

Patrik Saarelainen

**TIEDONLOUHINTATEKNOLOGIAN  
HYÖDYNTÄMINEN URHEILUVEDONLYÖNNISSÄ**



JYVÄSKYLÄN YLIOPISTO  
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA  
2017

## TIIVISTELMÄ

Saarelainen, Patrik

Tiedonlouhintateknologian hyödyntäminen urheiluedonlyönnissä

Jyväskylä: Jyväskylän yliopisto, 2017, 36 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja: Kyppö, Jorma

Tutkimus on systemaattisena kirjallisuuskatsauksena toteutettu kandidaatin tutkielma, jossa esitetyt määritelmät sekä löydetyt tulokset on saatu käyttäen lähdekirjallisuutena aiheesta aiemmin suoritetuista tutkimuksista. Tutkielman tarkoituksena on selvittää onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuutta saavuttaa tuottoja urheiluedonlyönnissä. Tutkielmassa esitellään vedonlyönnin teoriaa sekä määritellään alaan olennaisesti liittyvät käsitteet. Tämä tutkielma vastaa kaikkiaan kolmeen tutkimuskysymykseen: 1) Voidaanko vedonlyöntiä harjoittaa sijoitustoimintana? 2) Ovatko vedonlyöntimarkkinat tehokkaat? ja 3) Onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuus saavuttaa tuottoja urheiluedonlyönnissä? Sisällössä havaitaan urheiluedonlyönnin olevan lähempänä perinteistä sijoitustoimintaa kuin uhkapeliä. Lisäksi selvitettiin vedonlyöntimarkkinoiden olevan tehokkaat, mutta ei kuitenkaan täysin tehokkaat. Tutkimustuloksina todetaan, että käsitellyt 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. Joka tapauksessa teoreettinen mahdollisuus tuottoihin urheiluedonlyönnissä tiedonlouhintateknologiaa hyödyntämällä on olemassa.

Asiasanat: urheiluedonlyönti, vedonlyöntimarkkinat, tiedonlouhinta, tiedonlouhintateknologia

## ABSTRACT

Saarelainen, Patrik

Utilization of Data Mining Technologies in Sports Betting

Jyväskylä: University of Jyväskylä, 2017, 36 p.

Information Systems, Bachelor's Thesis

Supervisor: Kyppö, Jorma

This research paper is a Bachelor's Thesis which has made as a systematic literature review. The aim of this paper is to find out is it possible to earn some profit in sports betting by utilization of data mining. This paper presents theory of sports betting and defines the concepts which are related very essentially to sports betting. This paper answers to two research questions: 1) Can sports betting be conducted like an investing? 2) Are sports betting market efficient? 3) Is it possible to earn some profit in sports betting by utilization of data mining? The content consist the sports betting to be more investing than gambling. In addition this paper finds out that sports betting market is efficient but not absolutely efficient. Findings of this paper indicate that any algorithm achieved that kind of results it could be sure to earn profit by utilization of data mining technologies in sports betting. But the researches have indicated that it is possible to create an algorithm by data collection and data mining which achieve better results than just a guessing for example. Anyhow there is a theoretical chance at least to earn profit in sports betting by utilization of data mining technologies.

Keywords: sports betting, betting market, data mining, data mining technologies

## KUVIOT

KUVIO 1 Tiedonlouhintaprosessi .....	24
KUVIO 2 Tilastodatan kerääminen ja varastointi.....	24
KUVIO 3 Tietokannan tähtirakenne .....	25
KUVIO 4 Kaava osumatarkkuusprosentille .....	28

## TAULUKOT

TAULUKKO 1 Algoritmin tuottama datasetti .....	26
TAULUKKO 2 Selitykset taulukossa 1 käytetyille lyhenteille.....	27

# SISÄLLYS

TIIVISTELMÄ .....	2
ABSTRACT .....	3
KUVIOT .....	4
TAULUKOT .....	4
SISÄLLYS.....	5
1 JOHDANTO.....	6
2 URHEILUVEDONLYÖNNIN PERUSTEET .....	8
2.1 Mitä on urheiluedonlyönti? .....	8
2.2 Vedonlyöntimarkkina .....	8
2.3 Todennäköisyysarvio .....	9
2.4 Kerroin.....	10
2.5 Palautusprosentti .....	11
2.6 Odotusarvo .....	11
3 URHEILUVEDONLYÖNTI SJOITUSMUOTONA.....	13
3.1 Uhkapeli .....	13
3.2 Objekttiivinen todennäköisyys .....	14
3.3 Subjekttiivinen todennäköisyys.....	15
3.4 Panostaminen .....	16
3.5 Urheiluedonlyönnin suhde sijoittamiseen.....	17
3.6 Yhteenveto .....	18
4 TIEDONLOUHINTATEKNOLOGIA URHEILUVEDONLYÖNNISSÄ ....	20
4.1 Vedonlyöntimarkkinan tehokkuus .....	20
4.2 Tiedonlouhinta yleisesti.....	23
4.3 Tiedonlouhinta urheiluedonlyönnissä .....	23
4.4 Tiedonlouhintateknologiaan perustuva urheiluedonlyönti .....	27
5 YHTEENVETO .....	31
LÄHTEET .....	34

# 1 JOHDANTO

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

Urheiluedonlyönti on näennäisen helppoa - muutaman urheilutapahtuman lopputuloksen oikein ennustamista. Tehtävä on kuitenkin todellisuudessa vaikea. (Vuoksenmaa, 2016, 64.) Urheilutapahtumien lopputulosten ennustaminen on nykyaikana suosittua urheilunseuraajien kesken ympäri maailman, osittain vedonlyönnistä johtuen. Tänä päivänä urheilu tuottaa huomattavia määriä tilastollista informaatiota muun muassa pelaajista, joukkueista, otteluista sekä sarjakaudesta (Haghighat, Rastegari & Nourafza, 2013). Tutkielman tarkoituksena on selvittää, kuinka tätä tilastollista dataa voidaan hyödyntää urheiluedonlyönnissä.

Tutkielma toteutetaan systemaattisena kirjallisuuskatsauksena, jossa esitetyt määritelmät sekä havaitut tulokset perustuvat lähdekirjallisuuteen. Tutkimuksessa vastataan päätutkimuskysymykseen sekä tätä tukeviin kahteen apututkimuskysymykseen.

Päätutkimuskysymys tutkielmassa on:

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuutta saavuttaa tuottoja urheiluedonlyönnissä?

Apututkimuskysymykset ovat seuraavanlaiset:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?
- Ovatko vedonlyöntimarkkinat tehokkaat?

Tutkielmassa selviää, että tiedonloughintateknologiaa hyödyntämällä on mahdollista saavuttaa tuottoa urheiluedonlyönnissä, mutta koska vedonlyöntimarkkina on tieteellisten tutkimusten myötä osoitettu melko tehokkaaksi, minkäänlaisia takuita tuotosta ei ole. Useiden tutkimusten mukaan tiedonloughintateknologiaa hyödyntävät simulaatiomallit ovat pystyneet arvioimaan urheilutapahtumien lopputuloksia kohtuullisen hyvällä tarkkuudella, mutta näyttöä siitä olisiko kyseisillä malleilla saavutettu tuottoa vedonlyöntimarkkinoilla, ei ole.

Tutkielma sisältää tiivistelmän ja johdannon lisäksi kolme sisältölukua sekä yhteenvedon. Toisessa luvussa kuvataan urheiluedonlyönnin teoriaa sekä avataan vedonlyöntialaan olennaisesti liittyviä käsitteitä. Käsitteiden määrittelyn sekä konkreettisten esimerkkien avulla on pyritty avaamaan urheiluedonlyönnin lainalaisuuksia helpommin ymmärrettäviksi.

Kolmannessa luvussa käsitellään urheiluedonlyönnin ja sijoitustoiminnan yhtäläisyyksiä pyrkien erottamaan uhkapeli ja rationaalisesti harjoitettu urheiluedonlyönti toisistaan. Luvun lopussa vastataan tutkielman ensimmäiseen apututkimuskysymykseen: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?"

Tutkielman neljäs luku on varsinainen tutkimusluku, jossa vastataan tutkielman päätutkimuskysymykseen. Neljännen luvun alussa vastataan ensin toiseen apututkimuskysymykseen, jossa selvitetään ovatko vedonlyöntimarkkinat tehokkaat. Tämän jälkeen esitellään tiedonloughintateknologiaa sekä sen käyttöä urheiluedonlyönnin saralla. Neljännen luvun viimeisessä alaluvussa vastataan tutkielman päätutkimuskysymykseen pyrkien selvittämään onko tiedonloughintateknologiaa hyödyntämällä mahdollisuus saavuttaa tuottoja urheiluedonlyönnissä.

Viides luku on johdanto, jossa käydään läpi tutkielmassa käsitellyjä asioita luku kerrallaan sekä esitellään tutkielmassa havaitut tulokset perusteluineen. Lisäksi yhteenvetoluvussa esitellään myös mahdollisia aiheita jatkotutkimukselle.

## 2 URHEILUVEDONLYÖNNIN PERUSTEET

Tutkielman toisessa luvussa määritellään urheiluvedonlyöntiä käsitteenä, kuvataan sen teoriaa sekä avataan muita tutkielmassa käytettäviä vedonlyöntialaan olennaisesti liittyviä käsitteitä. Luvussa käytetään konkreettisia esimerkkejä helpottamaan urheiluvedonlyönnin lainalaisuuksien ymmärtämistä.

### 2.1 Mitä on urheiluvedonlyönti?

Kun puhutaan urheiluvedonlyönnistä, 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 urheiluvedonlyönti 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 voittokertoimella.

### 2.2 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 kilpailukykyänsä 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. Makropouloun ja Markellosen mukaan hyvin informoitujen vedonlyöjien eli järkivedonlyöjien ja viihdevedonlyöjien välinen ero syntyy hyvien 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).

## 2.3 Todennäköisyysarvio

Todennäköisyystarkastelussa pelitilanne formalisoidaan kertailmiöiden subjektiivisen todennäköisyyden lähestymistavan eli ns. Dutch Book-säännöstön 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 sovittun 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 lyövän 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 kerran kolmesta eli noin 33% todennäköisyydellä, hän panostaa kohteeseen vain silloin, kun saa kertoimeksi vähintään 3,00, ja niin edelleen.

## 2.4 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 2,00. (Annala, 2014.)

Kertoimella tarkoitetaan vedonvälittäjän tarjoamaa lukua, jonka perusteella vedonlyöjä tekee oman pelipäätöksensä. Jos vedonlyöjä esimerkiksi saa kohteelle kertoimeksi 2,20, 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 8.3.2017 (Veikkaus, 2017). Veikkaus tarjoaa jalkapallootteluun Borussia Dortmund - Benfica kertoimia 1,35 - 4,80 - 7,80. 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 = Borussia Dortmund (1/1,35) -> 0,74
- X = Tasapeli (1/4,80) -> 0,21
- 2 = Benfica (1/7,80) -> 0,13

Mikäli vedonlyöjän laatima todennäköisyysarvio ylittää yllämainitussa esimerkissä käyttämäni 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.

## 2.5 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 laatimaani esimerkkiä käyttäen, Veikkauksen teoreettinen tuotto-odottama Borussia Dortmundin ja Benfican väliseen jalkapallo-otteluun saadaan selville laskemalla ottelun eri lopputuloksille asetettujen kertoimien käänteislukujen summa. Veikkauksen teoreettinen palautusprosentti kyseisessä ottelussa on  $0,74 + 0,21 + 0,13 = 1,08$ . Tässä tapauksessa Veikkaus tekee kyseisessä pelikohteessa siis keskimäärin kahdeksan prosentin tuoton.

Vedonvälittäjän palautusprosentista voidaan johtaa myös vedonvälittäjän tarjoama palautusprosentti, joka kertoo, miten paljon se keskimäärin maksaa voittoja suhteessa asetettuihin panoksiin. Toisin sanoen se kertoo, miten paljon vedonlyöjät keskimäärin saavat voittona takaisin pelaamistaan panoksista. Vedonvälittäjän tarjoama palautusprosentti on vedonvälittäjän teoreettisen palautusprosentin käänteisluku. (Lahdensuo, 2015.) Käyttäen edelleen edellistä esimerkkiä, Veikkauksen tarjoama palautusprosentti esimerkikohteessa on  $1/1,08 = 0,93$ .

Mikäli oletetaan, että Veikkauksen laatima todennäköisyysarvio kyseisessä pelikohteessa on täysin oikea, vedonlyöjät saavat keskimäärin voittona takaisin pelaamistaan panoksista noin 93 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."

## 2.6 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 vedonlyöjä 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). Esimerkissäni toisena osapuolena toimii vedonvälittäjä ja toisena osapuolena puolestaan vedonlyöjä.

Käytän edelleen esimerkkinä Veikkauksen tarjoamaa jalkapallo-ottelua Borussia Dortmund - Benfica kertoimilla 1,35 - 4,80 - 7,80. Mikäli vedonlyöjä on arvioinut Dortmundin voittavan ottelun 70% todennäköisyydellä ja päättäisi pelata joukkueen voittoa Veikkauksen tarjoamilla kertoimilla, saisi hän vedolleen odotusarvon ( $0,7 \times 1,35$ ) 0,945. Tällöin, odotusarvon ollessa alle 1, hän tulisi panostaneeksi alikertoimeen, eikä vedon lyöminen olisi järkevää. Mikäli vedonlyöjä sen sijaan olisi arvioinut Dortmundin voiton todennäköisyydeksi 80%, hän saisi vedolleen odotusarvon ( $0,8 \times 1,35$ ) 1,08. Tässä tapauksessa kyseessä olisi ylikerroin, johon rationaalisen vedonlyöjän kannattaisi sijoittaa rahaa.

### 3 URHEILUVEDONLYÖNTI SJOITUSMUOTONA

Tutkielman kolmannessa luvussa 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 uhkapeli. 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 tutkielman ensimmäiseen apitutkimuskysymykseen: ”Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?”

#### 3.1 Uhkapeli

Helenius (2016) kirjoittaa, että Suomen arpajaislaki (23.11.2001/1047,2 §) määrittelee uhkapeli-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 kokemattomampiin pelaajiin verrattuna. Urheiluedonlyöntiä voi pitää taitopelinä, sillä tilastoja tarkastellen, laskelmia tehden ja korkeita kertoimia etsimällä voi tehdä elannon. (Helenius, 2016.)

Oikeaoppisesti toteutettua urheiluedonlyöntiä voidaankin pitää taitopelinä, ei niinkään uhkapelinä. Myös urheiluedonlyönnissä sattumalla on oma osuutensa, mutta taitava vedonlyöjä pystyy minimoimaan sen merkitystä. Näistä keinoista kirjoitetaan lisää tutkielman myöhemmässä vaiheessa.

### 3.2 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 urheiluedonlyönnissä 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, kerroin tapahtumalle 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 kolikonheitto-esimerkin 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 tappiollinen.

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.

### 3.3 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 subjektiiivisestä tietämyksestä, mikä ei tarvitse olla yhdenmukainen muiden henkilöiden tietämyksen kanssa. He mainitsevat tutkimuksessaan, että subjektiiivisen todennäköisyyden piirissä eriävä subjektiiivinen 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., 2014).

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

Koska yhtä oikeaa keinoa subjektiiivisten todennäköisyyksien arvioimiseen ei ole, on hyvin todennäköistä, että vedonlyöjän ja vedonvälittäjän arviot eroavat toisistaan. Kuten Malaska ja Virtanen esittivät, eriävä subjektiiivinen tietämys ei merkitse epärationaalisuutta, vaan erilaista evidenssiä päättelyssä. Vedonlyöjä ja vedonvälittäjä saattavat esimerkiksi

painottaa samaa informaatiota eri tavalla todennäköisyysarvioissaan. Teorian tasolla on siis täysin mahdollista, että vedonlyöjä onnistuu todennäköisyyksiä vedonvälittäjää tarkemmin arvioimalla saamaan palautusprosentikseen yli 100.

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

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 todennäköisyydeksi yli 50%, on kerroin tällöin suurempi kuin sitä vastaavan todennäköisyyden käänteisluku. 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.

### 3.4 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 voiton maksimoidakseen panostaa 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.

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.

### 3.5 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 (2016) näkee samaan tapaan urheiluedonlyönnin yhtenä kaupankäynnin muotona. Siinä missä osakemarkkinoilla käydään kauppaa pörssikursseilla, urheiluedonlyönnissä kaupankäynnin kohteena ovat todennäköisyysarviot.

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

Pickens (2006) on tutkinut urheiluedonlyönnin ja perinteisen sijoitustoiminnan välisiä eroja halutessaan 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 (2006) mukaan urheiluedonlyönti ja sijoittaminen ovat niin samankaltaisia aktiviteetteja, että rajanveto näiden välillä on käytännössä mahdotonta. Hän perustelee väitettään sillä, että vedonlyönnin sekä sijoittamisen haittapuolet ovat yhteneväiset. Vedonlyönnin kieltämistä ollaan 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 käytettävissä olevat varat. Mikäli sijoittaja kuitenkin toimii rationaalisesti joko urheiluedonlyönnissä tai perinteisessä sijoitustoiminnassa, on kaikkien toimintaan varattujen resurssien menettäminen äärimmäisen epätodennäköistä.

Moderni portfolioteoria opettaa, että hajauttamalla sijoituksiaan, sijoittaja voi aina pienentää sijoitustensa varianssia (Malkiel, 1999). Markkinoiden tehokkuuteen uskovan 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 nähden. Urheiluedonlyönnissä puolestaan ylisuurilla panoksilla pelaaminen suhteessa omiin käytettävissä oleviin resursseihin nähden täyttää uhkapelaamisen kriteerit.

### 3.6 Yhteenveto

Kolmannessa luvussa avattiin muun muassa uhkapelaamisen käsite ja pyrittiin tekemään ero uhkapelaamisen ja urheiluedonlyönnin välille. Luvussa tarkoituksena oli vastata tutkielman ensimmäiseen apututkimuskysymykseen: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?"

Sekä sijoittaminen että urheiluedonlyönti perustuvat aina myös osittain sattumaan. Urheiluedonlyönnissä ei ole koskaan ennalta selvää kuinka tapahtumassa tulee käymään. Hyvin informoidun, rationaalisesti toimivan urheiluedonlyöjän 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 urheiluedonlyöjä kykenee laskemaan todennäköisyyksiä vedonlyöntimarkkinaa paremmin ja panostaa ainoastaan kertoimiin, joissa odotusarvo on yli yhden. Lisäksi urheiluedonlyöjän tulee pitää panoskoko aina optimaalisena riskiin nähden, välttäen täten ylipanostamista. Kuten tämän luvun ensimmäisessä alaluvussa viitattu Buchdahl kirjoitti, on sijoittaminen pitkäkestoista toimintaa. Täten myös urheiluedonlyöntiin sijoittaminen täytyy nähdä pitkäkestoisena toimintana, ei yksittäisinä uhkapeleinä.

Lyhyellä aikavälillä sattuman merkitys on puolestaan suurempi. Tämän vuoksi optimaalinen panoskoko onkin todella olennainen osa urheiluedonlyönnin 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 klaavan tai kruunan esiintymisen todennäköisyys kuitenkin silti ainoastaan 50%. Mikäli vedonlyöjä panostaa kruunaan, klaava saattaa toteutua vaikka 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 urheiluedonlyöntikin. Lyhyellä aikavälillä sen sijaan riskit ovat suurempia, sillä kurssit 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 urheiluedonlyönti on lähempänä perinteistä sijoitustoimintaa kuin uhkapelaamista. Täten urheiluedonlyöntiä voidaan pitää yhtenä kaupankäynnin muotona, jota on mahdollista harjoittaa sijoitustoimintana, jossa on tarkoitus tehdä voittoa riittävän pitkällä aikavälillä.

## 4 TIEDONLOUHINTATEKNOLOGIA URHEILUVEDONLYÖNNISSÄ

Tutkielman neljännessä luvussa tutkitaan ensin vedonlyöntimarkkinoiden tehokkuutta. Ensimmäisessä alaluvussa vastataan tutkielman toiseen apututkimuskysymykseen, joka on: ”Ovatko vedonlyöntimarkkinat tehokkaat?”

Tämän jälkeen selvitetään millä tavoin tiedonlouhintateknologiaa voidaan käyttää hyödyksi urheiluedonlyönnissä. Tiedonlouhintaprosessi on pääpiirteittäin samankaltainen jokaisessa urheilutapahtumien lopputulosten ennustamista simulaatiomallien avulla käsittelevässä tutkimuksessa, joten havainnollistamisen apuna tässä tutkielmassa käytetään Caon (2012) kyseisestä aiheesta tekemää tutkimusta.

Neljännessä alaluvussa vastataan tutkielman päätutkimuskysymykseen: ”Voidaanko tiedonlouhintateknologiaa hyödyntämällä saavuttaa tuottoja urheiluedonlyönnissä?” Vastausta päätutkimuskysymykseen haetaan saaduista tutkimustuloksista, joissa on testattu urheilutapahtumien lopputuloksia ennustavia simulaatiomalleja. Vastaus ja pohdinnat päätutkimuskysymykseen esitetään luvun lopussa.

### 4.1 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 sekä osake- että vedonlyöntimarkkinoilla toimivan 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. Tämä johtuu siitä, että mitään sellaista informaatiota ei olisi enää saatavilla, jota ei olisi jo 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 tulkinnanvaraa.

Useat tutkimukset ovat osoittaneet, että vedonlyöntimarkkinoilta löytyy epäjohtonmukaisuutta, joka avaa tuottomahdollisuuksia. Epäjohtonmukaisuutta 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 urheiluedonlyönnissä. Hän tuli tutkimuksessaan tulokseen, että amerikkalaisen jalkapallon vedonlyöntimarkkina täyttää markkinatehokkuuden heikot ehdot. Pankoff löysi markkinasta epäjohtonmukaisuutta, 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 epäjohtonmukaisuuksia, 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 varmistaakseen 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 nollasummapeliä 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 (1986) kuitenkin antavat edelleen mahdollisuuden hevosvedonlyöntimarkkinoiden tehottomuudelle, vaikkakin suhtautuvat siihen aiempaa skeptisemmin.

Golec ja Tamarkin (1991) puolestaan löysivät tehottomuutta tutkiessaan amerikkalaisen jalkapallon vedonlyöntimarkkinoita. He havaitsivat, että vedonlyöjät aliarvioivat kotiedun merkitystä NFL:ssä. Lisäksi heidän tutkimuksensa osoitti, että vedon asettaminen kunkin ottelun altavastajan 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. He havaitsivat tutkimuksessaan, että vedon asettaminen kotijoukkueena toimivan altavastajan puolesta on palauttanut heidän aineistonsa perusteella panokset takaisin yli neljän prosentin tuotolla. Neljän prosentin tuotto on heidän mukaansa tilastollisesti merkittävä havainto ja riittävä kertomaan markkinoiden tehottomuudesta.

Malarić, Katić 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.

Esimerkiksi Malarić ym. onnistuivat saavuttamaan 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.

## 4.2 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 uudenslaisin 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).

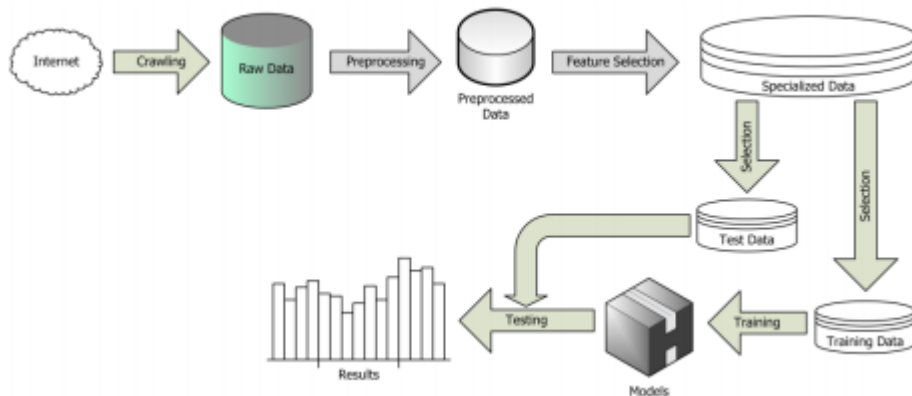
Seuraavissa alaluvuissa selvitetään kuinka tiedonloughintaa voidaan käyttää hyväksi urheiluedonlyönnin saralla ja onko sitä hyödyntämällä mahdollisuus kyetä voitolliseen tulokseen urheiluedonlyöntimarkkinoilla.

## 4.3 Tiedonloughinta 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 jäänyttä, potentiaalisesti hyödyllistä informaatiota. Heidän tutkimuksessaan todetaan, että tiedonlouhinta urheilun tietokannoista auttaa luomaan ennusteita tulevaisuuden otteluista käyttämällä aikaisemmista otteluista kerättyä dataa.



KUVIO 1 Tiedonlouhintaprosessi (Zdravevski & Kulakov, 2010)

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

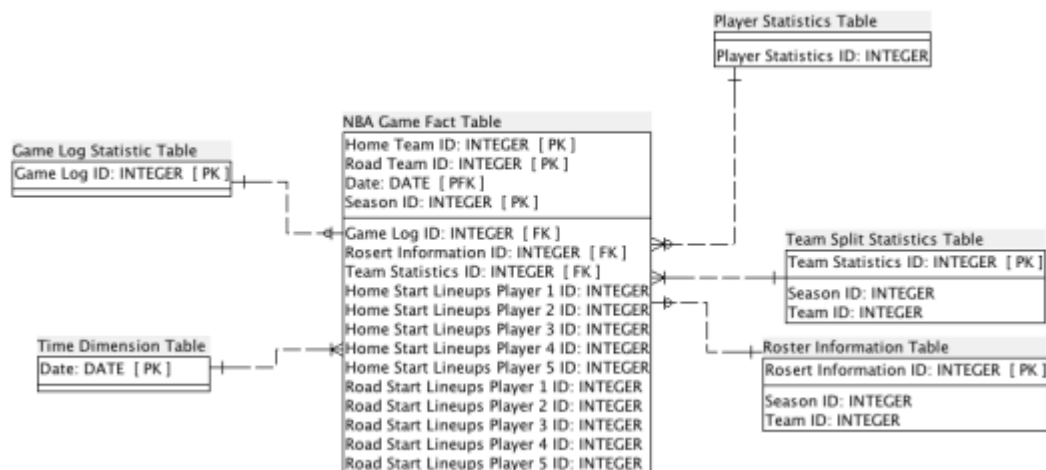
Tietokannoista louhitulla datalla kyetään rakentamaan erilaisia simulaatiomalleja, joilla voidaan ennustaa urheilutapahtumien lopputuloksia. Cao (2012) rakensi simulaatiomallin ennustamaan koripallo-otteluiden lopputuloksia. Mallin tarkoituksena oli hakea edellisten pelien tilastot kahden toisiaan vastaan 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, 2012)



Cao keräsi dataa NBA-aiheisilta (National Basketball Association) internet-sivuilta, jotka julkaisivat tilastodataa kyseisestä koripalloliigasta. Hän keräsi tilastollisen informaatiossa internetistä automaattisin menetelmin hyödyntäen prosessissa Ruby-kieltä. Kun data on kerätty paikalliselle tietokoneelle pääasiallisesti CSV-tiedostomuodossa, Cao latasi 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 (Cao, 2012)

Oheisessa kuviossa on nähtävillä Caon luoma tähtirakenteinen paikallisvarasto. Tietokantarakenteen keskiössä on "Game Fact" -taulu, joka sisältää kaiken oleellisen informaatiossa 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 tiedonlouhinta 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 algoritminsä 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 lisääminen 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.

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

TAULUKKO 1 Algoritmin tuottama datasetti (Miljković, Gajić, Kovačević, Konjović, 2010)

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. Alla olevasta taulukosta on nähtävissä 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

TAULUKKO 2 Selitykset taulukossa 1 käytetyille lyhenteille (Miljković, Gajić, Kovačević & Konjović, 2010)

#### 4.4 Tiedonlouhintateknologiaan perustuva urheiluedonlyönti

Tässä alaluvussa käydään läpi tuloksia tutkimuksista, jotka käsittelevät urheilutapahtumien lopputuloksien arviointia tietokonesimulaatioiden avulla. Näiden tutkimustulosten perusteella vastataan tutkielman päätutkimuskysymykseen: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollista saavuttaa tuottoja urheiluedonlyönnissä?"

Kuten ensimmäisessä luvussa kirjoitettiin, 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 tottakai 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 jalkapallo-otteluiden todennäköisyyksiä.

Tutkimuksissa on käytetty simulaatiomallin toteutunutta tarkkuutta arvioitaessa tarkkuusprosenttia, joka saadaan jakamalla historiallisesta datasta saadut oikein menneet ennusteet jaettuna kaikkien suoritettujen ennusteiden lukumäärällä. Saadun tarkkuusprosentin mukaan eri simulaatiomallien menestystä pystytään vertailemaan toisiinsa.

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

KUVIO 4 Kaava osumatarkkuusprosentille (Miljković, Gajić, Kovačević, Konjović, 2010)

Kahn (2003) tutki amerikkalaisen jalkapallon 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 arvioiden tarkkuuden kannalta on 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. Buursman tutkimusaineistona käyttämistä otteluista 48,66% päättyi kotijoukkueen voittoon, kun taas jokainen hänen testaamansa algoritmi saavutti vähintään 53% osumatarkkuuden. Buursma havainnollisti tutkimuksessa, 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 veikannut 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 tuottaa 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%), Australian rugby-liiga (63,2%), australialaisen jalkapallon AFL (65,1%) sekä rugbyyn 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.

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

Lisäksi on muistettava, ettei vedonlyönnissä vielä pelkästään riitä vaikka pystyisi laskemaan todennäköisyydet täysin tarkasti. Kuten aiemmin tässä tutkielmassa on havainnollistettu; vedonlyönnissä voi jäädä voitolle pitkällä aikavälillä ainoastaan ylikertoimia pelaamalla ja oikeassa suhteessa riskiin nähden panostamalla. Olisikin ollut mielenkiintoista nähdä tutkimus, missä olisi lyöty vetoa jonkin algoritmin laskemien todennäköisyyksien perusteella esimerkiksi yhden kauden ajan.

Tutkielmassa selvitettiin myös aiemmin, että vedonlyöntimarkkinat ovat tehokkaat, mutta eivät kuitenkaan täysin tehokkaat. Oleellista olisikin tutkia, että ovatko vedonlyöntimarkkinat niin tehokkaat ettei tiedonlouhintateknologiaa hyödyntämällä ole mahdollisuutta saavuttaa tuottoa

markkinoilta. Tehtävää vaikeuttaa vielä entisestään vedonvälittäjien palautusprosentti, joka toimii heille ikään kuin tietynlaisena virhemarginaalina. Tämä aiheuttaa sen, että edes tarkinkaan algoritmi ei välttämättä pysty saavuttamaan voittoja vedonlyöntimarkkinoilta 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. Muun muassa Buursma (2010) havainnollisti tutkimuksessaan, kuinka hänen luomansa algoritmi saavutti vedonlyöntimarkkinoilta paremman tuloksen kuin se, että olisi lyönyt jokaiseen otteluun vetoa ainoastaan kotijoukkueiden voiton puolesta. Mutta tieteellisiä todisteita tutkimustulosten muodossa ei kuitenkaan toistaiseksi ole sen puolesta, että tiedonlouhintateknologiaa hyödyntämällä voisi saavuttaa voittoja pitkällä aikavälillä vedonlyöntimarkkinoilta.

## 5 YHTEENVETO

Tässä tutkielmassa tutkittiin tiedonlouhintateknologian hyödyntämistä urheiluedonlyönnissä. Tutkielma toteutettiin systemaattisena kirjallisuuskatsauksena.

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

- Onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuutta saavuttaa tuottoja urheiluedonlyönnissä?

Apututkimuskysymykset olivat seuraavanlaiset:

- Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?
- Ovatko vedonlyöntimarkkinat tehokkaat?

Tutkielman toisessa luvussa määriteltiin urheiluedonlyönti käsitteenä. Lisäksi toisessa luvussa kuvattiin urheiluedonlyönnin teoriaa sekä avattiin tutkielmassa käytettyjä vedonlyöntialaan olennaisesti liittyviä käsitteitä, kuten vedonlyöntimarkkina, todennäköisyysarvio, kerroin, palautusprosentti sekä odotusarvo. Käsitteiden määrittelyn sekä konkreettisten esimerkkien avulla pyrittiin avaamaan urheiluedonlyönnin lainalaisuuksia helpommin ymmärrettäviksi.

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.)

Tutkielman kolmannessa luvussa perehdyttiin urheiluedonlyöntiin sijoitusmuotona. Luvussa vastattiin ensimmäiseen apututkimuskysymykseen, joka oli: "Voidaanko urheiluedonlyöntiä harjoittaa sijoitustoimintana?" Apututkimuskysymystä lähestyttiin määrittelemällä lähdekirjallisuuden avulla uhkapeli, minkä jälkeen määritelmää verrattiin rationaalisesti toteutetun urheiluedonlyönnin lainalaisuuksiin. Tämän jälkeen pohdittiin sijoitustoiminnan ominaispiirteiden soveltuvuutta urheiluedonlyönnin saralle.

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

Tutkielman neljännessä luvussa selvitettiin ensin vedonlyöntimarkkinoiden tehokkuutta ja pyrittiin vastaamaan toiseen apututkimuskysymykseen, joka oli: "Ovatko vedonlyöntimarkkinat tehokkaat?" Apututkimuskysymykseen vastatessa selvitettiin vedonlyöntimarkkinoiden 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 vedonlyöntikertoimissa. Lisäksi Katic ym. (2008) osoittivat tutkimuksessaan, että yksinkertaisella tietokoneohjelmalla pystyttiin hankkimaan tuottoa vedonlyöntimarkkinoilta.

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

Kun tutkielman apututkimuskysymyksiin oli vastattu, alettiin etsimään vastausta päätutkimuskysymykseen, joka oli: "Onko tiedonlouhintateknologiaa hyödyntämällä mahdollisuutta saavuttaa tuottoja urheiluedonlyönnissä?"

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 uudella 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.

Tutkielman neljännen luvun neljännessä alaluvussa esiteltiin havaittuja tutkimustuloksia tutkimuksista, jotka käsitelivät tiedonlouhintateknologiaa hyödyntäen rakenneltuja urheilutapahtumien lopputuloksia ennustavia simulaatiomalleja. Käsitellyt 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. Jatkotutkimuksena olisikin perusteltua tutkia tiedonlouhintateknologian hyödyntämistä nimenomaan urheiluedonlyönnin näkökulmasta, sillä läpikäydyistä aihetta koskevista tutkimuksista puuttui kokonaan vedonlyöntinäkökulma. Jatkotutkimuksena voisikin rakentaa urheilutapahtumien lopputuloksia ennustavan simulaatiomallin ja tutkia olisiko mallia hyödyntämällä kyennyt saavuttamaan tuottoa urheiluedonlyönnissä.

## 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.
- Annala, M. (2014). *Vedonlyönnin optimointityökalun kehittäminen*.
- 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.
- 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.
- Helenius, A. (2016). *Rahamonopoli–Tulevaisuuden tie vai menneisyyden jäännö.*
- Ioulianiou, M., Christofi, I., Talattinis, K., Chalkias, K., & Stephanides, G. (2011). *Automatic Identification of Low-Risk Trading Opportunities in Sport Betting Exchange Markets*. BALCOR 2011, 108.
- 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). *Markkinatohokkuus 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.
- 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.
- Rotando, L. M., & Thorp, E. O. (1992). *The Kelly criterion and the stock market*. *American Mathematical Monthly*, 922-931.
- 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.
- 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).
- 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.

- 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>