 8 Osumatarkkuus

	2011	2012	2013	2014	2015	2016	2017	2018	YHT.
Osumat	73	64	80	94	82	99	91	78	661
Pelikohteet	181	190	208	208	219	207	211	209	1633
Osuma%	40,3%	33,7%	38,5%	45,2%	37,4%	47,8%	43,1%	37,3%	
Osun. Kert.	174,24	153,52	178,52	215,74	213,20	210,22	190,10	204,23	
Avg_kerroin	2,39	2,40	2,23	2,30	2,60	2,12	2,09	2,62	

Taulukossa 8 on eritelty osumat jokaisen pelikauden osalta. Osumalla tarkoitetaan vedonlyönnissä oikein mennyttä pelikohdetta. Eli sellaista pelikohdetta, jossa vedonlyönnin kohteeksi odotusarvon perusteella valikoitunut joukkue on myös voittanut kyseisen ottelun. Taulukossa näkyy osumat, kaikki kauden aikana vedonlyönnin kohteeksi päätyneet ottelut sekä kauden vedonlyöntikohteiden osumaprosentti (Taulukko 8).

Jokaisella tutkimuksen kohteena olleella pelikaudella alle puolet lyödyistä vedoista meni oikein. Paras osumatarkkuus tuli kaudella 2016, osumaprosentin ollessa 47,8%. Pelkkä osumaprosentti ei tietenkään itsessään kerro juuri mitään, mutta antaa kuitenkin suuntaa siitä, kuinka suuri osa kauden aikana lyödyistä vedoista olisi voittanut.

Lisäksi taulukon kahdelta alimmalta riviltä on havaittavissa kunkin pelikauden aikana osuneiden pelikohteiden yhteenlasketut kertoimet sekä osuneen pelikohteen keskimääräinen kerroin. Tästä on havaittavissa se, että jokaisella tutkimuksen kohteena olleella pelikaudella osuneiden kohteiden keskiarvoinen kerroin on ollut yli 2,00. Tämä tarkoittaa sitä, että suuri osa kauden aikana lyödyistä vedoista on kohdistunut altavastajajoukkueisiin, joka taas selittää osaltaan verrattain suhteellisen matalia osumaprosentteja (Taulukko 8).

Olenneisempaa on joka tapauksessa tarkastella palautusprosenttia, joka kertoo sen, kuinka paljon vedonlyöjän voi odottaa saavan keskimäärin voittoina takaisin pelaamistaan panoksista (alaluku 2.1.5).

5.4 Palautusprosentti

Tässä alaluvussa tarkastellaan tutkimuksen kohteina olleiden pelikausien palautusprosentteja vedonlyönnissä. Palautusprosentti kertoo, kuinka paljon rahaa prosentuaalisesti vedonlyöjä olisi saanut takaisin vetoihin sijoittamastaan summasta.

Taulukko 9 Palautusprosentit

	2011	2012	2013	2014	2015	2016	2017	2018
Palautus%	96,3%	80,8%	85,8%	103,7%	97,4%	101,6%	90,1%	97,7%
Vedonvälittäjän palautus%	95,2%	95,5%	95,5%	95,6%	95,5%	95,6%	95,7%	95,8%

Taulukossa 9 on esitetty palautusprosentit tutkimuksen kohteina olleiden pelikausien 2011-2018 ajalta. Taulukon ylempi rivi kertoo, kuinka paljon vedonlyöjä olisi kyseisillä vedoilla saanut takaisin sijoittamastaan summasta. Alempi rivi puolestaan kertoo, kuinka suuren tuoton keskimäärin vedonvälittäjät kyseisellä pelikaudella tarjoamallaan kertoimilla tekivät.

Taulukosta ylemmältä riviltä on havaittavissa, että ainoastaan kahdella pelikaudella kahdeksasta, vedonlyöjän tuotto olisi jäänyt positiiviseksi Accuscoren todennäköisyysarvioilla ja vedonlyöntimarkkinan keskiarvokertoimilla.

Kolmella kaudella kahdeksasta, vedonlyöjän vedoistaan takaisin saama palautusprosentti olisi jäänyt niukasti alle sadan. Kolmella kaudella puolestaan reilusti alle sadan, kaudella 2012 vedonlyöjän palautusprosentti oli todella vaatimaton 80,8% (Taulukko 9).

Taulukon alempi rivi puolestaan kertoo vedonvälittäjien keskimäärin vedoista tarjoaman palautusprosentin kultakin kaudelta. Vedonvälittäjien tarjoama palautusprosentti on jokaisella pelikaudella pysynyt melko tarkalleen samoissa lukemissa. Tutkimuksen kohteina olleiden kausien viimeisinä vuosina palautusprosentti on ollut hieman kiristymään päin, pienentäen vedonvälittäjille jäävää osuutta. Tämä kertonee kilpailun kiristymisestä alalla (Taulukko 9).

Kaikkiaan siis viidellä kaudella kahdeksasta, vedonlyöjän palautusprosentti olisi päihittänyt vedonvälittäjien palautusprosentin. Vedonlyöjien tuotto olisi kuitenkin jäänyt näistä viidestä kaudesta plussalle vain kahdella, johtuen juuri vedonvälittäjille jäävästä osuudesta (Taulukko 9).

Yllä olevan taulukon (Taulukko 9) palautusprosentit kertovat kuinka paljon vedonlyöjä olisi saanut panoksistaan takaisin, mikäli jokainen veto olisi lyöty samalla panoksella. Yllä olevat laskelmat eivät ota huomioon panoskoon vaihtelevuutta eri kohteiden välillä. Seuraavassa alaluvussa avataan enemmän vedonlyöjän tuottoa, jossa on otettu huomioon myös panoskokojen vaihtuvuus.

5.5 Pelikassan kehitys

Tässä alaluvussa selvitetään, millainen olisi ollut vedonlyöjän pelikassan kehitys kullakin tutkimuksen kohteena olleella pelikaudella. Tutkimuksessa panoskoon määrittämiseen käytettiin Kellyn kaavaa (alaluku 2.2.4), jossa tarkoituksena on maksimoida pääoman odotettu tuotto eli toisin sanoen maksimoida vedon odotusarvo (Thorp, 2010). Mitä suurempi on Accuscoren todennäköisyysarviosta ja vedonlyöntimarkkinan tarjoamasta kertoimesta johdettu odotusarvo, sitä suurempi on panoskoko kyseisessä pelikohteessa.

Tutkimuksessa käytettiin kuvitteellista 1000 yksikön pelikassaa, joka kohteen osumasta tai hudista riippuen joko kasvoi tai väheni. Tutkimuksessa jokaisen kohteen kuvitteellinen panos määritettiin Kellyn kaavalla, kohteen odotusarvosta johtaen. Täten jokaiselle pelikohteelle saatiin määriteltyä optimaalinen panoskoko. Tarkastellaan ensin panoskokojen suuruutta oheisen taulukon muodossa:

Taulukko 10 Panostus

	2011	2012	2013	2014	2015	2016	2017	2018
Pelikohteet	181	190	208	208	219	207	211	209
Panokset	1876	1827	2452	3201	3478	3809	2909	2997
Panos/kohde	10,36	9,61	11,79	15,39	15,88	18,40	13,79	14,34

Taulukko 10 kertoo pelikohteiden lukumäärän kausittain, yhteenlasketut panokset näihin sekä keskimääräisen panoksen per kohde. Kahdella ensimmäisellä tutkimuksen kohteena olleella pelikaudella niin pelikohteiden määrä, yhteenlasketut panokset sekä keskimääräinen panoskoko ovat jääneet pienemmiksi kuin myöhemmillä pelikausilla. Kuten Taulukko 7 kertoi, kausilla 2011-2012 pelikohteiden odotusarvot ovat olleet matalampia muihin kausiin verrattuna. Tästä johtuen myös Kellyn kaavalla optimoidut panoskoot ovat olleet näillä ensimmäisillä tutkimuksen kohteena olleilla pelikausilla matalampia (Taulukko 10).

Taulukko 11 Pelikassan kehitys

	2011	2012	2013	2014	2015	2016	2017	2018
Pelikassa kauden alussa	1000	1000	1000	1000	1000	1000	1000	1000
Pelikassa kauden lopussa	878,11	762,50	639,67	860,81	928,25	1298,08	427,23	610,81
Pelikassan muutos	-12,2%	-23,8%	-36,0%	-13,9%	-7,2%	+29,8%	-57,3%	-38,9%

Taulukossa 11 on esitetty pelikassan kehitys kunkin tutkimuksen kohteena olleen pelikauden osalta. Kuten aiemmin todettua, tutkimuksessa jokaisen pelikauden alussa luotiin kuvitteellinen 1000 yksikön pelikassa, jonka mukaisesti tehtiin kuvitteellisia panostuksia Kellyn kaavan avulla optimoiden.

Huomionarvoista on, että ainoastaan yhdellä pelikaudella kahdeksasta, rahallinen tuotto olisi ollut positiivinen. Tämä tapahtui kaudella 2016 tuoton jäädessä 29,8% voiton puolelle. Kaikilla muilla seitsemällä pelikaudella tuotto olisi ollut negatiivinen (Taulukko 11). Siitäkin huolimatta, että kauden 2016 lisäksi myös kaudella 2014 palautusprosentti olisi ollut yli 100 (Taulukko 9). Tämä selittyy panoskokojen muutoksilla. Vaikka lyötyjen vetojen palautusprosentti olisi yli 100, se ei itsessään vielä takaa rahallista tuottoa, mikäli panoskoko ei ole jokaisessa pelikohteessa sama.

Kaudella 2017 pelikassasta olisi lyötyjen vetojen jälkeen huvennut yli puolet. Tämä siitä huolimatta, että palautusprosentti kyseisellä kaudella ei ollut läheskään heikoin tutkimuksen kohteena olleista kausista (Taulukko 9). Sen sijaan kaudella 2012 saatiin ylivoimaisesti heikoin palautusprosentti tutkimuksen kohteena olleista kausista. Palautusprosentti kaudella 2012 oli vain 80,8%.

Tästä huolimatta rahallisesti mitattuna kyseinen kausi oli vasta neljänneksi heikoin (Taulukko 11).

6 Pohdinta

Tutkielman kuudennessa luvussa esitellään tutkimukseen, tutkimuskysymyksiin ja tutkimuksen tuloksiin liittyvää pohdintaa. Tutkielman empiirisen tutkimuksen osio sai hyvää tukea kirjallisuuskatsauksesta ja siihen kerätystä lähdeaineistosta aiempia aiheesta tehtyjä tutkimuksia koskien. Tässä luvussa esitellään ensin pohdintaa sekä vastaus tutkielman päätutkimuskysymykseen: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?". Tämän jälkeen seuraavissa alaluvuissa käsitellään vastaukset tutkielman apututkimuskysymyksiin "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?" sekä "Onko urheiluedonlyöntimarkkina tehokas?". Tutkimuskysymyksiin on ensin pyritty hakemaan vastauksia kirjallisuuskatsauksen puitteissa, aiempaa olemassa olevaa tutkimustietoa hyödyntäen. Tämän jälkeen tutkielman empiirisen tutkimuksen osiossa tätä tietoutta on pyritty syventämään ja hakemaan vastauksia tutkimuksessa käytetyn datan pohjalta. Luvun viimeisessä osiossa pohditaan tutkimuksen luotettavuutta, yleistettävyyttä sekä rajoituksia.

6.1 Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?

Tutkimuksen päätutkimuskysymykseen lähdettiin ensin hakemaan vastausta tutkielman kirjallisuuskatsauksen osiossa. Kirjallisuuskatsauksessa saatiin selville, että tietoa keräämällä ja sitä tietokannoista louhimalla kykenee rakentamaan algoritmeja, jotka saavuttavat paremman tuloksen kuin puhdas arvaaminen tai vetojen asettaminen tietyn kaavan mukaan, esimerkiksi ainoastaan kotijoukkueiden voittoon panostamalla. Tieteellisiä todisteita ei kirjallisuuskatsauksessa löydetty siitä, että tiedonlouhintateknologiaa hyödyntämällä voisi varmuudella saavuttaa voittoa pitkällä aikavälillä urheiluedonlyöntimarkkinoilta. Kirjallisuuskatsauksessa ei myöskään löydetty aiempaa tutkimustietoa läpi-

käymällä sellaista tutkimustietoa, että voisi varmuudella sanoa tietystä algoritmista olevan kiistatonta hyötyä urheiluvedonlyönnissä.

Tutkielman empiirisen tutkimuksen osiossa ylläolevia, aiempia tutkimustuloksia lähdettiin testaamaan Accuscorelta saadun todennäköisyysarvioidatan pohjalta. Tutkimustuloksissa kävi ilmi, että Accuscoren todennäköisyysarvioilla ei olisi kyennyt tekemään ylituottoa urheiluvedonlyöntimarkkinoilta. Tutkimustuloksista kävi ilmi, että edellä mainituilla arvioilla vetoa lyömällä ainoastaan yksi kahdeksasta vedonlyöntikaudesta olisi kääntynyt vedonlyöjälle rahallisesti voitolliseksi (Taulukko 11). Tämän lisäksi löytyi myös toinen kausi, jolloin lyötyjen vetojen palautusprosentti käytetyillä todennäköisyysarvioilla oli yli 100, mutta joka olisi ollut kuitenkin rahallisesti mitattuna tappiollinen (Taulukko 9).

Tutkimuksessa käytettiin vedonlyönnin panostuksessa yleisesti hyödynnettävää Kellyn kaavaa, jossa tarkoituksena on maksimoida vedon odotusarvo. Kellyn kaavan mukaan panostettaessa vedonlyöjä ei altistu ylipanostukselle, eikä myöskään altistu tilanteeseen, jossa mahdollinen häviö on epäsuhteessa hänen maksukykyynsä. Tavoitteena on siis optimoida panoskoko.

Panoskoon vaihtelusta johtuen, eräällä tutkimuksen kohteena olleena pelikaudella päädyttiin tilanteeseen, jossa palautusprosentti oli yli 100, mutta rahallinen tuotto olisi jäänyt tappiolliseksi. Muun muassa tästä epäsuhdasta voidaan päätellä Accuscoren todennäköisyysarvioiden eroavan melko reilustikin vedonlyöntimarkkinan konsensuksesta. Panoskoot tuolla kyseisellä kaudella olisivat olleet liian suuria riskiin nähden.

Tutkimuksessa havaittiin myös, että Accuscoren todennäköisyysarviot eroavat paikoin todella paljon markkinasta. Jokaisella pelikaudella löytyi reilusti kohteita, joissa vedon odotusarvo Accuscoren arvion mukaan olisi ollut yli jopa 1,50 (Taulukko 7). Jos vedon odotusarvo olisi näin suuri, tarkoittaisi tämä sitä, että vedonlyöntimarkkinan on oltava erittäin pahasti väärässä. On vaikea kuvitella, että koko vedonlyöntimarkkina on todella paljon väärässä. Todennäköisempää on, että Accuscoren todennäköisyysarvio on väärässä sen sijaan, että koko muu vedonlyöntimarkkina olisi väärässä. Tämä tilanne aiheuttaa ylipanostamista Kellyn kaavan mukaisesti panostaen ja mitä enemmän pelikassa altistuu ylipanostukselle, sitä nopeammin se hupenee.

Täytyy muistaa, että Kellyn kaavasta hyötyminen vaatii todennäköisyysarvioiden täsmällisyyttä. Urheilutapahtumissa todennäköisyysarviot ovat kuitenkin loppujen lopuksi aina vain juuri niitä, arvioita.

Kuten tässä tutkielmassa (ks. alaluku 2.1.6) on tullut ilmi, odotusarvo on todennäköisyysarvion ja kertoimen tulo. Havainnollistettuna esimerkiksi 1,50 odotusarvo tarkoittaisi sitä, että 50% todennäköisyysarviolla tulisi vedonlyöntimarkkinoilta saada kertoimeksi vedolle 3,00. Käytännössä tällaisen pitäisi olla mahdotonta, sillä vedonlyöntimarkkinan havaittiin aiemmin tässä tutkimuksessa olevan suhteellisen tehokas, vaikkakaan ei täysin tehokas. Joka tapauksessa, noin suuri vääristymä ei tämän päivän vedonlyöntimarkkinoilla ole millään tavalla realistinen, vaan johtuu pikemminkin todennäköisyysarvion virheellisyydestä.

Tutkimuksessa havaittiin myös, että etenkin vierasjoukkueiden osalta todennäköisyysarviot olivat välillä turhan ylioptimistisia. Jokaisella tutkimuksen kohteena olleella kahdeksalla kaudella vedonlyönnin kohteeksi olisi päätyntä selvästi enemmän vierasvoittoja, suhteessa kotivoittoihin (Taulukko 6). Tämä kertoo todennäköisesti siitä, että Accuscoren arvioissa kotiedulle annetaan liian pieni painoarvo verrattuna vedonlyöntimarkkinaa. Tutkimuksen kohteina olleilla kahdeksalla pelikaudella ainoastaan yhdellä kaudella kahdeksasta Accuscoren todennäköisyysarvioiden mukaan kotijoukkueina pelanneita ennakkosuosikkeja oli enemmän kuin mitä vedonlyöntimarkkinan mukaan olisi ollut (Taulukko 3). Myös tämä kertoo siitä, että Accuscoren todennäköisyysarvioissa annettiin todennäköisesti liian pieni painoarvo kotiedulle.

Lisäksi tutkimuksessa havaittiin, että jokaisella tutkimuksen kohteena olleella pelikaudella panostettiin enemmän altavastajien voittojen puolesta (Taulukko 8). Tästä voidaan päätellä, että joko vedonlyöntimarkkina aliarvostaa altavastajia tai sitten Accuscoren todennäköisyysarviot yliarvioivat. Todennäköisempää lienee se, että koko muu vedonlyöntimarkkina on oikeassa. Erityisen vaarallisia pelikassalle olivat tilanteet, joissa altavastajajoukkue pelaa vieraskentällä. Tällöin pelikassaan kohdistuu suurta ylipanostusta, kun todennäköisyysarvio ja vedonlyöntimarkkinan konsensus eroavat suurimmin toisistaan ja panoskoko nousee luonnottoman suureksi suhteessa pelikassaan.

Kellyn kaavalla pystyy kyllä optimoimaan panoskoon sopivaksi, mutta ongelmaksi muodostuvat juuri yksittäiset tilanteet, joissa todennäköisyysarvio on ollut pahasti väärässä. Tällöin on ensin ylipanostettu ja kohteen mennessä väärin menetetty "liian" suuri osa pelikassasta. Tämän jälkeen Kellyn kaava leikkaa panoskokoja optimoiden ne pelikassan mukaan, eikä kassa enää helpolla kasva takaisin entiselleen. Joten vaikka kaikki muut osa-alueet voitolliseen vedonlyöntiin olisivatkin kunnossa, ei niistä ole apua ilman täsmällisiä todennäköisyysarvioita. Jo yksi suuri ylipanostus voi aiheuttaa todella ison loven pelikassaan, jonka paikkaaminen on hankalaa, vaikka jatkossa arviot olisivatkin täsmällisiä.

Tutkimuksen päätutkimuskysymyksenä oli: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?". Tämän tutkimuksen ja tutkimuksessa käytetyn datan pohjalta ei voida sanoa, että tiedonlouhintateknologiaa hyödyntämällä olisi ylituottoa saavutettu. 1633 analysoinnin kohteena olleen ottelun tulisi olla kyllin suuri otos tämän johtopäätöksen tekemiseen. Lisäksi rahallisesti voitollisia kausia tutkimuksen kohteena olleiden kahdeksan kauden ajalta olisi ollut ainoastaan yksi.

6.2 Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?

Tutkielman kirjallisuuskatsauksessa selvitettiin perinteisen sijoitustoiminnan eroja uhkapelaamiseen nähden. Kirjallisuuskatsauksessa saatiin selville, että uhkapelin ja sijoitustoiminnan väliset keskeisimmät erot liittyvät panoksen kokoon suhteessa käytettävissä oleviin varoihin. Tutkielmassa havaittiin, että ha-

vainnollistavien määritelmien kautta, että oikealla tavalla harjoitettu urheiluviedonlyönti on lähempänä perinteistä sijoitustoimintaa kuin uhkapelaamista.

Tutkielman empiirisen tutkimuksen osiossa lähdettiin soveltamaan sijoitustoiminnan ominaispiirteitä vedonlyönnin saralla. Tutkimuksessa käytettiin vedonlyönnin panostuksessa hyödynnettävää Kellyn kaavaa, jonka puitteissa koko pelikassan menettäminen on mahdotonta, sillä kaava optimoi jokaisen vedon panoskoon suhteessa käytettävissä oleviin varoihin sekä vedon odotusarvoon. Täten myöskään uhkapelistä tuttua ylipanostamisen riskiä ei tule, sillä oletuksella, että todennäköisyysarviot ovat täsmällisiä.

Kuten kuitenkin edellisessä alaluvussa havaittiin, tutkimusdatan todennäköisyysarvioiden täsmällisyydessä oli puutteita, jonka johdosta myös vedonlyönnin tulos olisi ollut tappiollinen yhtä lukuun ottamatta jokaisella tutkimuksen kohteena olleella pelikaudella. Tässä syyttävä sormi osoittaa kuitenkin todennäköisyysarvioihin, joiden vuoksi myös panoskoko kasvoi välillä turhan suureksi riskiin nähden.

Tutkimuksessa joka tapauksessa havainnollistettiin, kuinka kurinalaisesti jokaiseen kohteeseen kaavan mukaan panostamalla vetoa tulisi oikeaoppisesti lyödä. Lisäksi paikoin heikoistakin palautusprosentteista huolimatta, pelikassa pysyi kuitenkin siedettävällä tasolla minimoiden rahalliset tappiot. Tässä suhteessa Kellyn kaava toimii mainiosti, sillä pelikassan pienentyessä, myös siihen suhteutetut panoskoot pienentyvät. Ongelmaksi tutkimuksessa muodostui vain paikoin reilustikin vedonlyöntimarkkinasta eronneet todennäköisyysarviot, joiden johdosta panokset nousivat joissain kohteissa liian suuriksi tehden ison loven pelikassaan.

Kirjallisuuskatsauksessa lähdekirjallisuuden kautta esiteltyjen määritelmien kautta tutkielman toiseen apututkimuskysymykseen pystytään vastaamaan, että urheiluviedonlyöntiä voidaan harjoittaa sijoitustoimintana. Tutkielman empiirisessä tutkimuksessa tätä kirjallisuuskatsauksessa tehtyä havaintoa havainnollistettiin. Vaikka itse vedonlyönnin tulos ei tutkimuksessa käytettyyn dataan pohjautuen olisikaan jäänyt voitolliseksi, ei se kuitenkaan muuta vastausta tähän apututkimuskysymykseen. Sijoitustoiminnan ja uhkapelaamisen lainalaisuudet ovat selvät ja urheiluviedonlyönnissä pystytään soveltamaan nimenomaan sijoitustoiminnasta tuttuja lainalaisuuksia. Täten urheiluviedonlyöntiä voidaan pitää yhtenä kaupankäynnin muotona, jota on mahdollista harjoittaa sijoitustoimintana, jossa on tarkoitus tehdä voittoa riittävän pitkällä aikavälillä.

6.3 Onko urheiluviedonlyöntimarkkina tehokas?

Viedonlyöntimarkkinan tehokkuudesta on tehty lukuisia aiempia tutkimuksia. Tutkielman kirjallisuuskatsauksessa aiempia tutkimuksia vedonlyöntimarkkinan tehokkuudesta lähdemateriaalina käyttäen havaittiin, että markkina on suhteellisen tehokas, mutta ei kuitenkaan täysin tehokas. Lisäksi tutkimuksissa muistutettiin vedonvälittäjien palautusprosentista, joka toimii ikään kuin vir-

hemarginaalina heidän hyväkseen ja täten vaikeuttaa ylituoton saavuttamista huomattavasti.

Tutkielman empiirisen tutkimuksen osiossa vedonlyöntimarkkinan tehokkuudesta saatiin lisää viitteitä. Kuten tutkimustulokset osoittivat, Accuscoren todennäköisyysarvioilla palautusprosentti ylitti vain kahdella tutkimuksen kohteena olleella pelikaudella kahdeksasta 100% rajan (Taulukko 9). Eli tasapainoksilla kahdella kaudella olisi tehnyt voitollisen tuloksen lyömällä vetoa Accuscoren todennäköisyysarvioiden mukaan.

Tutkimuksen kohteena olleiden kahdeksan kauden aikana pelattiin kaikkiaan 2136 NFL-ottelua. Näistä otteluista yhteensä 1633 ottelua päättyi niin sanotusti pelikelpoisiksi Accuscoren arvion mukaan. Eli sellaisia kohteita, joissa odotusarvo Accuscoren todennäköisyysarvion mukaan oli yli 1, oli kaikkiaan 1633 kappaletta. Kertooko tämä enemmän vedonlyöntimarkkinan tehottomuudesta vai Accuscoren todennäköisyysarvioiden epätarkkuudesta?

Pohdinta-luvun aiempien alalukujen perusteella voidaan todeta Accuscoren todennäköisyysarvioiden epätarkkuuden olevan todennäköisempi syy sille, minkä takia arviot eroavat markkinasta välillä suhteellisen reilustikin. Kuten kirjallisuuskatsauksessa selvitettiin, vedonlyöntimarkkinan katsottiin olevan tehokas, vaikkakaan ei aivan täysin tehokas. Joka tapauksessa on hyvin epätodennäköistä, että vedonlyöntimarkkinoilta löytyisi pelikohteita, joissa kohteen odotusarvo on jopa yli 1,50. Tällaisia kohteita löytyi tutkimuksessa käytetyn datan mukaan kaikkiaan 107 kappaletta, joka on täysin epärealistinen määrä. Vielä kun otetaan huomioon, että NFL:n vedonlyöntimarkkina kerää suurimman osuuden koko maailman vedonlyöntimarkkinasta ja on täten hyvin tehokas, kuten tutkielman luvussa 4.2.1 selvennettiin.

Tutkimuksessa käsitellyistä kahdeksasta pelikaudesta kahdella palautusprosentti oli niukasti yli 100%. Kolmella pelikaudella kahdeksasta palautusprosentti oli niukasti alle 100% ja kolmella kaudella reilusti alle 100%. Mikäli vedonvälittäjän palautusprosenttia ei otettaisi huomioon, vaan kuviteltaisiin vedonvälittäjien tarjoavan kohteita 100% palautuksella, olisi kolme niukasti tappiollista vedonlyöntikautta kääntyneet niukasti voitollisiksi. Tutkimuksessa käytetystä kerroindatasta (Taulukko 9) voi havaita vedonvälittäjien keskimääräisen palautusprosentin vaihdelleen kausittain 95,2-95,8% välillä, kun taas vedonlyöjän saama palautusprosentti olisi ollut viidellä tutkimuksen kohteena olleella kaudella näitä korkeampi.

Tutkielman empiirisen tutkimuksen osion perusteella voidaan sanoa kirjallisuuskatsauksessa havaitun vedonlyöntimarkkinan tehokkuutta koskevan väittämän pitävän paikkansa. Vedonlyöntimarkkinan voidaan katsoa olevan melko tehokas ja vedonvälittäjien palautusprosentti pitää huolen siitä, että muutoin niukasti voitolliset kaudet kääntyvät niukasti tappiollisiksi ja täten vaikeuttaa ylituoton saavuttamista huomattavasti.

6.4 Tutkimuksen luotettavuus, yleistettävyyys ja rajoitukset

Tässä tutkielman kuudennen luvun neljännessä alaluvussa pohditaan ja arvioidaan tutkimuksen luotettavuutta, yleistettävyyttä sekä rajoituksia. Tutkimus toteutettiin kvantitatiivisena eli määrällisenä tutkimuksena. Kvantitatiivista menetelmää käyttävä tutkimus kuvaa ja tulkitsee ilmiöitä mittaumenetelmillä, jotka keräävät numeerisia tutkimusaineistoja, kuten alaluvussa 4.2 Vilppaan (2018) määritelmän avulla kerrottiin. Samassa alaluvussa Vilka (2007) puolestaan määritteli kvantitatiivisen tutkimuksen menetelmäksi, joka antaa yleisen kuvan muuttujien (mitattavat ominaisuudet) välisistä suhteista ja eroista. Thomasin (2003) mukaan kvantitatiivisen tutkimuksen tarkoituksena on tuottaa yleistettävissä olevia tutkimustuloksia. Thomasin (2003) mukaan määrällisessä tutkimuksessa tutkijan rooli on tutkia ja mitata, objektiivisuuteen pyrkien. Vilkan (2007) mukaan tutkimustulos on objektiivinen silloin, kun se on tutkijasta riippumaton.

Tämän tutkielman empiirisen tutkimuksen osiossa käytettiin täysin tutkijasta riippumatonta dataa. Tutkimuksessa käytetty data todennäköisyysarvioista saatiin Accuscorelta ja kerroindata kerättiin Oddsportal.com -sivustoa hyödyntäen. Tutkimuksessa tutkijan rooliksi jäi tutkia ja mitata lukuarvoja sisältävää havaintoaineistoa. Täten tutkimus oli riippumaton tutkijasta itsestään ja tutkimustuloksia voidaan pitää täysin objektiivisena.

Tutkimuksessa käytettiin Accuscorelta hankittua dataa todennäköisyysarvioissa. Koska todennäköisyysarvioidataa kerättiin vain yhdeltä alalla toimivalta yritykseltä, voi tämän tutkimuksen luotettavuutta sekä yleistettävyyttä rajoittavana tekijänä. Kuten edellisissä alaluvuissa on pohdittu, kyseiset Accuscoren todennäköisyysarviot eivät välttämättä olleet täysin tarkkoja. Myös tämä saattaa vaikuttaa tutkimuksen luotettavuuteen, yleistettävyyteen sekä luoda rajoituksia. Tutkielman tarkoituksena oli havainnollistaa ja lisätä ymmärrystä aiheesta, josta aikaisempaa tutkimustietoa ei vielä ole olemassa. Oletuksena oli, että todennäköisyyslaskennan prosessi yrityksestä ja toimijasta riippumatta on suhteellisen samankaltainen. Eri toimijoiden prosessi on varmastikin pääpiirteittäin samankaltainen, mutta laskennassa käytettävät muuttujat ja arvot todennäköisesti vaihtelevat. Täten myös todennäköisyysarviot samoihin urheilutapahtumiin vaihtelevat eri toimijoiden välillä. Tutkimustuloksia ei siis suoraan pystytä yleistämään koskemaan kaikkia toimijoita, mutta tutkimus antaa kuitenkin riittävän yleistettävän kuvan aiheesta, josta aiempaa tutkimusta ei vielä ole olemassa.

7 Yhteenveto ja johtopäätökset

Pro gradu -tutkielman seitsemännessä luvussa esitellään tutkimuksen yhteenveto, johtopäätökset sekä jatkotutkimusaiheet. Lisäksi keskustellaan tutkimuksen merkityksestä. Seitsemännen luvun ensimmäisessä alaluvussa esitellään tutkielman yhteenveto, jossa käydään läpi tutkimuksen kulkua sekä sen sisältöä. Tämän jälkeen käydään läpi vielä tutkimuksen johtopäätökset, joita saatiin tutkielman kirjallisuuskatsauksen sekä empiirisen tutkimuksen osioissa. Lopuksi keskustellaan tutkimuksen merkityksestä sekä esitellään mahdollisia jatkotutkimusaiheita.

7.1 Tutkimuksen yhteenveto

Tämän pro gradu -tutkielman tarkoituksena oli selvittää, onko tiedonlouhintateknologian avulla mahdollista laatia vedonlyöntimarkkinaa täsmällisempiä todennäköisyysarvioita tulevien urheilutapahtumien lopputuloksista ja näitä hyödyntämällä saavuttaa rahallista ylituottoa vedonlyöntiin panostamalla. Tutkimuksen kohteena ollut aihepiiri on sikäli mielenkiintoinen, sillä aikaisempaa tutkimusta kyseisestä aihepiiristä, jossa yhdistetään todennäköisyysarvioiden hyödyntäminen vedonlyöntiin, ei ole tehty.

Tutkielman ensimmäinen luku oli johdanto, jossa kuvattiin lyhyesti tutkimuksen tausta, ongelmat, tutkimuksen tarkoitus, tavoitteet, tutkimusmenetelmä sekä saavutetut tulokset. Johdannossa esiteltiin myös tutkielman päätutkimuskysymys sekä kaksi apututkimuskysymystä.

Itse tutkimus koostui kahdesta osasta: kirjallisuuskatsauksesta sekä empiirisen tutkimuksen osiosta. Tutkielman toinen luku oli kirjallisuuskatsauksen osio. Kirjallisuuskatsauksen tarkoituksena oli luoda perusymmärrys käsiteltävästä aihepiiristä määrittelemällä urheiluvedonlyöntiä käsitteenä, kuvaamalla sen teoriaa sekä avaamalla tutkielmassa käytettäviä vedonlyöntialaan olennaisesti liittyviä käsitteitä. Luvussa käytettiin konkreettisia esimerkkejä helpottamaan urheiluvedonlyöntiin liittyvien lainalaisuuksien ymmärtämistä. Kirjalli-

suuskatsauksessa pyrittiin vastaamaan tutkielman tutkimuskysymyksiin, käyttäen taustatietona aikaisempia urheiluvedonlyöntialasta tehtyjä tutkimuksia. Kirjallisuuskatsaus ja siihen kerätty lähdemateriaali aikaisemmista vedonlyöntialaa käsittelevistä tutkimuksista antoivat hyvän pohjan tutkielman empiirisen tutkimuksen osiolle. Tutkielman kolmannessa luvussa käytiin läpi kirjallisuuskatsauksen yhteenveto sekä johtopäätökset.

Tutkielman neljännessä pääluvussa esiteltiin empiirisen tutkimuksen vaiheet. Ensin käytiin läpi empiirisen tutkimuksen tavoite, tutkimusmenetelmät sekä tutkimuksen toteutus. Tutkimuksen toteutus piti sisällään taustatietojen esittämisen tutkimuksessa käytetyn datan alkuperästä, aineiston keruun vaiheet sekä tutkimusaineiston analysoinnin. Luvussa käytiin läpi empiirisen tutkimuksen eteneminen vaihe vaiheelta.

Tutkimuksen toinen osa oli empiirisen tutkimuksen osio, jossa kirjallisuuskatsauksessa havaittuja tutkimustuloksia lähdettiin tutkimaan empiirisen tutkimuksen puitteissa. Empiirinen tutkimus toteutettiin kvantitatiivisena eli määrällisenä tutkimuksena, joka on täysin riippumaton tutkijasta itsestään. Tutkimus perustui Accuscorelta hankittuun dataan, jota täydennettiin muista lähteistä kerätyllä datalla. Tästä lukuarvoja sisältävästä havaintoaineistosta tehtiin erinäisiä laskelmia, joiden perusteella vastattiin tutkielman tutkimuskysymyksiin.

Pro gradu -tutkielman viidennessä luvussa esiteltiin ja käytiin läpi tutkielman empiirisestä tutkimuksesta havaitut tutkimustulokset. Löydettyjä tutkimustuloksia esitettiin taulukkoina, joita pyrittiin avaamaan sanallisesti. Tutkielman kuudes pääluku sisälsi tutkimustuloksiin liittyvää pohdintaa. Luvussa esiteltiin vastaukset sekä pohdintaa tutkielman tutkimuskysymyksiin. Kuidennan luvun lopussa pohdittiin tutkimuksen luotettavuutta, yleistettävyyttä sekä rajoituksia. Tutkielman seitsemännessä ja viimeisessä pääluvussa käytiin läpi tutkimuksen yhteenveto, johtopäätökset, merkitys sekä jatkotutkimusaiheet.

7.2 Tutkimuksen johtopäätökset

Tutkimuksen johtopäätökset esitellään käymällä läpi tutkimusaineistosta löydettyjä havaintoja sekä vastaamalla näiden avulla kolmeen tutkimuskysymyksen tuoden esille tutkimustuloksista tehdyt johtopäätökset. Tutkimuskysymyksiin pyrittiin ensin hakemaan vastauksia kirjallisuuskatsauksen puitteissa, aiempaa olemassa olevaa tutkimustietoa hyödyntäen. Tämän jälkeen tutkielman empiirisen tutkimuksen osiossa tätä tietoutta syvennettiin ja haettiin vastauksia tutkimuksessa käytetyn datan pohjalta.

Tämän pro gradu -tutkielman tarkoituksena oli selvittää, onko tiedonlouhintateknologian avulla mahdollista laatia vedonlyöntimarkkinaa täsmällisempiä todennäköisyysarvioita tulevien urheilutapahtumien lopputuloksista ja näitä hyödyntämällä saavuttaa rahallista ylituottoa vedonlyöntiin panostamalla. Edellä kuvatusta tutkimusongelmasta johdettiin tutkimuksen päätutkimuskysymys:

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?

Syvällisemmän ymmärryksen hankkimiseksi tutkimuksen kohteena olevasta aihepiiristä sekä tutkimusaineiston analysoinnin helpottamiseksi, tutkimuksessa määritettiin päätutkimuskysymyksen rinnalle kaksi apututkimuskysymystä:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?
- Onko vedonlyöntimarkkina tehokas?

Tutkimuksen kirjallisuuskatsauksen osiossa määriteltiin ensin olennaisimmat aihepiiriin liittyvät käsitteet tutkimuksen näkökulmasta katsottuna. Kirjallisuuskatsauksen tarkoituksena oli luoda perusymmärrys tutkimuksen kohteena olevasta aihepiiristä sekä tarjota hyvä teoriapohja aiheen syvällisempään ymmärtämiseen. Kirjallisuuskatsaus tarjoaa kokonaisvaltaisen kuvan vedonlyöntialan peruskäsitteistä, tiedonlouhintateknologiasta sekä sen hyödyntämiskeinoista urheiluedonlyönnissä.

Tutkielmassa esitettiin tutkimuskysymyksiin pyrittiin ensin vastaamaan kirjallisuuskatsauksen puitteissa, käyttäen taustatietona aikaisempaa urheiluedonlyöntialasta tehtyä tutkimustietoa. Tämän jälkeen empiirisen tutkimuksen osiossa samoihin tutkimuskysymyksiin lähdettiin hakemaan vastauksia havaintoaineiston pohjalta erinäisiä laskelmia suorittaen.

Tutkimuksen tavoitteena oli selvittää vastaus päätutkimuskysymykseen "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa ylituottoa urheiluedonlyönnissä?". Kirjallisuuskatsauksen puitteissa havaittiin ensin, että tiedonlouhintateknologiaa hyödyntämällä on teoriassa mahdollista saavuttaa ylituottoa urheiluedonlyönnissä, mutta tieteellisiä todisteita ei kuitenkaan löydetty sen puolesta, että voisi varmuudella sanoa tiedonlouhintateknologian hyödyntämisen tarjoavan kiistatonta hyötyä urheiluedonlyöntimarkkinoilla. Tutkielman empiirisen tutkimuksen osion perusteella tämä kirjallisuuskatsauksen perustella todettu johtopäätös pysyy samana. Tutkimuksessa käytetystä havaintoaineistosta ei löydetty mitään siihen viittaavaa, että voisi varmuudella vastata tiedonlouhintateknologian tarjoavan ylituottoa urheiluedonlyönnissä.

Tutkielmassa vastattiin myös molempiin aiemmin esitettyihin apututkimuskysymyksiin. Ensin vastattiin apututkimuskysymykseen: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana". Kirjallisuuskatsauksessa apututkimuskysymystä lähestyttiin määrittelemällä perinteisen sijoitustoiminnan ja uhkapelaamisen keskeisiä eroja. Kirjallisuuskatsauksessa esitettyjen määritelmien kautta havaittiin, että oikein harjoitettuna urheiluedonlyönti on lähempänä nimenomaan sijoitustoimintaa kuin uhkapelaamista. Täten urheiluedonlyöntiä voidaan pitää yhtenä kaupankäynnin muotona, jota on mahdollista harjoittaa sijoitustoimintana, jossa tarkoituksena on tehdä voittoa pitkällä aikavälillä. Kuten kirjallisuuskatsauksessa määriteltiin, keskeisimmät uhkapelaamisen ja

sijoitustoiminnan väliset erot liittyvät panosten kokoon suhteessa käytettävissä oleviin varoihin. Empiirinen tutkimus havainnollisti tätä konkreettisesti kurinalaisesti Kellyn kaavan mukaan panostaen. Kaava pyrkii optimoimaan panoskoon jokaiseen vedonlyöntikohteeseen sen odotusarvon maksimoiden. Empiirisen tutkimuksen perusteella havainnot ovat yhteneväisiä kirjallisuuskatsauksen kanssa, jotta myös sen perusteella pystytään pääsemään siihen johtopäätökseen, että urheiluedonlyöntiä voidaan harjoittaa sijoitustoimintana.

Tutkielman toinen apututkimuskysymys oli: ”Onko vedonlyöntimarkkina tehokas?”. Tutkimuskysymystä lähestyttiin ensin kirjallisuuskatsauksessa aikaisempaa tutkimustietoa lähdemateriaalina käyttäen. Kirjallisuuskatsauksen perusteella havaittiin, että vedonlyöntimarkkina ei ole täysin tehokas, tämän perusteella olisi siis teoriassa mahdollista saavuttaa tuottoa vedonlyöntimarkkinoilta. Kirjallisuuskatsauksessa käytiin läpi aiempia, vedonlyöntimarkkinoiden tehokkuutta käsitteleviä tutkimuksia, joista lähes kaikissa tulos oli vastaavanlainen. Moni esitelty tutkimus kykeni tietyllä aikavälillä osoittamaan vedonlyöntimarkkinan lievää tehottomuutta. Tämän tutkielman empiirisen tutkimuksen osion perusteella havaitut tutkimustulokset ovat samansuuntaisia. Vedonlyöntimarkkinan havaittiin olevan suhteellisen tehokas, mutta ei kuitenkaan välttämättä täysin tehokas. Lisäksi havaittiin vedonvälittäjän palautusprosentin vaikeuttavan vedonlyöjän tehtävää. Vedonvälittäjän palautusprosentin voi nähdä ikään kuin korjaavan vedonlyöntimarkkinan lievää tehottomuutta tietyllä aikavälillä. Empiirisessä tutkimuksessa käsiteltyjen kahdeksan pelikauden aikana saavutettiin vedonlyöntimarkkinan keskiarvoa parempi tulos viidellä pelikaudella, mutta vain kahdella silloin, kun otetaan huomioon vedonvälittäjän palautusprosentti. Täten myös empiirisen tutkimuksen perusteella voidaan osoittaa vedonlyöntimarkkinan olevan suhteellisen tehokas, vaikkakaan ei välttämättä täysin tehokas.

7.3 Tutkimuksen merkitys

Tutkimus on aihepiiriltään merkityksellinen ja ajankohtainen siinä mielessä, että datan rooli yhteiskunnassa kasvaa jatkuvasti. Dataa kyetään keräämään ja käyttämään hyödyksi lähes alalla kuin alalla, myös urheiluedonlyönnissä.

Tutkimuksen tarkoituksena oli havainnollistaa sekä lisätä ymmärrystä aiheesta, josta aikaisempaa tutkimustietoa ei vielä ole olemassa. Tutkimus tarjoaa syvällisempää ymmärrystä tiedonlouhintateknologian hyödyntämisestä urheiluedonlyönnissä, alaan liittyvistä keskeisistä käsitteistä sekä toimintamalleista. Tutkielmassa avattiin todennäköisyyslaskennan prosessia, todennäköisyysarvioita sekä niiden käyttämistä urheiluedonlyönnin tarpeisiin.

Tiedeyhteisön näkökulmasta tutkimus antaa uutta tietoa siitä, millaisia tuloksia tiedonlouhintateknologiaa hyödyntäen voidaan urheiluedonlyönnissä saavuttaa. Tutkimus avaa kokonaiskuvan tiedonlouhintateknologian hyödyntämisestä urheiluedonlyönnissä tavalla, jollaista ei aikaisemmin ole tehty. Tutkimuksessa on koottu yhteen aikaisempia tutkimuksia samasta aihepiiristä,

avattu todennäköisyysarvioiden laatimisen prosessia sekä tutkittu kuinka nämä soveltuvat vedonlyönnin tarpeisiin.

Vedonlyöntimarkkinan tehokkuudesta on tehty useita eri tutkimuksia, kuin myös tiedonloughintateknologiaa hyödyntävien simulaatiomallien rakentamisesta. Aikaisempaa tutkimusta näiden yhdistämisestä ja tiedonloughintateknologian hyödyntämisestä nimenomaan urheiluedonlyönnin saralla ei kuitenkaan ole tehty. Tutkimus antaa kaiken kattavan läpileikkauksen aihepiiristä sekä tarjoaa aiheita jatkotutkimukselle.

7.4 Jatkotutkimusaiheet

Aikaisempaa tutkimustietoa tiedonloughintateknologian hyödyntämisestä urheiluedonlyönnissä ei ennen tätä tutkimusta ole tehty. Tässä tutkimuksessa keskityttiin siihen, olisiko tiedonloughintateknologiaa hyödyntämällä kyennyt saavuttamaan ylituottoa urheiluedonlyönnissä. Tutkimuksessa annettiin yleiskuva prosessista, olettaen todennäköisyyslaskennan prosessin olevan suhteellisen yhteneväinen toimijasta riippumatta.

Tutkimuksessa ei kuitenkaan paneuduttu syvemmin todennäköisyyslaskelmien prosessiin, eikä siinä käytettyihin muuttujiin tai muuttujien vaikutukseen todennäköisyysarvioissa. Tutkimuksessa havaittiin, että tutkimuksessa käytetty data erosi paikoin melko merkittävästikin urheiluedonlyöntimarkkinan arvioista. Syihin siitä, miksi data erosi niinkin paljon vedonlyöntimarkkinan konsensuksesta, ei kuitenkaan syvennytty tämän tutkimuksen puitteissa.

Mahdollisia jatkotutkimusaiheita voisikin olla eri simulaatiomallien luominen nimenomaan urheiluedonlyönnin näkökulmasta. Näiden tavoitteena olisi tarjota syitä ja selityksiä erinäisten muuttujien vaikutuksesta todennäköisyysarvioihin ja sitä kautta saavutettaviin tuloksiin urheiluedonlyönnissä.

Tutkimustiedon luominen siitä, mitkä muuttujat vaikuttavat urheilutapahtumien lopputuloksiin ja minkälaisella painoarvolla, olisi arvokasta. Tältä pohjalta voisi rakentaa erilaisia urheilutapahtumien todennäköisyysarvioita laskevia simulaatiomalleja, joita voisi käyttää edelleen hyödyksi urheiluedonlyönnin tarpeisiin.

LÄHTEET

- Alaharjula, A. (2005). *Tiedonlouhinta vakuutusaineistosta asiakkuudenhallinnassa*. Joensuun yliopisto, tietojenkäsittelytiede, pro gradu, Joensuu.
- Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. *Journal of political Economy*, 85(4), 803-815.
- Andersen, S., Fountain, J., Harrison, G. W., & Rutström, E. E. (2014). Estimating subjective probabilities. *Journal of Risk and Uncertainty*, 48(3), 207-229.
- Anderson, S. (2019). *NFL Betting Market Efficiency: Finding a Profitable Betting Strategy*.
- Asch, P., Malkiel, B. G., & Quandt, R. E. (1984). Market efficiency in racetrack betting. *Journal of Business*, 165-175.
- Asch, P., Malkiel, B. G., & Quandt, R. E. (1986). Market efficiency in racetrack betting: Further evidence and a correction. *The Journal of Business*, 59(1), 157-160.
- Avery, C., & Chevalier, J. (1999). Identifying Investor Sentiment from Price Paths: The Case of Football Betting (Digest Summary). *Journal of Business*, 72(4), 493-520.
- Bonate, P. L. (2001). A brief introduction to Monte Carlo simulation. *Clinical pharmacokinetics*, 40(1), 15-22.
- Bose, I., & Mahapatra, R. K. (2001). Business data mining – a machine learning perspective. *Information & management*, 39(3), 211-225.
- Bradley, D., & Leitgeb, H. (2006). When betting odds and credences come apart: More worries for Dutch book arguments. *Analysis*, 66(290), 119-127.
- Buchdahl, J. (2003). *Fixed odds sports betting: Statistical forecasting and risk management*. Summersdale Publishers LTD-ROW.
- Buursma, D. (2010). Predicting sports events from past results. In *14th Twente Student Conference on IT*.
- Cao, C. (2012). Sports data mining technology used in basketball outcome prediction.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.
- Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy*, 87(1), 75-88.
- Flepp, R., Nüesch, S., & Franck, E. (2016). Does bettor sentiment affect bookmaker pricing?. *Journal of Sports Economics*, 17(1), 3-11.
- Galavotti, M. C. (2001). Subjectivism, objectivism and objectivity in Bruno de Finetti's Bayesianism. In *Foundations of Bayesianism* (pp. 161-174). Springer Netherlands.
- Golec, J., & Tamarkin, M. (1991). The degree of inefficiency in the football betting market: Statistical tests. *Journal of Financial Economics*, 30(2), 311-323.
- Gray, P. K., & Gray, S. F. (1997). Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance*, 52(4), 1725-1737.

- Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal*, 2(5), 7-12.
- Hájek, A. (2008). Dutch book arguments. *The Oxford handbook of rational and social choice*, 173-195.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.
- Harrison, R. L. (2010, January). Introduction to monte carlo simulation. In *AIP conference proceedings* (Vol. 1204, No. 1, pp. 17-21). AIP.
- Helenius, A. (2016). Rahamonopoli-Tulevaisuuden tie vai menneisyyden jääne.
- Hvattum, L. M. (2013). Analyzing information efficiency in the betting market for association football league winners. *The Journal of Prediction Markets*, 7(2), 55-70.
- Kahn, J. (2003). Neural network prediction of NFL football games. *World Wide Web electronic publication*, 9-15.
- Karvinen, N. (2015). Sattumaa satumaassa: todennäköisyyslaskentaa nopanheitosta mittateoriaan.
- Lahdensuo, O. (2015). Markkinatehokkuus vedonlyöntimarkkinoilla.
- Leung, C. K., & Joseph, K. W. (2014). Sports data mining: predicting results for the college football games. *Procedia Computer Science*, 35, 710-719.
- Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets?. *The Economic Journal*, 114(495), 223-246.
- Malaska, P., & Virtanen, I. (2007) Uskottavaan päättelyyn perustuvasta todennäköisyysteoriasta. Tutkijasta johtajaksi; Näkökulmia akateemiseen ja liike-elämän johtamiseen, 33-59. Turun kauppakorkeakoulun julkaisuja, Sarja C:2007
- Malkiel, B. G. (1999). *A random walk down Wall Street: including a life-cycle guide to personal investing*. WW Norton & Company.
- Makropoulou, V., & Markellos, R. N. (2011). Optimal Price Setting In Fixed-Odds Betting Markets Under Information Uncertainty. *Scottish Journal of Political Economy*, 58(4), 519-536.
- Malarić, R., Katić, T., & Sabolić, D. (2008). The market efficiency of the soccer fixed odds internet betting market. *Applied Economics Letters*, 15(3), 171-174.
- Mani, S. (2018). NFL Betting Market Efficiency: A Closer Look at the Final Day of Betting.
- Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance*, 46(2), 469-477.
- McCabe, A., Travathan, J. (2008). Artificial Intelligence in Sports Prediction, IEEE Computer Society Washington, DC, USA, pp. 1194-1197
- Miljković, D., Gajić, L., Kovačević, A., & Konjović, Z. (2010). The use of data mining for basketball matches outcomes prediction. In *Intelligent Systems and Informatics (SISY), 2010 8th International Symposium on* (pp. 309-312). IEEE.
- Pankoff, L. D. (1968). Market efficiency and football betting. *The Journal of Business*, 41(2), 203-214.

- Pickens, C. T. (2006). Of Bookies and Brokers: Are Sports Futures Gambling or Investing and Does It Even Matter. *Geo. Mason L. Rev.*, 14, 227.
- Raychaudhuri, S. (2008, December). Introduction to monte carlo simulation. In *2008 Winter simulation conference* (pp. 91-100). IEEE.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, 571-587.
- Sebenius, J. K., & Geanakoplos, J. (1983). Don't bet on it: Contingent agreements with asymmetric information. *Journal of the American Statistical Association*, 78(382), 424-426.
- Silvennoinen, M. (2012). *Uhkapelaaminen Internetissä*.
- Snyder, W. W. (1978). Horse racing: Testing the efficient markets model. *The Journal of finance*, 33(4), 1109-1118.
- Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*, 26(3), 482-488.
- Thomas, R. M. (2003). *Blending qualitative and quantitative research methods in theses and dissertations*. Corwin Press.
- Thorp, E. O. (2006). The Kelly criterion in blackjack, sports betting and the stock market. *Handbook of asset and liability management*, 1, 385-428.
- Thorp, E. O. (2010). Understanding the Kelly criterion. *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific Press, Singapore.
- Tiitu, T. (2016). Abnormal returns in an efficient market? Statistical and economic weak form efficiency of online sports betting in European soccer.
- Trawinski, K. (2010). A fuzzy classification system for prediction of the results of the basketball games. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on* (pp. 1-7). IEEE.
- Turunen, A. (2012). Data-analyysin monet mahdollisuudet. *Tieteessä tapahtuu*, 30(2).
- Vilkka, H. (2007). *Tutki ja mittaa: määrällisen tutkimuksen perusteet*.
- Vilpas, P. (2018). Kvantitatiivinen tutkimus. *Metropolia. Pdf-tiedosto. Saatavissa: <https://users.metropolia.fi/~pervil/koantsu/Moniste.pdf>. Viitattu, 27, 2018.*
- Virtanen, M., & Vänni, M. (2005). SUOMALAISTEN VEDONLYÖJIEN KÄYTTÄYTYMINEN-SIJOITTAJIA VAI PELUREITA?.
- Vuoksenmaa, J. (2016). Pelaajalta pelaajille. Unibet Group.
- Woodland, L. M., & Woodland, B. M. (1994). Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *The Journal of Finance*, 49(1), 269-279.
- Xu, J. S. (2011). *ONLINE SPORTS GAMBLING: A LOOK INTO THE EFFICIENCY OF BOOKMAKERS' ODDS AS FORECASTS IN THE CASE OF ENGLISH PREMIER LEAGUE* (Doctoral dissertation, University of California, Berkeley).
- Zdravevski, E., & Kulakov, A. (2010). System for Prediction of the Winner in a Sports Game. In *ICT Innovations 2009* (pp. 55-63). Springer Berlin Heidelberg.

- Accuscore 12.1.2020
Luettavissa: <http://www.accuscore.fi/tietoja.html>
- Arpajaislaki 23.11.2001/1047.
Luettavissa: <http://www.finlex.fi/fi/laki/ajantasa/2001/20011047>
- Statista (2015)
Luettavissa: <https://www.statista.com/statistics/270728/market-volume-of-online-gaming-worldwide/>
- Verohallinto 23.3.2017
Luettavissa: <https://www.vero.fi/fi-FI/Henkiloasiakkaat/Sijoitukset>

