



OULUN YLIOPISTO  
UNIVERSITY of OULU

# **Tilastollisten algoritmien soveltuvuus jääkiekkovedonlyöntiin**

Oulun yliopisto  
Tietojenkäsittelytieteen laitos  
Pro gradu -tutkielma  
Juha-Matti Tapani Hurnasti  
7.5.2013

# Sisällysluettelo

Sisällysluettelo.....	2
1. Johdanto.....	3
2. Kirjallisuuskatsaus ja teoria.....	4
2.1 Jääkiekko.....	4
2.2 Urheiluviedonlyönnin ternejä ja käsitteitä.....	6
2.3 Poissonin jakauma ottelutulosten mallintamisessa.....	7
2.4 Bayesilaiset menetelmät.....	11
2.5 Elo-luvut.....	12
2.6 Pi-rating.....	14
2.7 Kotiedusta ja sen merkityksestä.....	16
2.8 Yhteenveto kirjallisuudesta ja teoriasta.....	17
3. Mallin ja algoritmin kehittäminen.....	19
3.1 Vedonlyönnin simulointi.....	19
3.2 Poisson-malli 1.....	20
3.3 Poisson-malli 2.....	22
3.4 Elo-malli 1.....	22
3.4.1 Elo-malli 2.....	23
3.4.2 Painotettu Elo-malli.....	23
3.5 Pi-hockey.....	23
3.6 Aggregaatti-malli.....	24
3.7 Graafinen käyttöliittymä.....	25
3.8 Yhteenveto algoritmin kehitysprosessista.....	26
4. Tulokset.....	27
4.1 Poisson-mallit.....	28
4.2 Elo-malli 1.....	31
4.3 Elo-malli 2.....	34
4.4 Pi-hockey -malli.....	36
4.5 Aggregaatti-malli.....	37
4.6 Yhteenveto tuloksista.....	38
5. Vedonlyöntitulokset.....	40
5.1 Poisson-malli - vedonlyöntitulokset.....	40
5.2 Perus-Elo-malli - vedonlyöntitulokset.....	44
5.3 Maaliperustainen Elo-malli - vedonlyöntitulokset.....	47
5.4 Pi-hockey -malli - vedonlyöntitulokset.....	49
5.5 Aggregaatti-malli - vedonlyöntitulokset.....	49
5.6 Yhteenveto vedonlyöntituloksista.....	52
6. Yhteenveto ja loppusanat.....	53
Lähteet.....	55

# 1. Johdanto

Eri urheilulajien tilastojen tutkiminen ja mallintaminen on aiheuttanut suurta kiinnostusta tilastotieteilijöiden parissa jo jonkin aikaa. Tilastoja ja muuta dataa otteluista on tutkittu eri näkökulmista ja usein myös tulevien otteluiden lopputuloksia on pyritty ennustamaan. (Cattelan et al. 2012). Myös intohimoiset urheilunseuraajat ja fanit ovat aina olleet kiinnostuneita suosikkijoukkueensa seuraavan pelin tuloksien ennustamisesta. (Min et al., 2008). Otteluiden lopputulosten ennustaminen liittyy taas hyvin läheisesti vedonlyöntiin. Esimerkiksi Englannissa, jalkapallo-otteluiden ottelutuloksista vedon lyömisellä on pitkät perinteet. (Dixon & Coles, 1997). Internetin valtavan nopea kasvu on luonut erittäin hedelmällisen kasvualustan myös verkossa tapahtuvalle vedonlyönnille. Uusia verkossa toimivia vedonvälittäjiä on syntynyt erittäin runsaasti. (Constantinou et al., 2012).

Jääkiekko tarjoaa mielenkiintoisen vaihtoehdon perinteiselle jalkapallolle vedonlyöntikohteena, koska kauden aikana pelataan huomattavasti enemmän otteluita ja sarjoissa on usein myös paljon enemmän joukkueita kuin jalkapallosarjoissa. Lisäksi jääkiekko-ottelussa tehdään yleensä myös enemmän maaleja kuin jalkapallossa, mikä tuo oman lisäjännitteensä vedonlyöjille. Lisäksi esimerkiksi Skandinaviassa, toisin kuin vaikkapa Englannissa, jalkapalloa pelataan kesäisin, jolloin jääkiekko tarjoaa talvisen mahdollisuuden vedonlyöntiin.

Tämän tutkimuksen tutkimusongelma on, onko jääkiekko-ottelun lopputulos mahdollista ennustaa etukäteen algoritmisesti historiallisiin tilastotietoihin perustuen ja kuinka menestyksekkäästi. Tässä tutkimuksessa ottelun lopputuloksella tarkoitetaan, päätyikö ottelu kotivoittoon, tasapeliin vai vierasvoittoon. Tutkimuskysymyksiä on kaksi. Ensimmäinen niistä on, kuinka hyvällä onnistumisprosentilla jääkiekko-otteluiden lopputuloksia on algoritmisesti mahdollista ennustaa. Toinen tutkimuskysymys on, voiko algoritmiä hyväksikäyttäen muodostaa voitokkaan vedonlyöntistrategian. Vedonlyöntistrategia ymmärretään voitokkaaksi tässä tutkimuksessa, jos pelikassan koko on kauden lopussa selkeästi isompi kuin se oli kauden alkaessa. Kyseessä on näin ollen konstrukttiivinen tutkimus.

Tämä tutkielma on jäsennetty niin, että aluksi käsitellään jääkiekon sääntöjä ja periaatteita ja sen jälkeen esitellään urheiluvedonlyönnin termistöä ja käsitteitä. Seuraavaksi analysoidaan ja esitellään aiheeseen liittyviä tutkimuksia ja muuta kirjallisuutta. Tässä vaiheessa käydään läpi erilaisia tilastollisia menetelmiä, joita on hyödynnetty otteluiden mallintamisessa ja ottelutulosten ennustamisessa. Sen jälkeen esitellään kehitetyt algoritmit sekä sovellus, joka algoritmeja käyttää sekä sovelluksen graafinen käyttöliittymä.

Tulokset esitellään tämän jälkeen ja se tehdään kahdessa vaiheessa: ensin tulokset analysoidaan pelkästään ottelutulosten ennustamisen näkökulmasta. Tällöin kiinnostuksen kohteena on vain, kuinka algoritmi onnistui ennustamaan otteluiden lopputuloksia kun se hyödynsi ennustamisprosessissa eri tilastollisia malleja. Nämä mallit ovat Poissonin jakauma, Elo-järjestelmä, Elo-järjestelmään perustuva pi-hockey-malli, sekä kaikkia näitä kolmea hyödyntävä aggregaatti-, eli yhdistelmämalli. Tulosten toisessa vaiheessa esitellään tulokset, jotka saatiin kun algoritmin muodostamia todennäköisyysarvioita käytettiin vedonlyönnin simulointiin. Vedonlyönnin

simuloinneissa käytettiin todellisia vedonvälittäjien antamia kertoimia analysoituille otteluille. Tulosten jälkeen seuraa vielä keskustelu, sekä yhteenveto ja loppusanat.

## 2. Kirjallisuuskatsaus ja teoria

Joukkuelajien otteluiden lopputuloksen ennustamiseen löytyy kirjallisuudesta kaksi olennaista tutkimussuuntausta. Ensimmäinen keskittyy ennustamaan koti- sekä vierasjoukkueen maali- tai pistemäärää. Yleisimmin tähän päämäärään pyritään käyttämällä Poissonin jakauman eri variaatioita. Toinen suuntaus pyrkii ennustamaan päättyykö ottelu kotivoittoon, tasapeliin vai vierasvoittoon. Toteutuneiden maalien tai pisteiden määrällä ei tällöin ole merkitystä. Shakista peräisin olevaa Elo-mallia ja sen variaatioita käytetään tässä tutkimuksessa tämän suuntauksen esimerkkinä. Huomattavaa kuitenkin on, että ensin mainitussa koulukunnassakin lopulta pyritään ennustamaan kumpi joukkueista voittaa ottelun, vai päädytäänkö tasapeliin. Tähän pyritään kuitenkin epäsuorasti maali- tai pisteluku-ennusteiden kautta. (Goddard, 2005). Tässä osuudessa käsitellään myös lyhyesti jääkiekkoa ja sen perusteita ja sääntöjä. Lisäksi esitellään urheiluvetonlyönnin perustermistö ja -käsitteet selityksineen niiltä osin, kuin niitä tässä tutkimuksessa hyödynnetään.

### 2.1. Jääkiekko

Jääkiekko on talviurheilujoukkuelaji, jossa ottelun voittaja on enemmän maaleja tehnyt joukkue. Nykyisin viralliset ottelut eivät voi päättyä tasapeliin. Ottelu koostuu kolmesta 20 minuuttia tehokasta peliaikaa kestävästä erästä. Tehokkaalla peliajalla tarkoitetaan, että peliaika ei kulu katkojen aikana toisin kuin esimerkiksi jalkapallossa. Erien välillä pidetään Euroopassa 18 minuutin tauko, ja Pohjois-Amerikassa 15 minuutin tauko. Jos ottelu on kolmen erän jälkeen tasatilanteessa, pelataan viisi minuuttia tehokasta peliaikaa kestävä jatkoaika. Jos jatkoajan aikana jompi kumpi joukkue tekee maalin, ottelu päättyy ja maalin tehnyt joukkue voittaa. Jos jatkoajankin jälkeen ollaan yhä tasatilanteessa, seuraa voittolaukauskilpailu. Voittolaukauskilpailun yksittäisessä voittolaukauksessa kentällä ovat yhtäaikaan vain kaksi pelaajaa: toisen joukkueen kenttäpelaaja ja vastajoukkueen maalivahti. Kenttäpelaaja lähtee tuomarin vihellyksestä kuljettamaan kiekkoa kentän keskipisteestä kohti maalivahdin puolustamaa maalia tarkoituksenaan saada tehtyä maali. Tämän kuljetuksen aikana kenttäpelaaja ei saa pysähtyä eikä kuljettaa kiekkoa taaksepäin. Molemmilla joukkueilla on aluksi kolme voittolaukausta ja näiden maalit lasketaan yhteen. Enemmän maaleja tehnyt joukkue voittaa. Jos kolmen voittolaukauksen jälkeen ollaan kuitenkin edelleen tasatilanteessa, molemmat joukkueet laukovat voittolaukauksia vuorotellen niin kauan, että toinen joukkue epäonnistuu toisen onnistuessa.

Jääkiekko pelinä juontaa juurensa jo 1800-luvun alkupuolen Kanadaan. Jääkiekon ensimmäiset tunnetut säännöt julkaistiin Montreal Gazette -sanomalehdessä vuonna 1877. Jääkiekko esiteltiin olympialajina, mielenkiintoista kyllä, kesäolympialaisissa 1920, jonka jälkeen sen kansainvälinen suosio on kasvanut suuresti. Viralliseksi miesten olympialajiksi jääkiekko vakiintui Chamonix'n talviolympialaisissa 1924. Vuonna 1994, Lillehammerin talviolympialaisissa myös naisten jääkiekko otettiin olympialajiksi. (Albert et al., 2005).

Jääkiekkoa voidaan pelata sekä ulkona että halleissa sisätiloissa. Jääkiekkoa pelataan ovaalin muotoisella jäädytetyllä kentällä jota ympäröivät laidat. Jäätä ja laidoista

muodostuu kaukalo. Kaukalon molemmissa päädyissä on maali. Maalien maaliviiva on molemmissa päädyissä neljä metriä kaukalon takalaidasta. Maaliviiva ulottuu koko kentän poikki kaukalon sivulaidasta toiseen. Itse maali on luonnollisesti maaliviivan takana, tolpat maaliviivalla. Kaukalo on Euroopassa 61 metriä pitkä, Yhdysvalloissa se on muutamia kymmeniä senttejä lyhyempi. Eurooppalainen kaukalo on 30 metriä leveä, mutta Yhdysvalloissa kaukalon leveys on vain 25.9 metriä, mikä muuttaa Yhdysvalloissa pelattavan pelin luonnetta huomattavasti. (Albert et al., 2005).

Normaalipelitilanteessa jääkiekossa on yhtä aikaa kaksitoista pelaajaa jäällä. Molemmilta pelaavilta joukkueilta jäällä on viisi kenttäpelaajaa ja maalivahti. Jääkiekossa sääntörikkomuksista voi saada eri pituisia jäähyjä. Jäähyn saanut pelaaja joutuu tietyksi ajaksi jäähyaitioon ja jäähyn saaneen pelaajan joukkue joutuu pelaamaan tuon ajan vajaalukuisena. Myös useampi saman joukkueen pelaaja voi joutua yhtä aikaa jäähyille, mutta kentällä on aina kuitenkin vähintään kolme kenttäpelaajaa molemmista joukkueista. Maalivahdin, koko joukkueen tai joukkueen valmentajan saadessa jäähyn, joku jäähyyn johtaneen rikkeen sattuessa kentällä ollut pelaaja määrätään kärsimään jäähyä. Tämän määräyksen suorittaa jäähyn saaneen joukkueen valmentaja. Pienestä rikkeestä pelaaja saa kaksi minuuttia kestävän jäähyn. Jos pelaaja tekee saman pelijakson aikana kaksi pientä rikettä, hän voi saada myös 2+2 minuuttia kestävän jäähyn. Vakavammista rikkeistä pelaajalle voidaan tuomita viisi minuuttia kestävä jäähy. Pelaaja voi saada myös henkilökohtaisia rangaistuksia, jotka eivät aiheuta pelaajan joukkueelle alivoimatilannetta. Lyhin henkilökohtainen rangaistus on 10 minuuttia ja pisin 25 minuuttia, joka tarkoittaa, ettei pelaaja saa pelata kyseisissä ottelussa enää ollenkaan. Jäähyjä voidaan myös yhdistellä: esimerkiksi Suomen SM-liigassa päähän kohdistuneesta taklauksesta pelaaja saa 2+10 minuutin jäähyn. Tällöin pelaajan joukkue joutuu pelaamaan 2 minuuttia alivoimalla, ja itse pelaaja joutuu istumaan jäähyaitiossa yhteensä 12 minuuttia. Jos ylivoimalla pelaava joukkue tekee kyseisen jäähyn aikana maalin, pääsee jäähyllä ollut pelaaja pois jäähyiltä enenaikaisesti. Tämä ei koske kuitenkaan yli kahden minuutin jäähyjä. Jääkiekossa kentällä on yhtä aikaa neljä tuomaria: kaksi päätuomaria ja kaksi linjatuomaria. Päätuomarit vastaavat oikeuden jaosta, eli valvovat sääntöjen noudattamista ja antavat sääntörikkomuksista jäähyjä. Linjatuomareiden tehtävänä on pudottaa aloituksissa kiekko ja valvoa paitsioiden syntymistä. Joukkue voi päästä kenttäylivoimalle myös pelaamalla ilman maalivahtia, jolloin maalivahdin tilalle kentälle voidaan ottaa ylimääräinen kenttäpelaaja. Jos kumpikaan joukkue ei pelaa alivoimalla, on kentällä tällaisessa tilanteessa yksitoista kenttäpelaajaa kymmenen sijasta. Maalivahdin poisottamista käytetään maalintekotodennäköisyyden lisäämisekeinona yleensä tappiolla olevan joukkueen toimesta ottelun lopussa. (Albert et al., 2005).

Tässä tutkielmassa tutkimusmateriaalina käytetään Pohjois-Amerikan jääkiekkoliiga NHL:n (National Hockey League) ottelutuloksia ja -tapahtumia. NHL on perustettu vuonna 1917 ja tällä hetkellä siinä pelaa yhteensä 30 joukkuetta Yhdysvalloista ja Kanadasta. Kukin joukkue pelaa runkosarjan aikana yhteensä 82 ottelua, 41 kotiottelua ja 41 vierasottelua. Yhteensä koko sarjassa pelataan siis 1230 ottelua kauden aikana. Ottelun voittaja saa aina kaksi pistettä, jos voitto saavutettiin varsinaisella peliajalla ilman jatkoaikaa. Hävinnyt joukkue saa tällöin nolla pistettä. Jos ottelu ratkeaa vasta jatkoajalla tai voittolaukauskilpailussa, molemmat joukkueet saavat yhden pisteen, ja voittaja joukkue vielä yhden lisäpisteen, eli kaksi pistettä. (Albert et al., 2005).

Joukkueet on jaettu Itäiseen ja Läntiseen konferenssiin, joista kukin on jaettu vielä neljään divisioonaan. Kukin joukkue kohtaa toisen konferenssin joukkueista kolme joukkuetta kaksi kertaa, ja muut toisen konferenssin joukkueet kerran. Kukin joukkue kohtaa toiset oman konferenssin joukkueet neljä kertaa, ja vielä oman divisioonan joukkueet kaksi kertaa. Runkosarjan jälkeen pelataan pudotuspelit, joihin selviytyy kummankin konferenssin saavutettujen pisteiden perusteella kahdeksan parasta

joukkueita. Joukkueet jaetaan tietyn kaavan mukaan ottelupareihin jotka pelaavat ottelusarjan. Ottelusarja pelataan periaatteella paras seitsemästä. Tämä tarkoittaa, että se joukkue joka voittaa ensiksi neljä ottelua sarjassa, pääsee seuraavalle pudotuspeli-kierrokselle. Mestaruuteen, eli Stanley Cupin voittamiseen vaaditaan siis yhteensä 16 voittoa. (Albert et al., 2005).

## 2.2. Urheiluvedonlyönnin termejä ja käsitteitä

Urheiluvedonlyönnissä tyypillisesti lyödään vetoa jonkun tietyn urheilutapahtuman maaliluvuista, tapahtumista tai lopputuloksesta. Lopputuloksella tarkoitetaan, päätyykö ottelu koti- tai vierasvoittoon vai tasapeliin. Tällaisesta esimerkiksi käy tavallinen suomalainen pitkäveto. Maaliluvuilla tarkoitetaan joukkueiden tekemien ja päästämien maalien määrää tietyssä ottelussa. Maaliluvut ottelussa voivat olla esimerkiksi 3-2, joka tarkoittaa, että kotijoukkue teki 3 maalia ja päästi kaksi maalia, ja vierasjoukkue teki kaksi maalia ja päästi kolme maalia. Kotijoukkue siis voitti, maalieron ollessa yksi maali. Tässä tutkimuksessa lyödään vetoa vain ottelun lopputuloksista ja pyritään löytämään näihin kolmeen vaihtoehtoon liittyviä todennäköisyyksiä. On myös huomioitavaa, että jääkiekossa ottelun lopputulokseen lasketaan mukaan vain varsinaisella peliajalla tehdyt maalit. Jatkoajalla tai voittolaukauskilpailussa tehtyjä maaleja ei lasketa. Näin ollen ottelu voi päättyä myös tasapeliin vedonlyönnin näkökulmasta. (Vuoksenmaa et al., 1999).

Urheiluvedonlyönti on vedonlyöjän ja vedonvälittäjän kauppaa todennäköisyyksillä. Vedonvälittäjä antaa tietyn urheilutapahtuman eri lopputuloksille kertoimet, eli todennäköisyysarvot, että ottelu päättyy kyseiseen lopputulokseen. Kerroin tietylle ottelun lopputulokselle merkitään yleensä desimaalilukuna joka on aina vähintään yli 1,00. Käytännössä pienin kerroin on 1,01. Ottelun muiden kertoimien pitää tällöin olla hyvin suuria. Tällaiset kertoimet ovat kuitenkin hyvin harvinaisia. Kerroin muutetaan vedonlyöntitoimiston arvioimaksi todennäköisyydeksi jakamalla luku 1 kertoimella. Todennäköisyys on siis kertoimen käänteisluku. Kertoimen todennäköisyysarviosta saa luonnollisesti jakamalla luvun 1 todennäköisyysarviolla. Näin ollen kerroin 1,01 tarkoittaa, että kyseisen lopputuloksen todennäköisyys on vedonvälittäjän mielestä yli 0,99, eli yli 99%. Ottelun kaikkien kolmen eri lopputuloksen todennäköisyysarvioiden summan tulisi olla 1, mutta vedonvälittäjien tapauksessa tämä ei ole aivan näin. Yleensä, kun vedonvälittäjien todennäköisyysarvot lasketaan yhteen, saadaan summaksi jonkin verran yli yksi. Tämä yhden ylittävä osa on vedonvälittäjän voittomarginaali. Voittomarginaali vaihtelee yleensä välillä 0,04 – 0,10.

Urheiluvedonlyönnissä pelaaja asettaa tietynsuuruisen rahallisen panoksen jollekin ottelun lopputuloksesta riippuen siitä, minkä tuloksen pelaaja kokee itse kaikkein todennäköisimmäksi. Jos pelaajan ennustus osui oikeaan, voittaa hän maksamansa panoksen kertoimella kerrottuna. Esimerkiksi, jos pelaaja lyö vetoa Minnesota Wildin voiton puolesta kertoimella 2,05 panostaen 2,00 euroa ja Minnesota todella voittaa, on pelaaja voittanut 2,05 kertaa 2,00 euroa, eli 4,10 euroa. Jos Minnesota taas ei voitakaan, on pelaaja menettänyt panostuksensa. (Vuoksenmaa et al., 1999).

On mahdollista lyödä vetoa myös useamman ottelun lopputuloksesta yhtä aikaa, jolloin kertoimet kerrotaan toisillaan voittokertoimen saamiseksi. Tällöin luonnollisesti myös voiton todennäköisyys pienenee huomattavasti. Tässä tutkimuksessa vetoa lyödään kuitenkin aina vain yhdestä ottelusta ja yhdestä lopputuloksesta kerrallaan.

Jotta urheiluvedonlyönti voi olla voitollista pitkällä aikavälillä, täytyy vedonlyöjän pystyä tekemään parempia todennäköisyysarvioita kuin vedonvälittäjien. Vetoa

kannattaa lyödä siis vain silloin, kun pelaajan oma todennäköisyysarvio jollekin ottelun lopputulokselle on suurempi kuin vedonvälittäjän arvio. Tällöin on kyse niin sanotusta ylikertoimesta. Ylikerronta voidaan mitata myös odotusarvona, joka saadaan kun kerroin jaetaan pelaajan todennäköisyysarviolla. Jos odotusarvo on yli yksi (1), on kyseessä ylikerroin. Jos vedonlyöjä lyö vetoa alikertoimista, tai esimerkiksi lyö vetoa vain silloin, kun jollekin lopputulokselle on pieni kerroin, voi hän voittaa lyhyellä aikavälillä paljonkin rahaa, mutta pitkällä aikavälillä tällainen pelaaminen ei voi olla voitollista. Puhutaankin arvovedonlyönnistä (engl. value betting). (Vuoksenmaa et al., 1999). Tässä tutkimuksessa lyödään vetoa vain sellaisista otteluista, joiden jollekin lopputulokselle algoritmin arvioima todennäköisyys on suurempi kuin vedonvälittäjän antama. Jos kahdelle lopputulokselle saadaan samassa ottelussa samat todennäköisyysarvot, valitaan pelattava lopputulos järjestyksessä kotivoitto, vierasvoitto, tasapeli. (Vuoksenmaa et al., 1999).

Voitokkaan vedonlyönnin kannalta asetetun panoksen koko on lähes yhtä ratkaisevaa kuin itse todennäköisyysarvot ja kertoimet. Yleisin tapa määrittää panoskoko tiettyyn vetoon on käyttää niin sanottua Kellyn kaavaa. Kellyn kaava (engl. Kelly criterion, Kelly formula) on todennäköisyyslaskennan kaava, jolla on mahdollista määrittää optimaalinen panoskoko suhteessa pelikassan suuruuteen. Se on saanut nimensä amerikkalaisen matemaatikon J. L. Kelly Juniorin mukaan. (Vuoksenmaa et al., 1999). Optimaalinen panoskoko lasketaan Kellyn kaavalla

$$B = (pk - 1) / (k - 1) \quad (1)$$

Kaavassa B on optimaalinen panos esitettynä osuutena pelikassasta, p on pelaajan arvioima todennäköisyys tietylle lopputulokselle ja k on vedonvälittäjän kerroin. Yleensä Kellyn kaavalla saadaan melko suuria optimaalisia panostuksia suhteessa pelikassan kokoon, joten usein tuo suositettu panos jaetaan vielä niin sanotulla Kellyn jakajalla (engl. Kelly divider), joka kuvaa käytännössä pelaajan ottamaa riskiä. Mitä suurempaa Kellyn jakajaa käytetään, sitä pienemmän riskin pelaaja on valmis ottamaan. Yleensä Kellyn jakaja on joku kokonaisluku väliltä 1-10. On myös huomioitava, että vedonvälittäjillä on aina minimipanokset, joilla kohteita voi pelata. Suomalainen Veikkaus Oy tarjoaa minimipanokseksi yhden euron, mutta esimerkiksi englantilaisella William Hillillä minimipanos on vain 0,03 euroa. Tämä asettaa luonnollisesti vaatimuksia myös pelikassan koolle. Esimerkiksi kymmenen euron pelikassalla ei voisi Kellyn kaavaa käyttäen harrastaa arvovedonlyöntiä, jos lyötäisiin vetoa vain Veikkaus Oy:n tarjoamista peleistä. Tässä tutkimuksessa ei kuitenkaan oteta huomioon minimipanoksia. Sijoitettu panos voi näin olla hyvin pieni. (Vuoksenmaa et al., 1999).

### 2.3. Poissonin jakauma ottelutulosten mallintamisessa

Olkoon  $\underline{x}$  satunnaismuuttuja, joka voi saada ei-negatiivisia kokonaislukuarvoja. Jos  $\lambda > 0$  ja muuttujan frekvenssifuntio on

$$P(x) = (e^{-\lambda})(\lambda^x) / x! \quad (2)$$

niin  $\underline{x}$  noudattaa Poissonin jakaumaa parametrilla  $\lambda$ . Satunnaismuuttujan  $\underline{x}$  odotusarvo on tällöin  $\lambda$ . Poissonin jakaumalla voidaan mallintaa jonkin aikavälin kuluessa sattuvien melko harvinaisten tapahtumien lukumäärän todennäköisyyksiä. (Vasama & Vartia, 1971, s. 237-241).

Poissonin jakauman avulla voidaan mallintaa joukkuelajeissa joukkueiden maalimääriä. Oletetaan esimerkiksi, että joukkue A on tehnyt keskimäärin 2,34 maalia ottelussa ja joukkue B 1,95 maalia. Jos joukkueiden pelatessa vastakkain, niiden maalimäärät eivät riipu toisistaan, voidaan olettaa, että joukkueen A maalimäärä noudattaa Poissonin jakaumaa parametrilla 2,34 ja joukkueen B maalimäärä Poissonin jakaumaa parametrilla 1,95. Lasketaan todennäköisyys sille, että ottelu päättyy tulokseen 3-1.

Lasketaan ensin todennäköisyys sille, että joukkue A tekee 3 maalia. Kaavassa  $x$  on se muuttuja, jonka esiintymisen todennäköisyyttä lasketaan, eli annettussa esimerkissä 3.  $\lambda$  on tiettyjen tapahtumien keskiarvo tietyssä ajassa, tässä esimerkissä  $\lambda$  on joukkueen A tekemien maalien keskiarvo 2,34. Kaavassa  $e$  on vakio, niin sanottu Eulerin luku, jonka likiarvo on 2,71828. Näillä arvoilla todennäköisyys sille, että joukkue A tekee 3 maalia on 20,6 %. Samalla kaavalla todennäköisyys sille, että joukkue B tekee yhden maalin on 27,7 %. Tuloksen 3-1 todennäköisyys saadaan kertomalla joukkueiden todennäköisyydet toisillaan. Näin ollen todennäköisyys lopputulokselle 3-1 on 5,7 % ( $0,206 \cdot 0,277$ ). Samalla tavalla voidaan laskea todennäköisyydet muillekin ottelun mahdollisille lopputuloksille. Ne on esitetty alla olevassa taulukossa 1. Taulukon perusteella voidaan päätellä, että kyseisen ottelun todennäköisin lopputulos on 2-1, jonka todennäköisyys on 7,32 %.

Kirjallisuudessa Poissonin jakaumaa ei kuitenkaan ole käytetty edellä esitettyyn tapaan. Seuraavassa käydään läpi monimutkaisempia malleja Poissonin jakaumasta, joita kirjallisuudessa on käytetty.

**Taulukko 1.** Poissonin jakauman esimerkki.

Joukkue A								
Joukkue B	0	1	2	3	4	5	6	7
0	1.37%	3.21%	3.75%	2.93%	1.71%	0.80%	0.31%	0.10%
1	2.65%	6.25%	<b>7.32%</b>	5.71%	3.34%	1.56%	0.61%	0.20%
2	2.61%	6.10%	7.13%	5.56%	3.26%	1.52%	0.59%	0.20%
3	1.69%	3.96%	4.64%	3.62%	2.12%	0.99%	0.39%	0.13%
4	0.83%	1.93%	2.26%	1.76%	0.40%	0.48%	0.19%	0.06%
5	0.32%	0.75%	0.88%	0.69%	0.40%	0.19%	0.07%	0.03%
6	0.11%	0.25%	0.29%	0.22%	0.13%	0.06%	0.02%	0.01
7	0.03%	0.07%	0.08%	0.06%	0.04%	0.02%	0.01%	0.00%

Aiempi tutkimus ((Moroney, 1951) ja (Reep, Pollard & Benjamin, 1971)) oli todennut Poissonin jakauman sopimattomaksi jalkapallo-otteluiden (association football) lopputulosten mallintamiseen ja käyttäneet sen sijaan negatiivista binomijakaumaa, Maher (1982) halusi kuitenkin tutkia sen soveltuvuutta tarkemmin koska negatiivinen binomijakauma ei mahdollista joukkueiden eri ominaisuuksien huomioon ottamista.

Aluksi Maher (1982) käytti riippumatonta Poissonin mallia (independent Poisson distribution) lopputulosten mallintamiseen. Tutkimuksessa oletetaan joukkueiden  $i$  ja  $j$  pelatessa vastakkain, että ottelun maaliluvut ( $x_{ij}$  ja  $y_{ij}$ ) ova toisistaan riippumattomia ja noudattavat Poisson-jakaumia seuraavasti:  $x_{ij} \sim \text{Poisson}(\alpha_i \beta_j)$  ja  $y_{ij} \sim \text{Poisson}(\gamma_i \delta_j)$ . Näistä  $\alpha_i$  edustaa joukkueen  $i$  hyökkäysvoimaa kotiottelussa,  $\beta_j$  joukkueen  $j$  puolustuksen heikkoutta vierasottelussa,  $\gamma_i$  joukkueen  $i$  puolustuksen heikkoutta kotiottelussa, sekä  $\delta_j$  joukkueen  $j$  vierasotteluiden hyökkäysvahvuutta. Lopulta tutkimuksessa huomattiin



kuitenkin, että merkityksellisiä olivat vain  $\alpha_i$ , eli joukkueen  $i$  hyökkäyksen vahvuus kotiotteluissa sekä  $\beta_j$ , eli joukkueen  $j$  puolustuksen heikkous vierasottelussa. Tämä tarkoittaa, että joukkueiden suhteellinen hyökkäysvoima on sama sekä kotona että vieraissa pelattaessa. Sama on totta myös puolustuksen heikkouden suhteen. (Maher, 1982). Hyökkäysvoima ja puolustuksen heikkous joukkueille on tutkimuksessa laskettu maximum likelihood -menetelmää käyttäen.

Tutkimuksessaan Maher (1982) testasi riippumattoman Poissonin mallin soveltuvuutta yhteensä kahteentoista data-aineistoon, jotka oli kerätty Englannin neljän ylimmän sarjatason tilastoista vuosilta 1973-1975. Tutkimuksessa huomattiin, että riippumattoman Poissonin jakauman antamat todennäköisyydet vastaavat melko hyvin datasettien muodostamaa todellisuutta. Riippumatonta Poissonin mallia käytettäessä huomattiin kuitenkin, että se aliarvioi sellaisten otteluiden määrän, jossa tehdään vain yhdestä kahteen maalia, sekä yliarvioi sellaisten otteluiden määrän, joissa tehdään nolla maalia, tai vähintään neljä maalia. Kun tätä tutkittiin vielä syvemmin ja tarkemmin, havaittiin, etteivät mallin antamat todennäköisyydet lopulta vastanneetkaan todellisuutta tarpeeksi hyvin. Tästä johtuen Maher päätyi käyttämään kaksiulotteista Poissonin mallia (bivariate Poisson distribution), joka paransi saatuja tuloksia huomattavasti.

Huomattavaa on, että Maher ei tutkimuksessaan pyrkinyt niinkään ennustamaan tulevien otteluiden tuloksia, vaan pelkästään mallintamaan ja kuvaamaan jo pelattuja otteluita ja niiden tuloksia, ja tutkimaan, ovatko nämä tulokset Poisson-jakautuneita. Dixon & Coles (1997) käyttävätkin Maherin mallia pohjana omalle tutkimukselleen, jossa he pyrkivät tuottamaan todennäköisyyksiä, joita voidaan käyttää hyväksi jalkapallo-otteluiden tulosten ennustamiseen.

Dixon ja Coles (1997) pyrkivät parantamaan Maherin mallia erityisesti juuri vähämaalisten otteluiden osalta käyttämällä erityistä riippuvuusmuuttujaa. Tällä riippuvuusmuuttujalla he pyrkivät lisäämään maalilukujen riippumattomuutta vähämaalisissa otteluissa, erityisesti otteluissa jotka päättyvät tuloksiin 0-0, 1-0, 0-1 ja 1-1. Lisäksi he lisäävät malliin kotietua kuvaavan muuttujan. Kotietu-muuttujan käyttöönottoa Dixon ja Coles perustelevat sillä, että kotivoiton todennäköisyys jalkapallo-ottelussa on 0,46, kun taas todennäköisyys tasapelille ja vierasvoitolle on molemmille 0,27. (Dixon & Coles, 1997). Tässä mallissa ongelmana on, että se olettaa joukkueiden hyökkäys- ja puolustusvoiman olevan staattisia ajan kuluessa, vaikka todellisuudessa nämä joukkueiden ominaisuudet ajan kuluessa usein muuttuvat eri asioista ja tapahtumista johtuen. Mallia parantaakseen Dixon ja Coles (1997) päätyvät käyttämään painotusfunktioita, joka antaa enemmän painoarvoa joukkueiden tuoreille suorituksille, eli otteluiden tuloksille, ja vähemmän painoarvoa vanhemmille tuloksille.

Dixon & Coles (1997) testasivat malliaan aineistoon, joka koostui yhteensä yli 6000 Englannin eri sarjatasojen ottelusta vuosilta 1992-1995. Mallin toimivuus oli lupauksia herättävä. Mallin kehittämisen lisäksi tutkimuksessa oli tarkoitus testata myös kehitetyn mallin soveltuvuutta vedonlyöntiin, eli otteluiden tulosten ennustamiseen. Tarkoitus oli siis yrittää kehittää toimiva vedonlyöntistrategia kehitettyä mallia hyväksi käyttäen. Testimateriaalina on käytetty tuloksia Englannin jalkapallon eri sarjatasoilta kaudelta 1995-96. Tutkimuksessa tarkoituksena oli pelata vain sellaisia vedonlyöntikohteita, joille tutkijoiden kehittämä malli antaa paremmat todennäköisyydet kuin vedonlyöntitoimiston antamat todennäköisyydet olivat. Lopputuloksena oli, että mallia käyttäen voitaisiin saada positiivisia tuloksia vedonlyönnissä ja näin ollen kehitetty vedonlyöntistrategia oli toimiva. (Dixon & Coles, 1997).

Maherin (1982) tavoin Chu (2003) pyrkii osoittamaan, että jalkapallo-ottelun maalit noudattavat Poissonin jakaumaa. Chu käyttää tutkimusaineistona 232 jalkapallo-ottelua jalkapallon maailmanmestaruuskisoista vuosilta 1990, 1994, 1998 ja 2002. Erityisen

tarkkailun kohteena on aika tehtyjen maalien välillä. Tutkimuksessaan Chu olettaa, että kaikki ottelut on pelattu peräkkäin järjestyksessä. Näin ollen edellisen ottelun viimeisen maalin tekoajan ja ottelun keston erotus lisätään seuraavan ottelun ensimmäisen maalin synty aikaan ja näin ollen saadaan aikaan maalintekoaikojen jatkumo läpi kaikkien tutkittujen turnausten.

Chu (2003) olettaa, että jos ajat toteutuneiden maalien välillä ovat jakautuneet eksponentiaalisesti, silloin maalien lukumäärän tietyssä ajassa tulisi olla Poissonin jakauman mukainen. Tämän väitteen todenperäisyyttä tutkimuksessa on mitattu ja arvioitu siten, että oletettujen otteluiden määrä, jossa tehdään yhteensä  $x$  maalia voidaan laskea kertomalla  $x$  maalin Poisson todennäköisyys luvulla 232, eli kaikkien havainnoidujen otteluiden määrällä. Näitä määriä verrataan toteutuneisiin maalimääriin otteluissa.  $\chi^2$ -testin perusteella  $p$ -arvoksi saatiin 0,9745, jonka perusteella ei voida kumota oletusta, että maalimäärät ja niiden tekoajat noudattavat kutakuinkin Poissonin jakaumaa. (Chu, 2003).

Chun (2003) mukaan eräs Poissonin jakauman ominaisuus on, että tapahtumien määrä vaihtoehtoisten aikavälien sisällä on myös Poisson-jakautunut, mutta sillä on suhteellisesti säätyvä keskiarvo. Chu testaa tämän ominaisuuden esiintymistä jalkapallo-otteluissa tutkimalla vertaamalla toteutuneita ja arvioituja maalimääriä 1, 5, 10, 15 ja 45 minuutin aikaväleillä. Havaittiin, että eroavuudet toteutuneiden ja oletettujen maalimäärien välillä oli merkityksettömiä, joten tämä Poissonin jakauman ominaisuus toimii myös jalkapallo-ottelussa. (Chu, 2003).

Julkaisemattomaksi jääneessä tutkimuksessaan Lock & Danehy (1997) tutkivat Poissonin jakauman soveltuvuutta tulosten mallintamiseen ja ennustamiseen Yhdysvaltojen jääkiekon yliopistosarjoissa. Tutkimuksessa oletetaan, että ottelun lopputulos, eli joukkueiden tekemät maalit ottelussa, noudattavat kaksiulotteista Poissonin jakaumaa. Mallissaan Lock & Danehy (1997) lähtevät siitä, että yksittäisen joukkueen kyky tehdä maaleja ottelussa muodostuu sekä joukkueen hyökkäyksen että puolustuksen tehokkuudesta. Lisäksi he ovat ottaneet malliinsa mukaan kotietumuuttujan, joka on jokaiselle joukkueelle sama. Näin ollen saadaan kaava

$$\lambda = O_1 D_2 H^1 / \mu \quad (3)$$

Kaavassa  $O$  kuvaa joukkueen hyökkäysvoimaa ja  $D$  sen vastustajan puolustusvoimaa.  $H$  on kotietumuuttuja ja  $\mu$  kuvaa kaikkien joukkueiden keskimääräistä maalintekotehokuutta. Poissonin jakaumaa soveltaen voidaan laskea todennäköisyys sille, että Joukkue A tekee  $m$  maalia Joukkue B:tä vastaan kaavalla

$$P(m) = \lambda^m e^{-\lambda} / m! \quad (4)$$

Samalla tavalla voidaan luonnollisesti laskea myös todennäköisyys sille, että Joukkue B tekee  $m$  maalia Joukkue A:ta vastaan. Huomattavaa tässä mallissa on, että hyökkäys- ja puolustusvoiman tai -tehon sekä kotietumuuttujan vuorovaikutus ymmärretään moninkertaistavaksi. (Lock & Danehy, 1997).

Testatakseen mallinsa soveltuvuutta ottelun lopputuloksen ennustamiseen, Lock & Danehy (1997) käyttivät aineistona kauden 1996-1997 tilastoja USA:n jääkiekon yliopistoliigasta. He laskivat kaikille joukkueille koko kauden tilastojen perusteella hyökkäys- ja puolustustehokkuuksien luvut sekä kokonaistehokkuusluvun jakamalla hyökkäystehokkuusluvun puolustustehokkuusluvulla. Näitä lukuja hyväksikäyttäen

koko kauden 1996-97 jokaisen ottelun voittaja pyrittiin ennustamaan periaatteella, että joukkue jonka kokonaistehokkuusluku on suurempi tulisi voittaa. Lock & Danehyn (1997) mallin ennustama suosikki voitti 487 ottelua 758 pelatusta ottelusta, eli 64%. Huomioitavaa on, että jokaisen pelin ennustamiseen käytettiin koko kauden tilastoihin perustuvia tietoja, joten tämä ennustus ei ollut täysin realistinen. Saadakseen kokeesta realistisemman tutkijat simuloivat samat ottelut niin, että vain siihen mennessä pelattujen otteluiden tilastoja käytettiin. Tällä menetelmällä malli onnistui ennustamaan oikean voittajan ottelussa 59% tapauksista.

## 2.4. Bayesilaiset menetelmät

Bayesilainen päättely on todennäköisyyslaskennan ja tilastotieteen menetelmä, jossa Bayesin sääntöä käyttämällä päivitetään jonkin tapahtuman tai hypoteesin todennäköisyysarviota kun uutta tietoa opitaan tai uusia todisteita löydetään. Bayesin sääntö on seuraavanlainen:

$$P(H | E) = P(E | H) P(H) / P(E) \quad (5)$$

Todennäköisyys  $P(H)$  on tapahtuman  $P$  prioritodennäköisyys, ts. tapahtuman  $H$  todennäköisyys ennenkuin evidenssiä  $E$  on tarkasteltu. Ehdollinen todennäköisyys  $P(H | E)$  on ns. posterioritodennäköisyys, ts. todennäköisyys että  $H$  sattuu sillä ehdolla, että  $E$  sattuu. Ehdollinen todennäköisyys  $P(E | H)$  voidaan tulkita niin, että se mittaa kuinka hyvin evidenssi  $E$  selittää tapahtuman  $H$  esiintymisen. Todennäköisyydestä  $P(E)$  käytetään nimitystä marginaalitodennäköisyys (marginal likelihood). (Vasama & Vartia, 1971).

Min et al. (2008) kehittivät viitekehysten joukkuelajien otteluiden lopputulosten ennustamiseen. Se hyödyntää sekä sääntöpohjaista että bayesilaista päättelyä. Min et al. (2008) soveltavat viitekehystään luomalla Java-ohjelmointikielellä ohjelmoidun Football Result Expert Systemin (FRES), joka muodostuu sääntöpohjaisesta päättelykomponentista sekä Bayes-verkkokomponentista. Sovellus keskittyy vain jalkapallo-turnausten otteluiden lopputulosten ennustamiseen. Sääntöpohjainen päättelykomponentti pyrkii simuloimaan valmentajan tekemiä päätöksiä ottelun aikana. Data komponentin käyttämien sääntöjen muodostamiseen kerättiin alan kirjallisuudesta sekä haastattelemalla asiantuntijoita. Itse päättelyn sovellus tekee kahdessa vaiheessa, jotka ovat strategian laatiminen ja lopputuloksen laskeminen. Strategia laaditaan asiantuntija-arvioihin ja tilastoihin perustuen, joiden mukaan sovellus ”valmentaa” joukkueita ottelun aikana. Itse strategia ottaa huomioon neljä jalkapallon osa-aluetta: joukkueen hyökkäyksen, puolustuksen, pallonhallinnan ja pelaajien väsymyksen. Nämä osa-alueet on esitetty kukin omana Bayes-verkkonaan. Bayes-verkko antaa todennäköisyysarvioita ottelun kulusta, joiden perusteella sääntöpohjainen päättelykomponentti muuttaa strategiaa tarvittaessa ottelun kuluessa. (Min et al., 2008).

FRES-sovellus pystyi simuloimaan kohtalaisen hyvin jalkapalloturnauksen ottelutuloksia ja tapahtumia. Se pystyi myös ennustamaan lopputuloksia kohtalaisen hyvällä onnistumisprosentilla. Sen sovellettavuus yleisesti on kuitenkin aika pieni, koska se on suunniteltu vain turnausmuotoisten kilpailujen mallintamiseen. (Min et al., 2008).

Suzuki et al. (2010) kehittivät bayesilaisiin menetelmiin perustuvat mallin jalkapallo-ottelun lopputulosten ennustamiseen. Mallia sovellettiin vuoden 2006 jalkapallon

maailmanmestaruuskilpailuihin. Malli perustuu asiantuntijoiden arvioihin otteluista, FIFA:n (Kansainvälinen jalkapalloliitto) maarankingiin sekä turnauksessa aiemmin pelattujen otteluiden tuloksiin. Ottelussa toteutuneet maalimäärät pyritään ennustamaan Poissonin jakauman avulla, jossa muuttujina käytetään pelaavien joukkueiden FIFA-ranking -arvoja. Nämä Poissonin jakaumat sekä asiantuntijoiden ennakoarvot otteluiden lopputuloksista tai voittajista muodostavat Bayes-päätelyn prioritodennäköisyyden.

Itse ottelun lopputuloksen ennustamiseen Suzuki et al. (2010) käyttävät hyväksi asiantuntija-arvioita sekä turnauksessa jo pelattujen otteluiden lopputuloksia, jos otteluita on jo pelattu. Näiden avulla saadaan muodostettua Bayesin säännön likelihood-osa. Jos otteluita ei ole pelattu vielä yhtään, likelihood muodostetaan käyttäen negatiivista binomijakaumaa pelkkiä asiantuntija-arvioita hyväksi käyttäen. Tilanteessa, joissa tietoja aiemmista otteluista on käytössä, likelihood muodostetaan Poissonin jakaumaa käyttäen. Nämä laskelmat tehdään aina molemmille pelaaville joukkueille erikseen ja tämän jälkeen laskettiin todennäköisyydet, että ottelu päättyy kotivoittoon, tasapeliin tai vierasvoittoon. Tutkimuksessa esitetään tuloksia otteluennustuksista vain kuuteen otteluun, joista esitetty malli onnistui ennustamaan viisi ottelua, eli 83%, oikein. (Suzuki et al., 2010).

## 2.5. Elo-luvut

Elo-luku on alunperin shakkia varten kehitetty menetelmä, jolla tilastoihin perustuen pelaajalle lasketaan vahvuusluku, joka kuvaa pelaajan suhteellista taitotasoa. Elo-luvun perusteella voidaan ennakoida pelaajan tulos jossain tulevassa ottelussa jota sitten verrataan toteutuneeseen tulokseen. Jos pelaaja häviää toiselle pelaajalle, jonka Elo-luku on pienempi, pelaajan oma Elo-luku pienenee tästä johtuen. Jos pelaaja taas voittaa isomman Elo-luvun omaavan pelaajan, pelaajan oma Elo-luku vastaavasti nousee. Elo-luku on saanut nimensä kehittäjänsä Arpad Elon mukaan. (Elo, 1978).

Vaikka Elo-järjestelmä alunperin kehitettiin lajeihin, joissa kaksi yksilöä kohtaavat, on siitä kehitetty versioita myös muihin urheilulajeihin kuvaamaan ottelun osapuolien vahvuuksia toisiinsa nähden. Hvattum & Arntzen (2010) kehittivät Elo-luku-järjestelmän jalkapallo-otteluiden lopputulosten ennustamiseen. Elo-lukujen perusteella saatuja kovariantteja sen jälkeen hyödynnetään ordered logit -regressiomallissa, joka tuottaa lopulliset ennustukset. He käyttävät kahta eri Elo-luvun laskemismenetelmää, joiden menestystä verrattiin kuuteen muuhun ottelun lopputuloksen ennustamiseen kehitettyyn menetelmään. Heidän perus-Elo -järjestelmänsä perustuu seuraavaan kaavaan. Oletetaan, että joukkue saa ottelussa voitosta yhden pisteen, tasapelistä puoli pistettä ja häviöstä nolla pistettä. Elo-järjestelmä näin ollen olettaa, että keskimäärin tietyssä ottelussa, kotijoukkueen saavuttama pistemäärä  $\gamma^H$  voidaan laskea kaavalla

$$\gamma^H = 1 / (1 + c^{(R_A - R_H)/d}) \quad (6)$$

Vierasjoukkueen saavuttama pistemäärä  $\gamma^A$  lasketaan vastaavasti kaavalla

$$\gamma^A = 1 / (1 + c^{(R_H - R_A)/d}) \quad (7)$$

Kaavoissa  $R_H$  kuvaa kotijoukkueen Elo-lukua ja  $R_A$  vierasjoukkueen vastaavaa. Kaavassa  $c$  on kerroin, jonka arvo on 10, kuten alkuperäisessä Elo-järjestelmässäkin.

Kaavassa  $d$  on jakaja, joka myös vastaa alkuperäistä Elo-järjestelmää, eli sen arvo on 400. Nämä kaksi muuttujaa luovat mitta-asteikon Elo-luvuille. (Elo, 1978).

Huomioitavaa on että  $\gamma^H + \gamma^A = 1$ . Jos  $\gamma^H$  on esimerkiksi 0,75 ja  $\gamma^A$  on 0,25 voidaan olettaa, että pitkässä sarjassa näiden joukkueiden välisissä otteluissa kotijoukkue saisi keskimäärin 0,75 pistettä ottelua kohti eli voittaisi 75% otteluista. Vierasjoukkueen vastaavasti oletetaan saavan 0,25 pistettä peliä kohden eli voittavan 25 % peleistä. Tätä ennakoitua joukkueen pistemäärää verrataan joukkueen todellisuudessa kyseisessä ottelussa saavuttamaan pistemäärään ja sen mukaan joukkueen Elo-lukua päivitetään kaavalla

$$R_H = R_H + k(\alpha^H - \gamma^H) \quad (8)$$

jossa  $\alpha^H$  on joukkueen todellisuudessa saamat pisteet kyseisestä ottelusta. Muuttujan  $k$  arvosta Hvattum & Arntzen (2010) toteavat, että jos  $k$  on liian pieni, joukkueen Elo-luku ei pysty muuttumaan tarpeeksi nopeasti kun sen suoritustaso nousee. Toisaalta liian korkea  $k$  tekee Elo-luvusta epäluotettavan, koska se vaihtelee silloin liian paljon otteluiden välillä. Perus-Elo-järjestelmään Hvattum & Arntzen (2010) määrittivät  $k:n$  arvoksi 20. Alkuperäisessä Elo-järjestelmässä  $k$  on mestaruustason pelaajille 16, ja heikompi tasoisille pelaajille 32. (Hvattum & Arntzen, 2010).

Hvattum & Arntzen (2010) esittelevät tutkimuksessaan myös maaleihin perustuvan Elo-järjestelmän. Se eroaa perus-Elo-järjestelmästä niin, että Elo-luvun päivitysfunktion  $k$ -kerroin muuttuu muotoon

$$k = k_0(1 + \delta)^\lambda \quad (9)$$

Kaavassa  $\delta$  kuvaa joukkueen pelaaman edellisen ottelun maalieroa. Jos ottelu päättyi esimerkiksi joukkueen 3-1 voittoon, on  $\delta$  2, jos taas joukkue hävisi 1-2  $\delta$  on -1. Tässä maaleihin perustuvassa Elo-järjestelmässä  $k_0$  on 10 ja  $\lambda$  on 1. Näin ollen päivitysfunktion  $k$ -kerroin on lopulta

$$k = 10(1 + \text{maaliero}) \quad (10)$$

Itse ottelun lopputuloksen ennustamiseen Hvattum & Arntzen (2010) käyttivät ordered logit -menetelmää, jonka kovarianttina käytetään koti- ja vierasjoukkueiden Elo-lukujen erotusta kotijoukkueen eduksi. Kovariantti saadaan siis laskemalla  $R_H - R_A$ . Tätä samaa periaatetta sovellettiin tutkimuksessa sekä perus-Elo-järjestelmässä että maaleihin perustuvassa Elo-järjestelmässä. Alustavien Elo-lukujen laskemiseen kullekin joukkueelle tutkimuksessa käytettiin dataa kahden kauden ajalta, kausilta 1993-1995 Englannin jalkapallon neljältä ylimmältä sarjatasolta. Seuraavien viiden kauden 1995-2000 data käytettiin ordered logit -regressiomallin alustavien muuttujien löytämiseksi. Itse mallin testaukseen datana käytettiin tilastoja Englannin jalkapallon neljältä korkeimmalta sarjatasolta kausilta 2000-2008. Hvattum & Arntzen (2010) havaitsivat, että suhteessa verrokkimenetelmiin, heidän kehittämänsä Elo-järjestelmät olivat otteluiden lopputulosten ennustamisessa selkeästi huonompia kuin kaksi verrokkimenetelmää, jotka hyödynsivät datana vedonvälittäjien kertoimia. Tulokset olivat kuitenkin selkeästi parempia kuin muut neljä verrokkimentelmää. (Hvattum & Arntzen, 2010).

## 2.6. Pi-rating

Constantinou & Fenton (2012) esittelevät artikkelissaan Elo-järjestelmän tyyppisen pi-rating -järjestelmän (pi = probability intelligence) joissa joukkueille lasketaan pi-luku. Pi-luku lasketaan kotiotteluille ja vierasotteluille erikseen, eli kaikilla joukkueilla on kaksi eri pi-lukua. Itse järjestelmän perusteena ovat ajatukset kotiedun hyvin todistetusta olemassaolosta, faktasta, että joukkueen tämän hetkistä vahvuutta arvioitaessa joukkueen viimeaikaiset tulokset ovat tärkeämpiä kuin vanhemmat tulokset, sekä ajatuksesta, että voitto on joukkueelle tärkeämpi asia kuin maalieron kasvattaminen tai parantaminen. Maalierolla tarkoitetaan tässä joukkueen tehtyjen maalien ja päästettyjen maalien erotusta koko kaudella johonkin tiettyyn ajanhetkeen mennessä. (Constantinou & Fenton, 2012).

Esitellyt kolme ajatusta otetaan mallissa huomioon seuraavasti. Ensimmäiseksi, kaikille joukkueille on siis arvioitu erikseen koti- ja vieras-pi-luvut ja lisäksi on määritelty oppimisvakio  $\gamma$ , joka määrittää kuinka paljon uusi informaatio joukkueen suorituksista kotiotteluissa vaikuttaa joukkueen suorituskyykyyn vierasotteluissa ja toisinpäin. Constantinou & Fenton (2012) määrittelevät myös toisen oppimisvakion. Tämä vakio  $\lambda$  määrittää, kuinka paljon uusi informaatio joukkueen tekemistä ja päästetyistä maaleista viime otteluissa kumoaa vanhaa informaatiota itse pi-lukujen osalta. Kolmanneksi, suuret maalierot yksittäisessä ottelussa pienennetään eksponentiaalisesti ennen pi-lukujen päivittämistä. (Constantinou & Fenton, 2012).

Pi-rating -malli pyrkii ennustamaan, mikä on tietyn ottelun maaliero joukkueiden välillä. Jos maaliero arvioidaan positiiviseksi, ennakoidaan ottelun päättyvän kotivoittoon, jos se arvioidaan negatiiviseksi, ennakoidaan ottelun päättyvän vierasvoittoon. Molemmille pelaaville joukkueille arvioidaan oma osuutensa ennakoidusta maalierosta, jotka sitten lasketaan yhteen. Ottelun jälkeen kotijoukkueen uuteen kotiotteluiden pi-lukuun vaikuttavat kotijoukkueen kotiotteluiden pi-luku ennen kyseistä ottelua, vierasjoukkueen vierasotteluiden pi-luku ennen kyseistä ottelua sekä ottelun lopputulos, eli tehdyt ja päästetyt maalit. Pi-luvut kehittyvät siis dynaamisesti kauden aikana otteluiden lopputuloksista riippuen. Joukkueen kokonais-pi-luku on sen koti- ja vieras-pi-lukujen keskiarvo. (Constantinou & Fenton, 2012).

Pi-rating -malli eroaa monesta muusta ottelun lopputuloksen ennustamiseen ja joukkueiden arviointiin luodusta järjestelmästä siinä, että pi-rating -mallissa joukkueen kotiotteluiden tuloksilla on vaikutusta myös sen suorituksiin myös vierasotteluissa ja toisinpäin. Näin ollen yksittäisen ottelun jälkeen molempien joukkueiden koti- ja vieras-pi-lukuja päivitetään, tosin eri kaavalla. Kotijoukkueen koti-pi-luku  $R_{1\alpha H}$  päivitetään kaavalla

$$R_{1\alpha H} = R_{0\alpha H} + \Psi_H(e) * \lambda \quad (11)$$

Kotijoukkueen vierasotteluiden pi-luku  $R_{1\alpha A}$  päivitetään kaavalla

$$R_{1\alpha A} = R_{0\alpha A} + (R_{1\alpha H} - R_{0\alpha H}) * \gamma \quad (12)$$

Kaavoissa  $R_{1\alpha H}$  on päivitetty koti-pi-luku ja  $R_{0\alpha H}$  on ottelua edeltävä koti-pi-luku.  $R_{1\alpha A}$  on vastaavasti ottelun jälkeinen päivitetty kotijoukkueen vieras-pi-luku ja  $R_{0\alpha A}$  on ottelua edeltävä kotijoukkueen vierasotteluiden pi-luku. Vierasjoukkueen ottelun jälkeinen vierasotteluiden pi-luku  $R_{1\beta A}$  saadaan kaavalla

$$R_{1\beta A} = R_{0\beta A} + \Psi_H(e) * \lambda \quad (13)$$

ja vierasjoukkueen kotiotteluiden päivitetty pi-luku  $R_{1\beta H}$  kaavalla

$$R_{1\beta H} = R_{0\beta H} + (R_{1\beta H} - R_{0\beta H}) * \gamma \quad (14)$$

Kaavoissa  $\psi(e)$  on painotettu maalierovirhe. Se lasketaan kaavalla

$$\Psi(e) = c * \log_{10}(1 + e) \quad (15)$$

jossa  $e$  on ennakoidun ja toteutuneen maalieron erotus ja  $c$  on vakio jolle Constantinou ja Fenton käyttävät arvoa 3. Painotetulla maalierovirheellä eksponentiaalisesti pienennetään suuria maalieroja ja niiden vaikutusta. Kaavoissa oppimisvakio  $\lambda$  saa aina arvon 0.1 ja oppimisvakio  $\gamma$  saa aina arvon 0.3. (Constantinou & Fenton, 2012).

Kotijoukkueen  $i$  ennustettu osuus odotetusta maalierossa tietyssä ottelussa saadaan kaavalla

$$b^{R_{i\alpha H} / c} - 1 \quad (16)$$

Koska kaavassa 15 käytetään 10-kantaista logaritmia, sen nojalla  $b$  on aina 10. Vierasjoukkueen  $j$  ennustettu osuus odotetusta maalierosta tietyssä ottelussa saadaan vastaavasti kaavalla

$$b^{R_{j\alpha A} / c} - 1 \quad (17)$$

Koko ottelun odotettu maaliero saadaan vähentämällä vierasjoukkueen odotettu maaliero kotijoukkueen odotetusta maalierosta. Ottelun lopputuloksen ollessa selvillä, odotettua maaliero verrataan ottelun toteutuneeseen maalieroon vähentämällä toteutuneesta maalierosta odotetun maalieron itseisarvo. Näin saadaan toteutuneen ja odotetun maalieron välinen virhe, josta voidaan laskea sitten painotettu maalierovirhe. Jos odotettu maaliero oli pienempi kuin toteutunut maaliero, kotijoukkueen painotettu maalierovirhe on laskettu painotettu maalierovirhe suoraan. Kotijoukkue suoriutui siis odotuksia paremmin. Tässä tilanteessa vierasjoukkueen painotettu maalierovirhe on varsinaisen maalierovirheen vastaluku: se suoriutui ottelussa odotettua heikommin. Jos odotettu maaliero oli suurempi kuin toteutunut, kotijoukkueen painotettu maalierovirhe on varsinaisen maalierovirheen vastaluku. Kotijoukkue suoriutui siis odotettua heikommin. Vastaavasti tässä tilanteessa, vierasjoukkue suoriutui ottelusta odotettua paremmin, ja se saa painotetuksi maalierovirheeseen varsinaisen painotetun maalierovirheen. (Constantinou & Fenton, 2012).

Kun pelatun ottelun jälkeinen painotettu maalierovirhe on tiedossa, voidaan molempien pelanneiden joukkueiden sekä kotiotteluiden että vierasotteluiden pi-luvut päivittää. Oletetaan, että kotijoukkueen  $\alpha$  ja vierasjoukkueen  $\beta$  pi-luvut ovat  $\{ R_{\alpha H} = 1,6, R_{\alpha A} = 0,40 \}$  ja  $\{ R_{\beta H} = 0,3, R_{\beta A} = -1,2 \}$ . Tällöin ottelun odotettu maaliero on 3,9264 kaavojen 16 ja 17 perustella. Oletetaan, että ottelu päättyy tulokseen 4-1. Toteutunut maaliero on siis +3, joka on 0.9264 pienempi kuin odotettu maaliero, joka on siis

maalierovirhe. Painotettu maalierovirhe on näin ollen vierasjoukkueelle 0.8542 ja koska kotijoukkue suoriutui odotettua heikommin, sen painotettu maalierovirhe on -0.8542. Kotijoukkueen kotiotteluiden pi-luku oli ennen ottelua 1.6 ja sen vierasotteluiden pi-luku 0.4. Vierasjoukkueen vierasotteluiden pi-luku ennen ottelua oli -1.2 ja sen kotiotteluiden pi-luku 0.3. Kotijoukkueen kotiotteluiden pi-luku päivitetään kaavalla  $1.6 + (-0.8542) * 0.1$  josta saadaan 1.5145, eli se laski ottelua edeltävästä luvusta. Kotijoukkueen vierasotteluiden pi-luku on ottelun jälkeen  $0.4 + (1.5145 - 1.6) * 0.3 = 0.3743$ . Eli myös vierasotteluiden pi-luku laski. Vierasjoukkueen vierasotteluiden pi-luku on ottelun jälkeen  $-1.2 + (+0.8542) * 0.1 = -1.1145$ , eli se parani hiukan. Sen kotiotteluiden pi-luku nousi myös hiukan ollen ottelun jälkeen  $0.3 + (-1.1145 - (-1.2)) * 0.3 = 0.3256$ . (Constantinou & Fenton, 2012).

Constantinoun ja Fentonin (2012) artikkelissa ei suoraan esitetä tai kerrota, kuinka pi-lukujen perusteella laskettua odotettua maaliero voidaan käyttää hyväksi ottelutulosten ennustamisessa. Henkilökohtaisesti sähköpostilla asiasta Constantinoulta kysyttäessä, vastaukseksi saatiin, että todennäköisyydet eri otteluiden lopputuloksille saadaan monimutkaisen Bayes-päätelyn avulla. Myöhemmin tässä tutkimuksessa ja sen algoritmissa todennäköisyydet arvioidaan historiatietoihin perustuen päättelemällä.

Constantinou & Fenton (2012) vertaavat pi-rating -malliansa Hvattum & Arntzen (2010) artikkelissa esitettyihin perus-Elo-järjestelmään ja maaliperustaiseen Elo-järjestelmään. Constantinoun & Fentonin (2012) mukaan pi-rating -malli on huomattava parannus Hvattumin & Arntzenin (2010) Elo-järjestelmiin verrattuna ja sen avulla päästiin pitkällä aikavälillä varsin kannattavaan vedonlyöntitulokseen. Pi-rating -mallin saavuttamat voittoprosentit vaihtelivat 28 prosentista 37 prosenttiin. Constantinou & Fenton (2012) testasivat malliansa jalkapallon Englannin Valioliigaan kausilla 2007-2012. (Constantinou & Fenton, 2012).

## 2.7. Kotiedusta ja sen merkityksestä

Kotiedulla joukkueurheilussa tarkoitetaan ilmiötä, jossa kotona pelaavalla joukkueella katsotaan olevan suurempi todennäköisyys voittaa ottelu kuin vieraissa pelaavalla joukkueella. Kotietuun vaikuttavia ja sen muodostavia tekijöitä on monia. Oberhofer et al. (2010) havaitsivat, että pelaavien joukkueiden kotipaikkojen etäisyydellä toisistaan on merkitystä jalkapallo-ottelun lopputulokselle. He havaitsivat, että kotipaikkojen etäisyys lisää vierasjoukkueen taipumusta päästää maaleja ja vaikuttaa myös vierasjoukkueen kykyyn tehdä maaleja. Tosin jälkimmäisen vaikutus todettiin varsin pieneksi. (Oberhofer et al., 2010). Syy, miksi joukkueiden kotipaikkojen etäisyydellä on merkitystä on, että vierasjoukkue joutuu matkustamaan toiselle pelipaikkakunnalle. Toisinaan jopa ottelua edeltävänä päivänä. Riippuen vierasjoukkueen taloudellisesta tilanteesta ja sarjasta, tehdään matkat yleensä joko bussilla tai lentämällä. Näistä luonnollisesti ensin mainittu on raskaampi pelaajan ja joukkueen kannalta, koska bussimatkat kestävät pitempään kuin lentomatkat. Toisaalta, esimerkiksi Venäjän jääkiekon KHL-liigassa etäisyydet pelipaikkojen välillä voivat olla tuhansiakin kilometrejä, jolloin lentomatoistakin muodostuu pitkiä, useita tunteja kestäviä. Yhdysvaltojen NHL-liigassa taas joukkeet tekevät usein pitkiäkin vierasottelukiertueita, joissa otteluita vieraissa saatetaan pelata peräkkäisinä päivinä, toisinaan samaa vastustajaa vastaan, toisinaan eri vastustajaa vastaan. Smith et al. (2000) tulivat tutkimuksessaan kuitenkin siihen tulokseen, että NHL-jääkiekossa vierasjoukkueen vierasotteluun matkustamien kilometrien määrällä ei ole positiivista vaikutusta kotietuun. Sen sijaan he havaitsivat, että kilometrimäärän kasvaessa, kotiedun merkitys väheni. Eräksi syyksi tälle he arvelivat, että matkustaessa joukkue on paljon tiiviimmin toistensa kanssa tekemisissä, joka kasvattaa joukkuehenkeä. (Smith et al., 2000).



Toinen kotietuun ja sen suuruuteen ja muodostumiseen vaikuttava seikka on yleisö. Kun joukkue pelaa kotona, on katsojista yleensä suurempi osa, usein merkittävän suuri osa, kotijoukkueen kannattajia, jotka kannustuksellaan saavat kotijoukkueen pelaamaan paremmin. Agnew & Carron (1994) tutkivat kotiedun merkitystä jääkiekossa. Heidän mukaansa yli 58% heidän analysoimistaan jääkiekko-otteluista kotiedulla oli vaikutusta. Kun otteluista tasapeliottelut jätettiin analysoimasta, kotietu vaikutti yli 61% otteluista. Kotiedun merkittävimmäksi syyksi Agnew & Carron (1994) löysivät juuri yleisömäärän. Kun yleisömäärä kasvoi, myös kotietu kasvoi. (Agnew & Carron, 1994).

Myös itse kotikenttä voi kirjaimellisesti tuottaa etua kotijoukkueelle, eli sille joukkueelle, joka on tottunut pelaamaan kyseisellä kentällä. Barnett & Hilditch (1993) tutkivat keinonurmen vaikutusta jalkapallo-otteluiden lopputuloksiin, ja erityisesti sen tuottamaa mahdollista etua kotijoukkueelle. Tutkimuksen kohteena oli neljä eri seuraa eri sarjatasoilla Englannin jalkapallo-sarjoissa, jotka ottivat keinonurmen käyttöön 1980-luvun puolivälissä. Dataa analysoitiin kahdeksan kauden ajalta. Tutkimuksessa havaittiin, että keinonurmi tuotti etua kotijoukkueelle. Etu oli tilastollisesti merkittävä, mutta käytännössä aika pieni. (Barnett & Hilditch, 1993).

Myös tuomarit saattavat suosia kotijoukkuetta päätöksissään, mikä myöskin edistää kotiedun muodostumista. Greer (1983) tutki yleisön buuaksen ja reagoinnin vaikutusta tuomareihin koripallo-otteluissa ja havaitsi, että kun yleisö oli reagoinut voimakkaan negatiivisesti johonkin ottelutilanteeseen, tämän tilanteen jälkeen tuomarit usein suosivat kotijoukkuetta jonkin aikaa, joka johti kotietuun. (Greer, 1983). Nevill et al. (2002) tutkivat myös yleisön aiheuttaman metelin vaikutusta tuomareiden työskentelyyn. He havaitsivat, että yleisön metelöinnillä oli dramaattisia vaikutuksia tuomareiden päätöksiin. Tuomarit, jotka toimivat suuressa melussa olivat huomattavasti epävarmempia tuomioissaan ja tuomitsivat 15,5% vähemmän rikkeitä kotijoukkueelle. (Nevill et al. 2002).

Kotietua käytetään joissain joukkuelajeissa myös kannustimena. Esimerkiksi jääkiekossa, pudotuspelien alkaessa sarjataulukossa korkeammalla ollut joukkue saa pelata ensimmäisen pudotuspeliottelunsa tietyssä pudotuspelisarjassa kotona. Pudotuspelisarjoissa merkitsevät vain voitot, joita tulee jatkoon päästäkseen saada yleensä kolme tai neljä.

Kotiedun mallintamiseen ja sen mittaamiseen on olemassa monia eri lähetymistapoja. Lock & Danehy (1997) olettavat, että kotietu on läpi kauden sama, ja se on myös joka joukkueelle sama. Heidän mallissaan kotietu lasketaan maaleissa: kuinka monta maalia kotijoukkue keskimäärin tekee enemmän kuin vierasjoukkue. Vastaavasti vierasjoukkueen katsotaan tekevän kotiedun verran vähemmän maaleja ottelussa. Itse tutkimuksessa kotietu oli 1.078 maalia. Käytännössä itse mallissa kotietu katsotaan prosentuaaliseksi kasvuksi kotijoukkueen maalintekovoimassa, ja laskuksi vierasjoukkueen maalintekovoimassa. (Lock & Danehy, 1997). Clarke & Norman (1995) taas laskivat kotiedun jokaiselle Englannin jalkapallon neljällä ylimmällä sarjatasolla pelaavalle joukkueelle. He laskivat maalien ja voittojen erot koti- ja vierasotteluissa ja sen jälkeen pienimmän neliösumman menetelmää hyväksikäyttäen määrittivät kotiedun kullekin joukkueelle. (Clarke & Norman, 1995).

## 2.8. Yhteenveto kirjallisuudesta ja teoriasta

Edellä on laajasti käsitelty jääkiekon ja urheiluviedonlyönnin perusteita. Lisäksi on käyty läpi kolme kirjallisuudessa käsiteltyä eri menetelmää, jotka pyrkivät ennustamaan otteluiden lopputuloksia. Sekä Maher (1982) että Dixon & Coles (1997) käyttivät

menestyksekkäästi tutkimuksissaan hyväksi Poissonin jakauman eri variaatioita. Molemmat tutkivat Poissonin jakauman soveltuvuutta jalkapallo-otteluiden lopputulosten ennustamisessa. Lock & Danehy (1997) loivat myös Poissonin jakaumaan perustuvan mallin ja onnistuivat varsin hyvällä todennäköisyydellä ennustamaan Yhdysvaltojen yliopistojääkiekkosarjan otteluiden lopputuloksia.

Hvattum & Arntzen (2010) ottivat mallissaan täysin toisenlaisen lähestymistavan verrattuna Poissonin jakaumaan. Heidän mallinsa noudattaa melko tarkasti Arpad Elo Elo-luku -järjestelmää, joka alunperin kehitettiin mittamaan shakinpelaajien keskinäistä paremmuutta. (Elo, 1978). Hvattum & Arntzen (2010) kehittivät järjestelmänsä jalkapallojoukkueiden paremmuuden vertailuun ja ordered logit -mallia hyväksikäyttäen myös loivat tavan ennustaa otteluiden lopputuloksia. He loivat kaksi eri Elo-perustaista mallia, joista ensimmäinen noudatti alkuperäistä Elo-luku-järjestelmää täysin, ja toinen otti huomioon myös otteluiden toteutuneen maalieron Elo-lukujen päivityksen yhteydessä. Saadut tulokset eivät ole menetelmävertailussa parhaiten menestyviä, mutta eivät myöskään huonoiten menestyviä menetelmiä. (Hvattum & Arntzen, 2010).

Constantinou & Fenton (2012) käyttivät Hvattumin & Arntzenin (2010) Elo-perustaista järjestelmää perustana omalle pi-rating -mallilleen. Sen ytimenä on ajatus kotiedun merkittävydestä, viimeisimpien tulosten tärkeydestä verrattuna vanhoihin tuloksiin sekä voiton suuremmasta tärkeydestä verrattuna maalieron parantamiseen. Heidän mallissaan kaikille joukkueille on omat pi-luvut sekä kotiotteluille että vierasotteluille. Molempia myös päivitetään jokaisen ottelun jälkeen, olipa kyse sitten koti- tai vierasottelusta. Pi-lukujen päivitykseen vaikutti pi-lukujen perustella arvioidun ennakoitun maalieron ja toteutuneen maalieron erotus. Jos toteutunut maaliero oli suurempi kuin arvioitu, katsottiin voittaneen joukkueen suoriutuneen odotuksia paremmin ja sen pi-lukuja kasvatettiin. Jos taas toteutunut maaliero oli pienempi kuin ennakoitu, katsottiin voittaneen joukkueen suoriutuneen odotuksia heikommin ja näin ollen sen pi-lukuja pienennettiin. Kun maaliero oli ennakoitua suurempi, katsottiin hävinneen joukkueen luonnollisesti silloin suoriutuneen vielä odotuksiakin huonommin. Maalieron ollessa odotettua pienempi, oli hävinnyt joukkue suoriutunut odotuksia paremmin.

Tutkimuksessaan Constantinou & Fenton (2012) havaitsivat, että heidän pi-rating järjestelmänsä suoriutui jalkapallo-otteluiden lopputulosten ennustamisesta paremmin kuin Hvattum & Arntzenin (2010) molemmat Elo-järjestelmät. Pi-rating -mallin avulla he onnistuivat harrastamaan voitollista vedonlyöntiä. (Constantinou & Fenton, 2012).

Kotiedun merkitys joukkueurheilussa on kirjallisuuden perusteella suuri. Syitä tälle ovat esimerkiksi, että vierasjoukkue joutuu matkustamaan ottelupaikkakunnalle, pahimmissa tapauksissa jopa tuhansia kilometrejä. Myös kotijoukkueen yleisöllä ja sen toiminnalla on usein vaikutusta vierasjoukkueen suoritukseen. Greer (1983) havaitsi, että joissain tapauksissa jopa tuomarit suosivat kotijoukkuetta, usein yleisön pauhun aiheuttaman paineen seurauksesta. (Greer, 1983). Lock & Danehy (1997) käyttivät tutkimuksessa samaa kotietuvakiota kaikille joukkueille joka ottelussa läpi koko kauden, Clarke & Norman (1995) sen sijaan laskivat erilliset kotietumuuttajat jokaiselle joukkueelle. (Clarke & Norman, 1995).

Seuraavaksi käydään yksityiskohtaisesti läpi luodut algoritmikonstruktiot ja niiden kehitys. Lisäksi käydään läpi myös ohjelman graafinen käyttöliittymä ja sen kehitysprosessi.

### 3. Mallin ja algoritmin kehittäminen

Tässä tutkimuksessa ja kehitetyssä algoritmissa käytetään hyväksi kaikkia kolmea edellisessä kappaleessa mainittua menetelmää, joita on kirjallisuudessa käytetty otteluiden lopputulosten ennustamiseen. Aluksi hyödynnetään hyvin yksinkertaista Poissonin jakauman mallia, jossa molemmille joukkueille lasketaan riippumattomalla Poissonin mallilla todennäköisyydet maaliluvuille. Kotietu otetaan tässä huomioon niin, että kotijoukkueen maalikeskiarvoon, joka syötetään Poisson-funktiolle, lasketaan mukaan vain kotiotteluiden maalit, ja vierasjoukkueen maalikeskiarvoon vain vierasotteluissa tehdyt maalit. Sen jälkeen hyödynnetään suoraan Hvattum & Arntzenin (2010) Elo-järjestelmää, mutta jalkapallon sijasta sen soveltuvuutta testataan siis NHL-jääkiekkoon. Molempia tutkimuksessa esitettyjä Elo-malleja, perus-Elo-järjestelmää ja maaliperustaista Elo-järjestelmää hyödynnetään algoritmin kehittämisessä. Lopputuloksien ennustamista varten luodaan yksinkertainen historiatietoihin perustuva luokittelu, jonka perusteella arvioidaan tasapelien todennäköisyydet. Itse Elo-järjestelmien antamat odotetut pistemäärät otteluihin koti- ja vierasjoukkueelle käsitetään suoraan todennäköisyyksiksi koti- ja vierasvoitoille.

Kolmanneksi kehitetään algoritmi, joka hyödyntää Constantinoun & Fentonin (2012) pi-rating -menetelmää ja siitä käytetään nimitystä pi-hockey. Myös pi-hockey soveltaa täysin Constantinoun & Fentonin (2012) artikkelissaan esittelemää pi-rating -mallia, mutta sovelluskohde on siis eri. Todennäköisyysarvioiden muodostamiseksi otteluiden eri lopputuloksille luodaan samantyyppinen menetelmä kuin Elo-järjestelmiäkin varten. Tässä tapauksessa historiatietoihin perustuen luodaan frekvenssitaulukko algoritmin muodostamille eri odotetuille maalieroille otteluissa. Todennäköisyysarviot muodostetaan sen mukaan, kuinka suuri osa otteluista joille pi-hockey -mallin avulla arvioitiin joukkueiden välille tietty maaliero, päättyi kotivoittoon, tasapeliin tai vierasvoittoon. Näitä prosenttiosuuksia käytettiin sitten otteluiden ennustamiseen.

Lopuksi kehitettiin vielä aggregaatti- eli yhdistelmämalli, joka käyttää hyväksi kaikkia kolmea mallia, Poisson-mallia, Elo-mallia ja pi-hockey -mallia. Jokaiseen otteluun lasketaan jokaisella näillä mallilla todennäköisyydet eri lopputuloksille, ja lopulliseksi todennäköisyyksiksi lasketaan kaikkien näiden kolmen todennäköisyysarvion keskiarvot. Esimerkiksi siis kotivoiton todennäköisyys ottelussa saadaan silloin kaavalla

$$(p1 + p2 + p3) / 3 \quad (18)$$

jossa  $p1$  vastaa Poisson-mallin antamaa kotivoiton todennäköisyyttä,  $p2$  on Elo-mallin kotivoiton todennäköisyys ja  $p3$  pi-hockey mallin arvioima todennäköisyys kotivoitolle. Sama toistetaan tasapelin ja vierasvoiton todennäköisyyksille. Näidenkin todennäköisyysarvioiden summa on aina yksi.

Varsinainen algoritmin ja sitä hyödyntävän sovelluksen ohjelmistokehitystyö tehtiin Java-ohjelmointikielellä Eclipse-kehitysympäristöä käyttäen. Algoritmi hyödyntää MySQL-tietokantaa.

#### 3.1. Vedonlyönnin simulointi

Aivan aluksi algoritmiin kehitettiin vedonlyönnin simuloinnin mahdollistava osa. Se on varsin yksinkertainen. Wallet -niminen luokka kuvaa pelaajan pelikassaa ja sen kulloistakin tilaa. Luokka pitää kirjaa pelikassan saldosta, voittojen lukumääristä ja

rahallisesta määrästä, sekä lyötyjen vetojen määrästä ja niiden rahallisesta suuruudesta. Luokasta löytyy myös metodi, jonka avulla voidaan rajoittaa tai kasvattaa otettavan riskin määrää. Tämä tapahtuu niin sanotun Kellyn jakajan avulla. Kellyn kaavan anatama suosituspäätös jaetaan Kellyn jakajalla, jolloin saadaan lopullinen päätös. Yleensä Kellyn jakaja on positiivinen kokonaisluku väliltä 1-10. Oletuksena Wallet -luokka asettaa Kellyn jakajaksi 5.

Wallet -luokka on toteutettu Singleton -suunnittelumallia hyväksi käyttäen joten siitä voidaan luoda vain yksi ajonaikainen olio. Tämä mahdollistaa sen, että pelikassan saldon ja muiden muuttujien päivittäminen on helppoa ja esimerkiksi saldo säilyy oikeana vaikka luokkaa käytettäisiin ja kutsuttaisiin useammastakin luokasta ohjelman ajan aikana.

Vedonlyönnin simuloimiseen tarvitaan myös vedonvälittäjien antamia todellisia kertoimia otteluille. Tietokannassa vedonvälittäjien kertoimia löytyy kaudesta 2008-2009 kauteen 2011-2012. Kertoimiin liittyvät toiminnot ja metodit on koottu Odd -nimiseen luokkaan. Odd -luokan bet-metodilla tapahtuu varsinainen vedonlyönti. Se saa syötteenään kertoimet, algoritmin muodostamat todennäköisyydet ottelulle, ottelun oikean lopputuloksen sekä Wallet -olion. Aluksi metodi vähentää Kellyn kaavan mukaisen panoksen pelikassasta. Sen jälkeen tarkastetaan menikö veto oikein. Jos veto meni oikein, lisätään voitto pelikassan saldoon. Lisäksi voittoja ja vetoja laskevia laskureita päivitetään asianmukaisesti. Odd -luokassa on myös metodi, joka päättää, löytyykö tietyistä otteluista pelattavaa, eli onko algoritmin muodostamista todennäköisyyksistä jokin isompi kuin vedonvälittäjien antama todennäköisyys. Jos tällainen ylikerroin löytyy, metodi kehottaa lyömään vetoa. Luokassa on myös metodi, joka hakee parhaat kertoimet kyseessä olevaan otteluun tietokannasta.

Jotta algoritmin saavuttamien tulosten seuraaminen olisi helpompaa ja vaivattomampaa, luotiin myös Counter-luokka, joka peritään Model-luokasta. Sen tehtävänä on laskea, kuinka moni algoritmin kotivoitoiksi ja vierasvoitoiksi ennustamista onnistutaan ennustamaan oikein, ja kuinka moni ennustus menee väärin. Counter-luokka hyödyntää Singleton-suunnittelumallia, joten sen käyttö on helppoa, koska voidaan olla varma, että kaikki tulokset tallennetaan aina samaan paikkaan. Luokalla on print -niminen metodi, jota käytetään hyväksi kun tulokset tulostetaan graafisen käyttöliittymän avulla käyttäjälle. Counter-luokka palautetaan takaisin käyttöliittymälle, kun halutut simuloinnit on suoritettu.

### 3.2. Poisson-malli 1

Seuraavaksi kehitettiin mahdollisimman yksinkertainen algoritmi ennustamaan otteluiden lopputuloksia. Tämä algoritmi käyttää hyväkseen kahden satunnaismuuttujan riippumattomia Poissonin jakaumaa, jonka avulla lasketaan todennäköisyydet sekä koti- että vierasjoukkueen maalimäärille. Näiden todennäköisyyksien perusteella lasketaan todennäköisyydet kotivoitolle, tasapelille ja vierasvoitolle, kuten Poisson-esimerkissä luvussa 2.3 tehtiin. Parametreinä Poissonin jakaumalle käytetään kotijoukkueella sen kaudella siihen mennessä pelaamien kotiotteluiden tehtyjen maalien keskiarvoa, sekä vierasjoukkueelle sen kyseisellä kaudella siihen mennessä pelaamissa vierasotteluissa tekemien maalien keskiarvoa. Rajoitteena on, että kunkin pelaavan joukkueen on pitänyt pelata vähintään kolme ottelua, kotijoukkueen kolme kotona ja vierasjoukkueen kolme vieraisissa, jotta todennäköisyydet lopputuloksille lasketaan. Tällä tavalla saadaan varmistettua, että kauden alku ei tuota vääristymiä saatuihin tuloksiin.

Käytettävissä mallissa on huomioitavaa, että se ottaa huomioon mahdollisen kotijoukkueen kotiedun sekä huomioi myös vierasjoukkueen kyvyt vierasotteluissa. Lisäksi malli ottaa huomioon, että joukkueiden kyvyt, taidot ja taktiikka muuttuvat kauden aikana säilymättä stabiileina koko kautta.

Vaadittava algoritmi on varsin yksinkertainen. Aluksi se hakee tietokannasta pelaavien joukkueiden kyseisellä kaudella pelaamat ottelut, kotijoukkueen kotiottelut sekä vierasjoukkueen vierasottelut, ja laskee joukkueen tekemien maalien keskiarvon noissa otteluissa. Sen jälkeen kyseistä keskiarvoa parametrinä käyttämällä muodostetaan Poissonin jakauma maalimäärille nollasta seitsemään (0-7). Tämä tehdään molemmille joukkueille. Tämän jälkeen algoritmi muodostaa molempien joukkueiden maalilukujen kaikki mahdolliset kombinaatiot ja laskee todennäköisyydet näille tuloksille. Todennäköisyys kotivoitolle saadaan laskemalla yhteen sellaisten tulosten todennäköisyydet, jossa kotijoukkueen maalimäärä on suurempi kuin vierasjoukkueen. Samalla periaatteella saadaan todennäköisyydet myös tasapelille ja vierasvoitolle. Nämä laskelmat algoritmi tekee jääkiekon NHL-liigan otteluille kaudelta 2002-2003 kaudelle 2011-2012, lukuun ottamatta kautta 2004-2005, jota ei pelattu. Otteluille, joissa joukkueista ei ole käytettävissä kotijoukkueelle kolmen kotiottelun ja vierasjoukkueelle kolmen viimeisimmän vierasotteluiden tuloksia, todennäköisyyksiä ei lasketa. Ottelun ennustettu lopputulos on se, jolle algoritmi laskee suurimman todennäköisyyden. Tätä tulosta algoritmi vertaa kyseisen ottelun oikeaan, toteutuneeseen tulokseen.

Algoritmi toteutetaan Java-kielellä model-view-controller -arkkitehtuurimallia hyödyntäen. Algoritmin luokkarakenne koostuu kolmesta yliluokasta Controller, Model ja View. Kaikki muut algoritmin luokat peritään näistä kolmesta luokasta. Esimerkiksi jokaista tietokantataulua vastaa yksi Model-luokasta peritty aliluokka. Tietokannassa jokaisen taulun nimi on monikossa englanniksi, kuvaten alkioita, joita kyseiseen tauluun on tallennettu. Esimerkiksi kaikkien otteluiden perustiedot on tallennettuna games-tauluun, jonka yksittäistä alkioita algoritmissa mallintaa Game-luokka. Game-luokkaa käyttää GameController-luokka, joka peritään Controller-yliluokasta. Games-paketti sisältää kaikki peleihin liittyvät graafisen käyttöliittymän tiedon esittämiseen tarvittavat luokat.

Käytettävä tietokanta sisältää taulut games ja teams. Games-tilusta löytyy viittaukset ottelussa pelanneisiin joukkueisiin home\_team ja away\_team. Näiden joukkueiden nimet löytyvät vastaavasti teams-tilusta. Lisäksi games-tilu sisältää kotijoukkueen tekemät maalit ottelussa (home\_goals) sekä vierasjoukkueen tekemät maalit (away\_goals). Lisäksi ottelusta on tallennettu sen pelipäivämäärä (date) sekä kausi, jolloin ottelu on pelattu (season). Yksittäiseen otteluun muissa taulussa viitataan game\_id -muuttujan avulla. Teams-tilussa on tallennettuna joukkueiden nimet joihin viitataan muissa tauluissa id -muuttujaa käyttäen.

Algoritmiä hyödyntävä Java-sovellus toimii niin, että ohjelman avausikkunasta voidaan vetoalvikosta valita kausi, jonka otteluita algoritmin halutaan simuloivan. Kun kausi on valittu ja painettu OK-painiketta, algoritmi simuloi kyseisen kauden ottelut, ja vertaa simuloitujen lopputulosten vastaavuutta ja osuvuutta oikeisiin tuloksiin. Simuloinnin päätyttyä ohjelma näyttää, kuinka suuri osuus simuloitujen otteluiden lopputuloksista onnistuttiin ennustamaan oikein.

Koska tässä tutkimuksessa on tarkoitus selvittää, pystytäänkö algoritmin avulla historiatietoihin pohjaten ennustaa jääkiekko-otteluiden tuloksia, algoritmi on tarkoitus säilyttää mahdollisimman yksinkertaisena. Tästä syystä käytetään Java-kielen double tietotyyppiä, vaikkakin sen käyttö aiheuttaa jonkin verran epätarkkuuksia tehdyissä laskelmissa. Nämä epätarkkuudet esimerkiksi todennäköisyyksiä laskettaessa ovat vain promillen kokoisia, joten niiden vaikutus varsinaisiin ennustuksiin on merkityksetön.

### 3.3. Poisson-malli 2

Toiseen konstruktion algoritmiin tehdään pieni muutos. Ensimmäisessä Poisson-mallissahan algoritmi laski maalikeskiarvot kaikista siihen mennessä pelatuista otteluista. Seuraavassa Poisson-mallissa on tarkoitus ottaa huomioon paremmin joukkueen sen hetken kunto. Tämä toteutetaan niin, että aluksi otetaan huomioon vain kolme viimeistä ottelua, kotijoukkueelle kolme viime kotiottelua sekä vierasjoukkueelle kolme viime vierasottelua. Tämän jälkeen kokeillaan saadaanko ennustuksia parannettua ottamalla huomioon viimeiset neljä, viisi ja kuusi ottelua. Algoritmin osalta tämä on varsin pieni muutos, koska se jo osaa ottaa huomioon, kuinka monen viime pelin maalimäärät keskiarvoon otetaan mukaan.

### 3.4. Elo-malli 1

Kolmas konstruktio eli Elo-malli 1 hyödyntää suoraan paperissa Hvattum & Arntzen (2010) esiteltyä Elo-lukuihin perustuvaa mallia, ja aluksi nimenomaan  $ELO_b$  mallia, eli pelkästään Elo-lukuihin perustuvaa perusmallia. (Hvattum & Arntzen, 2010). Ainoa poikkeus on, että sen sijaan, että todennäköisyydet eri lopputulosvaihtoehdoilla olisi johdettu ordered logit -mallia käyttäen, on mallin antamat odotettavissa olevat pisteet joukkuetta kohden ymmärretty suoraan todennäköisyyksiksi. Jos mallin perusteella kotijoukkueen odotetaan saavan tietyssä ottelussa 0,55 pistettä ja vierasjoukkueen 0,45 pistettä, ymmärretään nämä aluksi tarkoittamaan, että kotijoukkueen voittotodennäköisyys on 55% ja vierasjoukkueen voittotodennäköisyys 45%. Tasapelin todennäköisyys arvioidaan aiempien kausien tilastoihin perustuen ja tämän todennäköisyyden puolikas vähennetään sekä kotivoiton että vierasvoiton todennäköisyydestä. Kunkin joukkueen oletus-Elo-luku kauden alussa on 400, eikä siihen siis vaikuta esimerkiksi edellisen kauden Elo-luku. Näin ollen jokaisen joukkueen katsotaan olevan kauden alussa yhtä vahva suhteessa toisiinsa. Tämä ei luonnollisestikaan pidä paikkaansa, ja saattaa vaikuttaa saatuihin ennustuksiin.

Elo-lukujen laskemista ja käsittelyä varten kehitettiin EloRatingsController-luokka joka peritään Controller-luokasta. Lisäksi luotiin oma Model-luokka, EloRating, suorittamaan itse Elo-lukujen laskutoimitukset. EloRatingsController-luokka laskee kaikki todennäköisyydet eri lopputuloksille. Tämä on varsin hidas prosessi, koska todennäköisyyksien laskemiseen käytetään tietokantaa. Tätä voisi ehkä tulevaisuudessa nopeuttaa lukemalla kaikki Elo-lukujen todennäköisyyskategoriat yhdellä kerralla esimerkiksi taulukkoon. Vaikka taulukostakin tiedot pitäisi joka kerta aina erikseen hakea, olisi se kuitenkin selkeästi nopeampi ja tehokkaampikin tapa tietokantahakuihin verrattuna.

Tasapelin todennäköisyyden määrittämiseksi, kausien 2002-2008 jokaisen otteluun laskettiin kunkin pelaavan joukkueen Elo-luvut. Tämän jälkeen laskettiin joukkueiden Elo-luvun erotus jokaiseen otteluun. Jos erotus oli negatiivinen, muutettiin se luvun vastaluvuksi ja pyöristettiin lähimpään kokonaislukuun. Saadut luvut järjestettiin suurimmasta pienimpään, ja jaettiin sen jälkeen suuruutensa perusteella kategorioihin  $> 250$ , 200-249, 150-199, 100-149, 50-99 ja  $< 50$ . Kustakin kategoriasta laskettiin tasapeleihin päättäneiden otteluiden prosentuaalinen osuus kyseisessä kategoriassa. Tämän prosenttiosuuden katsotaan olevan myös tasapelin todennäköisyys otteluissa, joissa pelaavien joukkueiden Elo-lukujen erotus kuuluu kyseiseen kategoriaan. Lopullisten todennäköisyyksien laskemiseksi kotivoitolle, tasapelille ja vierasvoitolle, tasapelin todennäköisyys jaettiin kahdella, ja sekä koti- että vierasvoiton todennäköisyyksistä vähennettiin tasapelintodennäköisyyden puolikas. Näin todennäköisyyksien summa on yksi.

Kaudella 2008-2009 tasapelien todennäköisyydet laskettiin siis kausien 2002-2008 datan perusteella. Siitä seuraaville kausille noihin laskelmiin oli otettu mukaan myös edellisen kauden tulokset. Kauden 2009-2010 laskelmissa myös kauden 2008-2009 data oli siis laskettuna mukaan. Tämä vaikutti jonkin verran laskettuihin tasapelitodennäköisyyksiin. Lisäksi huomioitavaa on, että algoritmi alkaa laskemaan otteluiden todennäköisyyksiä heti ensimmäisestä ottelusta alkaen, jolloin kaikkien joukkueiden Elo-luvut ovat samoja. Lisäksi algoritmi ei erottele toisistaan koti- ja vierasotteluista Elo-luvuissa, eli Elo-luku kuvastaa joukkueen kaikkien otteluiden tuloksia ja sen kokonaisvahvuutta tietyllä hetkellä.

### 3.4.1. Elo-malli 2

Elo-malli 2 vastaa Hvattum & Arntzen (2010) artikkelissa esiteltyä maaliperustaista Elo-mallia suoraan. Sen sijaan, että Elo-lukuja päivitetäisiin vain ennustettujen ja toteutuneiden pistemäärien perusteella, ottelun toteutunut maaliero muodostaa pohjan Elo-lukujen päivitykselle. Muutoin Elo-malli 2 ei eroa Elo-mallia 1:stä.

### 3.4.2. Painotettu Elo-malli

Elo-malleja päädyttiin vielä monipuolistamaan niin, että ne ottavat huomioon kotiedun vähän samalla tavalla kuin Poisson-mallitkin. Tämä kehittyneempi Elo-malli laskee kaikille pelaaville joukkueille sekä koti- että vieras-Elo-luvut. Tämä on varsin pieni muutos algoritmiin ja hoitui muutaman uuden metodin määrittelyllä.

Elo-mallia 1 testatessa heräsi myös ajatus, että olisi ehkä mahdollista saada parempia tuloksia aikaan, jos kauden alussa jokaisen joukkueen Elo-luvut eivät olisi samat. Tästä syystä algoritmiin kehitettiin ominaisuus, niin sanottu painotettu Elo-luku, joka laskee jokaiselle joukkueelle koko edellisen kauden Elo-luvut ja nämä samat Elo-luvut ovat voimassa myös seuraavan kauden alkaessa. Tämä tuntui idea-tasolla varsin pieneltä muutokselta algoritmiin, mutta lopulta sen toteuttaminen oli varsin työläs prosessi. Turhalta koodin toistolta lopulta kuitenkin vältyttiin kun metodi laitettiin kutsumaan edellisen kauden Elo-luvut laskettuaan uudestaan itseään rekursiivisesti. Molemmat uudet ominaisuudet toimivat sekä Elo-malli 1:ssä että 2:ssa.

## 3.5. Pi-hockey

Pi-hockey -malli noudattaa suoraan Constantinoun ja Fentonin (2012) artikkelissa esiteltyä pi-rating -mallia, mutta sovelluskohde on luonnollisesti jalkapallon sijasta jääkiekko. Mallin toteuttamiseksi luotiin PiRatingsController-luokka joka peritään Controller-luokasta sekä sille Model-luokka PiRating. PiRating-luokka huolehtii sekä koti- että vierasotteluiden pi-lukujen päivityksestä. Verrattuna esimerkiksi Elo-malleihin, tämä päivitysprosessi on hiukan monimutkaisempi, koska odotetun maalieron ja maalierovirheen lisäksi täytyy laskea vielä painotettu maalierovirhe, jonka perusteella varsinaiset päivitykset tehdään. Myös odotetun maalieron laskemisen suorittaa PiRating-luokka.

Koska Constantinoun ja Fentonin (2012) artikkelissa ei selkeästi esitetä, kuinka he muodostavat todennäköisyydet pi-lukujen perusteella ottelun eri lopputuloksille, täytyi sitä varten kehittää oma menetelmänsä aivan alusta saakka. Se muistuttaa hiukan Elo-

mallien tasapelin todennäköisyyksien määrittämiseksi kehitettyä menetelmää, mutta on kuitenkin kehittyneempi ja pitemmälle viety versio siitä. Todennäköisyyksien estimointi perustuu arvioon, että kaikki ottelut, joissa ottelun odotettu maaliero on jokin tietty luku, todennäköisyydet kotivoitolle, tasapelille ja vierasvoitolle ovat samat. Näin ollen eri lopputulosten tilastolliset osuudet kaikista otteluista, joissa odotettu maalieroon ollut tämä sama tietty luku, toimivat myös ennustuksina tulevaisuudesta. Esimerkiksi, jos odotettu maaliero on 1,35, lasketaan yhteen kaikki ottelut tietokannasta, joissa odotettu maaliero oli 1,35 ja sen jälkeen lasketaan kuinka moni niistä päättyi kotivoittoon, kuinka moni tasapeliin ja kuinka moni vierasvoittoon. Niistä odotetuista maalieroista, jotka olivat negatiivisia, laskettiin mukaan niiden vastaluvut. Näistä muodostui tavallaan mittari, jossa lähes joka sadasosalle nollan ja 2,55 väliltä löytyi otteluita, joissa tämä luku oli ollut ennakoitu maaliero. Näistä tuloksista muodostettiin kategoriat, joissa yhdistettiin lähellä toisiaan olevia maalierokategorioita suuremmiksi kategorioiksi. Lopulta jäi jäljelle 18 kategoriaa, joiden vaihteluväli oli 0,15 alkaen nolasta ja lopputen 2,55. Viimeinen, kahdeksastoista, kategoria käsitti odotetut maalierot jotka ovat suurempia kuin 2,55.

Kategorioihin ei laskettu kuitenkaan kaikkia otteluita suoraan, koska tällöinhän ennustuksia olisi tehty sellaisilla tiedoilla, joita ei vielä pitäisi olla tiedossa. Koska vedonlyöntitoimistojen kertomisia oli tietokannassa vasta kaudesta 2008-2009 alkaen, laskettiin tuolle kaudelle maalierokategorioihin mukaan kaikki tuota edeltävät kaudet mukaan. Siitä eteenpäin mukaan laskettiin lisäksi aina edellisen pelattavan kauden tulokset, jolloin kaudella 2011-2012 maalierokategorioiden frekvensseissä oli mukana kausien 2002-2011 maalierot ja otteluiden lopputulokset.

Taulukossa 2 on esitettynä kaikki odotettujen maalierojen kategoriat kaudelle 2011-2012, ja näihin kategorioihin laskettujen kotivoittojen, tasapelien ja vierasvoittojen määrät sekä prosenttiosuudet.

Kun huomattiin, että pi-hockey -mallin muodostamien todennäköisyysarvioiden ja vedonlyöntitoimistojen kertomien suhdetta kuvaavat odotusarvot olivat joissain tapauksissa epänormaalin suuria, algoritmia kokeellisesti muokattiin niin, että se huomioi vain tietyntyyppiset odotusarvot. Aluksi kokeiltiin ratkaisua, jossa algoritmi löi vetoa vain otteluista joiden odotusarvo sijoittui välille 1-1,25. Sen jälkeen kokeiltiin lyödä vetoa vain otteluista joiden odotusarvo oli yli 1,25. Tätä ylärajaa nostettiin vielä muutaman pykälän. Lisäksi kokeiltiin vielä laittaa ylärajaksi odotusarvoille 1,6, 2,0 ja 3,0. Millään näillä muutoksilla ei kuitenkaan saatu parempia tuloksia kuin versiolla, jossa odotusarvoja ei mitenkään otettu huomioon tai rajoitettu, joten nämä rajoitukset lopulta poistettiin algoritmista.

Koska pi-rating -malli otti automaattisesti huomioon joukkueiden suoriutumisen sekä koti- että vieraskentillä, tätä erillistä kotietumuuttujaa ei tässäkään tapauksessa tarvinnut pi-hockey algoritmiin kehittää.

### 3.6. Aggregaatti-malli

Aggregaatti-malli on yhdistelmä Poisson 1:stä, perus-Elo-mallista ja pi-hockey -mallista. Se toimii niin, että jokaiseen otteluun kukin mainituista algoritmeista muodostaa ottelun lopputuloksille omat todennäköisyysarvionsa, joista sitten lasketaan keskiarvot, jotka muodostavat todennäköisyydet kotivoitolle, tasapelille ja vierasvoitolle. Muutoin tämäkin konstruktio hyödyntää jo Poisson-malli 1:ä varten kehitettyjä vedonlyöntialgoritmeja.



Aggregaatti-mallin toteuttamiseen johti havainto, että esimerkiksi pi-hockey -mallin todennäköisyydet olivat poikkeuksellisen ja epänormaalinkin suuria suhteessa vedonlyöntitoimistojen arvioihin. Toisaalta nämä arviot tuottivat kuitenkin parhaat tulokset eri konstruktioiden kesken. Kuitenkin myös muut konstruktioet kuin pi-hockey onnistuivat kohtalaisen hyvin arvioissaan vedonlyöntivoitoissa laskettuna, joten yhdistelmä näistä kaikista voisi ehkä onnistua kaikkein parhaiten kaikista konstruktioista.

Ratkaisu käyttää todennäköisyyksien keskiarvoja sen sijaan, että esimerkiksi valittaisiin eri konstruktioiden muodostamista keskiarvoista suurimmat, johtui puhtaasti siitä, että keskiarvoja käyttämällä minkäänlaisia painotuksia ja tasaamisia eri todennäköisyyksien välillä ei tarvitse tehdä koska myös keskiarvotodennäköisyyksien summa on 1. Jos sen sijaan olisi valittu aina suurin mahdollinen todennäköisyys kaikille lopputulosvaihtoehdolle, todennäköisesti olisi jouduttu tekemään tasapainottamisia, koska todennäköisesti usein todennäköisyyksien summa olisi ollut joko suurempi tai pienempi kuin 1. Tasapainottaminen olisi voinut olla hyvin hankala prosessi määrittellä ja sen onnistumisesta ja menestyksestä ei olisi ollut minkäänlaisia takeita. Näin ollen siis päädyttiin helpoimpaan ja varmimpaan ratkaisuun, eli keskiarvoihin.

### 3.7. Graafinen käyttöliittymä

Sovelluksen graafisesta käyttöliittymästä haluttiin tehdä mahdollisimman yksinkertainen, käytettävyydestä kuitenkaan tinkimättä. Alunperin ideana oli tehdä erilliset ikkunat jokaiselle eri mallille tarvittavine valikkoineen. Tarkemman mietinnän jälkeen lopputulos oli kuitenkin, että kaikki tarvittava pystytään helposti sovittamaan yhteen ikkunaan, jos tietyt toiminnot voidaan sulkea pois käytöstä tietyn mallin ollessa valittuna. Jokainen konstruktio on rakennettu ja testattu erillään toisistaan, mutta ne tieteenkin kuitenkin hyödyntävät paljon samaa ohjelmakoodia keskenään. Tässä testaus- ja rakennusvaiheessa esimerkiksi pelikassana käytettiin aina 1000 rahayksikköä. Graafiseen käyttöliittymään kuitenkin tehtiin mahdolliseksi vaihtaa pelikassan kokoa. Vaihtoehdot pelikassan koolle ovat 100, 500, 1000, 1500, 2000 ja 2500 rahayksikköä. Kaikki muut käyttöliittymässä esillä olevat valikot olivat tarjolla jo konstruktioiden testivaiheessa, tosin nämä muutokset tuli yleensä tehdä koodia muokkaamalla.

Käyttöliittymä suunniteltiin tarkoituksella niin, ettei käyttäjän tarvitse, eikä hän myös voi, syöttää mitään tietoja näppäimistöltä. Kaikki valinnat tehdään siis valikoista, joissa vaihtoehdot ovat annettu valmiiksi. Tämä luonnollisesti vähentää valinnan mahdollisuuksia, mutta tämä ei ole tärkeää sovelluksen käytön eikä koko tutkimuksen kannalta.

Graafisen käyttöliittymän ylläladasta valitaan ensin malli jota halutaan käyttää simuloimiseen. Vaihtoehdot ovat Poisson, Pi-hockey, perus-Elo-malli, maaliperustainen Elo-malli sekä aggregaatti-malli. Nämä valinnat tehdään radionappeja käyttämällä, joten vain yksi malli voidaan valita kerralla käytettäväksi. Mallin valinnan alta voidaan määrittää pudotusvalikoita käyttämällä haluttu kausi, Kellyn jakaja, pelikassa, sekä lyödäänkö vetoa kaikista otteluista, vai vain joko kotivoitoista tai vierasvoitoista. Viimeinen tarkoittaa siis, veikataanko vain otteluita, joiden algoritmi arvioi päättyvän kotivoittoon tai otteluista, jotka se arvioi päättyvän vierasvoittoon. Kausi vaihtelee niin, että vain Poisson-mallille voidaan simuloida kaudet 2002-2008. Näiden alla oleva valikko koskee vain Poisson-malleja ja minkä tahansa muun mallin ollessa valittuna tälle valikolle ei voi tehdä mitään, koska sen käyttö on estetty. Tästä valikosta valitaan otetaanko Poisson-mallin maalikeskiarvoja laskiessa huomioon kaikki kauden viime ottelut, vai viimeiset 1, 2, 3, 4, 5 tai 6 ottelua.

**Valitse:**

**Malli**

☒ Poisson ☐ Pi-Hockey

☐ Agregaatti-malli ☐ Elo-malli - perus

☐ Elo-malli - maaliperusteinen

**Kausi** **Kellyn jakaja**

2008-2009 5

**Pelikassa** **Lyödään vetoa**

2000.0 Kaikista ottelui...

**Huomioitavien viimeotteluiden määrä**  
(vain Poisson, 0 = kaikki viime ottelut)

4

**Vain elo-mallit**

☐ Koti- ja vieras-elo-luvut erikseen

☐ Painotettu - edellisen kauden elo-luvut muk...

**Poisson-malli:**

**Kausi: 2002-2003**

Oikein: 425	Väärin: 626	Yhteensä: 1051
Kotivoitoiksi ennustetut:		
Oikein: 269	Väärin: 351	Yhteensä: 620
Vierasvoitoiksi ennustetut:		
Oikein: 156	Väärin: 275	Yhteensä: 431

**Poisson-malli:**

**Kausi: 2003-2004**

Oikein: 425	Väärin: 629	Yhteensä: 1054
Kotivoitoiksi ennustetut:		
Oikein: 281	Väärin: 356	Yhteensä: 637
Vierasvoitoiksi ennustetut:		
Oikein: 144	Väärin: 273	Yhteensä: 417

**Poisson-malli:**

**Kausi: 2005-2006**

Oikein: 433	Väärin: 632	Yhteensä: 1065
Kotivoitoiksi ennustetut:		
Oikein: 288	Väärin: 347	Yhteensä: 635
Vierasvoitoiksi ennustetut:		
Oikein: 145	Väärin: 285	Yhteensä: 430

**Poisson-malli:**

**Kausi: 2007-2008**

Oikein: 408	Väärin: 650	Yhteensä: 1058
Kotivoitoiksi ennustetut:		
Oikein: 250	Väärin: 366	Yhteensä: 616
Vierasvoitoiksi ennustetut:		
Oikein: 158	Väärin: 284	Yhteensä: 442

**Lopeta** **Simuloi**

**Tyhjennä tekstikenttä**

**Kuva 1.** Sovelluksen graafinen käyttöliittymä.

Poisson-mallia koskevan valikon alapuolella on kaksi valintaruutua, jotka ovat käytettävissä vain kun jompi kumpi Elo-malleista on valittuna. Näillä valikoilla määritetään, laskeeko algoritmi erikseen sekä koti- että vieras-Elo-luvut joukkueille ja käytetäänkö painotettuja Elo-lukuja, eli Elo-lukuja, joihin kauden alussa on laskettu mukaan edellisen koko kauden aikana kertyneet Elo-luvut. Näiden valintaruutujen alapuolella on kaksi tavallista nappia, joista toisella lopetetaan ohjelman käyttö, ja toisella käynnistetään itse simulointi. Näiden kaikkein valikoiden vieressä on iso tekstialue, jonne saadut tulokset aina tulostetaan. Tekstialueen alapuolella on nappi, jolla tekstialue voidaan tyhjentää kaikesta tekstistä. Kuvassa 1 on kuvattuna koko graafinen käyttöliittymä tilanteessa, jolloin joitain saatuja tuloksia on tulostettu tekstialueelle.

### 3.8. Yhteenvedon algoritmin kehitysprosessista

Yllä on käyty läpi koko algoritmin kehitysprosessi alusta alkaen lopulliseen toimivaan sovellukseen, jolla on graafinen käyttöliittymä. Algoritmin kehitys lähti liikkeelle perusluokkarakenteen toteuttamisesta joka noudattelee aika pitkälti model-view-controller -mallia. Tämän jälkeen kehitettiin ensimmäinen varsinainen algoritmi, joka käytti hyväkseen mahdollisimman yksinkertaista Poissonin jakaumaa. Poisson-mallia kehitettiin tämän jälkeen älykkäämmäksi ja monipuolisemmaksi lisäämällä mahdollisuus rajoittaa Poisson-mallille syötettyyn maalikeskiarvoon laskettavien viimeisten otteluiden määrää. Jotta Poisson-mallin todellista toimivuutta otteluiden lopputulosten ennustamisessa pystyttiin mittaamaan, algoritmiin kehitettiin myös vedonlyönti-osa. Tämän osan tehtävänä oli laskea, oliko ottelussa tarjolla ylikerroin ja jos oli, mille lopputulokselle. Seuraavaksi vedonlyönti-osaa algoritmista kehitettiin niin, että se osaa suositella vain tiettyyn lopputulokseen loppuvaksi ennustettuja otteluita.

Seuraavaksi alettiin kehittää Elo-mallien algoritmeja. Aluksi kehitettiin Hvattumin ja Arntzenin (2010) artikkelissa esitetty perus-Elo-malli, joka noudattaa täysin kyseissä artikkelissa esitettyä Elo-mallia. Perus-Elo-mallin antamia ennakoituja pistemääriä seuraavaan otteluun käytettiin suoraan todennäköisyysarvioina koti- ja vierasvoitolle, mutta tasapelin todennäköisyyden määrittämiseksi tarvittiin täydennystä algoritmiin. Tätä varten tarvittiin myös erillinen tietokantataulu drawcategories, joihin eri tasapelikategorioiden todennäköisyydet tallennettiin. perus-Elo-mallin jälkeen kehitettiin Hvattumin & Arntzenin (2010) artikkelissa esitelty maaliperustainen Elo-malli. Tämä ei algoritmin kannalta ollut olennainen tai suuri lisäys, koska vain Elo-lukujen päivitykseen tarvitsi tehdä muutoksia.

Elo-mallien jälkeen kehitysvuorossa oli Constantinoun ja Fentonin (2012) tutkimuksessa esitetty pi-rating -malli, joka sai tässä paperissa nimekseen pi-hockey. Pi-hockey muodosti jälleen tavallaan oman haaransa kehitykselle, koska suurin osa koodista tätä mallia varten täytyi kehittää alusta alkaen. Vain runko koodille saatiin aiemmista malleista. Pi-hockey -malli muistuttaa suuresti Elo-malleja idealtaan, mutta on kuitenkin niitä jonkin verran monimutkaisempi.

Aluksi kehitettiin algoritmi aivan perusmallille. Tämän jälkeen havaittiin tarve mallille, jossa pi-luvut eivät olisi samat joka joukkueelle kauden alussa, vaan ne kuvaisivat paremmin joukkueiden todellisia tasoeroja. Tämä toteutettiin niin, että laskettiin ensin edellisen kauden pi-luvut kaikille joukkueille ja nämä toimivat lähtöarvoina seuraavalle kaudelle. Edellisen kauden simulointi ennen varsinaisen kauden simulointia osoittautui ehkä haastavimmaksi osuudeksi koko ohjelmointityössä tätä tutkimusta varten. Tähän oli syynä lähinnä se, ettei alkuperäinen pi-hockey -mallin koodi ollut mitenkään huomioinut tällaista vaihtoehtoa, mutta toisaalta oli liian riskialtista alkaa tätä koodia liikaa muuttamaankaan. Lopputuloksena oli pätkä ohjelmakoodia, joka ei ole kovin eleganttia, mutta kuitenkin hoitaa tehtävänsä.

Viimeinen kehitettävä malli oli yhdistelmä Poisson-mallista, perus-Elo-mallista sekä pi-hockey mallista. Tämäkin toimi omana haaranaan kehityksessä, mutta pystyi käyttämään todella paljon jo olemassa olevaa koodia hyväkseen. Tämän osan ohjelmaa tarvitsi oikeastaan vain hakea tarvittavat tiedot ensin tietokannasta ja sen jälkeen syöttää nämä tiedot jo olemassa olevia eri malleja ja niiden ohjelmakoodia. Ainoa uusi asia oli todennäköisyydet lopputuloksille laskeva palanen, joka laski keskiarvot eri lopputuloksille kaikkein eri mallien todennäköisyysarvioista.

Lopuksi kehitettiin vielä graafinen käyttöliittymä hallitsemaan kaikkia malleille ja helpottamaan niiden käyttöä. Käyttöliittymä koostuu vain yhdestä ikkunasta, josta eri malleille voidaan valikkoja käyttämällä antaa kaikki niiden tarvitsema tieto. Osa käyttöliittymän valikoista on käytettävissä vain tietyille malleille, muulloin ne eivät ole käytössä. Käyttöliittymä myös tulostaa algoritmien saamat tulokset käyttöliittymäikkunassa olevaan tekstialueeseen.

## 4. Tulokset

Seuraavaksi käsitellään tuloksia, jotka on saatu käyttämälle Poisson-mallia, perus-Elo-mallia, maaliperustaista Elo-mallia, pi-hockey -mallia sekä aggregaatti-mallia. Aluksi tuloksia käydään läpi vain siitä näkökulmasta, onnistuiko algoritmi ennustamaan ottelun lopputuloksen oikein vai ei. Seuraavassa kappaleessa käsitellään tuloksia vedonlyönti-simulaation näkökulmasta.

## 4.1. Poisson-mallit

Algoritmin ensimmäinen versio käyttää hyväkseen yksiulotteista Poissonin jakaumaa. Parametrinä käytetään kotijoukkueelle sen aiemmissa kotiotteluissa tekemien maalien keskiarvoa, ja vierasjoukkueelle sen aiemmissa vierasotteluissa tekemien maalien keskiarvoa. Algoritmin laskemia todennäköisyyksiä verrattiin kyseisen ottelun toteutuneeseen lopputulokseen. Simulointi tehtiin kaikkiin NHL-otteluihin kaudesta 2002-2003 kauteen 2011-2012 asti lukuun ottamatta kautta 2004-2005, jota ei pelattu ollenkaan. Sellaisia otteluita, joissa molemmille joukkueille ei voida laskea vähintään kolmen aiemman ottelun tehdyistä maaleista muodostuvaa maalikeskiarvoa, ei otettu huomioon.

**Taulukko 2.** Poisson-mallin simulointitulokset.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Otteluita yht.
2002-2003	453	635	41.64%	58.36%	1088
2003-2004	473	617	43.39%	56.61%	1090
2005-2006	481	612	44.01%	55.99%	1093
2006-2007	466	619	42.95%	57.05%	1085
2007-2008	438	653	40.15%	59.85%	1091
2008-2009	474	614	43.57%	56.43%	1088
2009-2010	462	628	42.39%	57.61%	1090
2010-2011	445	649	40.68%	59.32%	1094
2011-2012	448	638	41.25%	58.75%	1086
<b>Yhteensä:</b>	4140	5665	42.22%	57.78%	9805

Taulukossa 2 on esitetty algoritmin ensimmäisen version antamat ottelutulosten ennustukset yhteenlaskettuna kausittain. Kuten sarakkeesta "Oikein" voidaan havaita, algoritmi ei pysty vielä ennustamaan otteluiden lopputuloksia riittävällä tarkkuudella. Prosentuaalisesti algoritmi onnistui ennustamaan oikein huonoimmillaan 40,15% otteluista (kausi 2007-2008) ja parhaimmillaan 43,92% otteluista (kausi 2005-2006). Keskimäärin algoritmi onnistuu ennustamaan oikein 42,12% otteluista. Huomion arvoista on kuitenkin, ettei algoritmin oikein ennustamien otteluiden määrissä esiinny kovinkaan suuria eroja kausien välillä, vaan tulokset ovat melko yhdenmukaisia läpi kausien.

Toinen huomionarvoinen, joskin valitettava seikka on, että algoritmin onnistumisprosentti on kohtalaisen hyvä. Jos se olisi huomattavan pieni, esimerkiksi 20%, voitaisiin asetelmää kääntää ikään kuin ylösalaisin. Tämäkin on ongelmallista siinä mielessä, että kun pyritään ennustamaan oikea vaihtoehto kolmesta vaihtoehdosta, vaikkakin algoritmi onnistuisi erottamaan kaksi väärää vaihtoehtoa 80% onnistumistodennäköisyydellä, siltikin vielä pitäisi yrittää jotenkin ennustaa kumpi väärästä lopulta olisi todennäköisimmin se oikea vaihtoehto.

Kun algoritmin antamia todennäköisyyksiä otteluiden lopputuloksille tarkasteltiin syvemmin, mielenkiintoisia asioita tuli esille. Jokaisen kauden yksittäinen ottelu arvioitiin sen mukaan ennustiko algoritmi sen päättyvän koti- vai vierasvoittoon ja jaettiin tällä perusteella kahteen kategoriaan. Tämän jälkeen laskettiin, kuinka usein algoritmi ennusti oikein sellaiset ottelut, jotka se ennusti päättyvän kotivoittoon, ja

kuinka usen sellaiset, jotka se ennusti päättyvän vierasvoittoon. Tulokset näkyvät taulukoissa 3 ja 4.

**Taulukko 3.** Ottelut, joiden Poisson-malli ennusti päättyvän kotivoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	302	367	45.14%	54.86%	669
2003-2004	355	429	45.28%	54.72%	784
2005-2006	339	366	48.09%	51.91%	705
2006-2007	334	389	46.20%	53.80%	723
2007-2008	309	429	41.87%	52.04%	738
2008-2009	365	396	47.96%	54.20%	761
2009-2010	360	426	45.80%	57.53%	786
2010-2011	316	428	42.47%	56.12%	744
2011-2012	326	417	43.88%	54.82%	743
<b>Yhteensä:</b>	3006	3647	45.18%	54.82%	6653

Kuten taulukoista 3 ja 4 voidaan huomata, kun algoritmi ennusti, että ottelu päättyy vierasvoittoon, osuivat nämä ennustukset oikeaan parhaimmillaan vain vajaa 39% tapauksista (kausi 2003-2004). Heikoimmin algoritmi onnistui ennustuksissaan kaudella 2008-2009, jolloin vierasvoitoiksi ennustetuista ainoastaan 33,33% osui oikeaan. Keskimäärin algoritmin vierasvoittoon ennustamista otteluista kaikilla analysoidyilla kausilla oikeaan osui 36,95%. Tämän perusteella siis algoritmi onnistuu ennustamaan kotivoittoja paremmin kuin vierasvoittoja. Ero tulosten välillä on keskimäärin noin yhdeksän prosenttiyksikköä.

**Taulukko 4.** Ottelut, joiden Poisson-malli ennusti päättyvän vierasvoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	151	268	36.04%	63.96%	419
2003-2004	118	188	38.56%	61.44%	306
2005-2006	142	246	36.60%	63.40%	388
2006-2007	132	230	36.46%	63.54%	362
2007-2008	129	224	36.54%	63.46%	353
2008-2009	109	218	33.33%	66.67%	327
2009-2010	102	202	33.55%	66.45%	304
2010-2011	129	221	36.86%	63.14%	350
2011-2012	122	221	35.57%	64.43%	343
<b>Yhteensä:</b>	1134	2018	35.98%	64.02%	3152

Otteluiden, joiden algoritmi ennusti päättyvän kotivoittoon, onnistumisprosentit ovat hiukan parempia kuin kokonaisuonnistumisprosentit, mutta ne ovat todennäköisesti kuitenkin liian huonoja vedonlyönnissä hyväksikäytettäväksi.

**Taulukko 5.** Poisson-mallin tulokset, kun vain 3 viimeistä ottelua on otettu huomioon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	427	661	39.25%	60.75%	1088
2003-2004	445	645	40.83%	59.17%	1090
2005-2006	445	648	40.71%	59.29%	1093
2006-2007	425	658	39.24%	60.76%	1083
2007-2008	432	659	39.60%	60.40%	1091
2008-2009	459	629	42.19%	57.81%	1088
2009-2010	461	629	42.29%	57.71%	1090
2010-2011	435	658	39.80%	60.20%	1093
2011-2012	422	664	38.86%	61.14%	1086
<b>Yhteensä:</b>	3951	5851	40.31%	59.69%	9802

On myös huomattavaa, ettei algoritmi antanut yhdessäkään ottelussa suurinta todennäköisyyttä tasapelille. Analysoiduista otteluista, eli niistä, joissa sekä kotijoukkueella oli ennen kyseistä ottelua vähintään kolme kotiottelua pelattuna, ja vierasjoukkueella vähintään 3 vierasottelua pelattuna, keskimäärin 23,9 % päättyi tasapeliin. Kaikista tietokannan otteluista 23,7% päättyi tasapeliin. Tämä on selkeä rajoite algoritmissä, mutta toisaalta myös potentiaalinen parannuskohta, jolla oikeaan osuneiden ennustusten määrää voidaan parantaa.

**Taulukko 6.** Poisson-mallin tulokset, kun vain 4 viimeistä ottelua on otettu huomioon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	425	626	40.44	59.56	1051
2003-2004	425	629	40.32	59.68	1054
2005-2006	433	632	40.66	59.34	1065
2006-2007	408	650	38.56	61.44	1058
2007-2008	452	608	42.64	57.36	1060
2008-2009	442	615	41.82	58.18	1057
2009-2010	417	642	39.38	60.62	1059
2010-2011	417	642	39.38	60.62	1059
2011-2012	419	637	39.68	60.32	1056
<b>Yhteensä:</b>	3838	5681	40.32	59.68	9519

Taulukoissa 5-8 on tulokset simuloinneista Poisson-mallia käyttäen, joissa otettiin huomioon 3-6 viimeistä ottelua Poisson-funktiolle syötettäviä maalikeskiarvoja laskettaessa. Kun näitä tuloksia verrataan taulukossa 2 esitettyihin tuloksiin, huomataan, että Poisson-malli, joka ottaa koko kauden siihen asti pelatut ottelut huomioon maalikeskiarvoissa, pärjää paremmin verrattuna malleihin, jotka ottavat vain 3-6 viime ottelua huomioon. Tämä on oikeastaan hiukan intuition vastaista, koska kun otetaan huomioon vain muutama viime peli maalikeskiarvoa laskettaessa, tämän pitäisi huomioida joukkueen tämän hetkinen suorituskunto paremmin, kuin jos koko kauden siihen astiset pelatut pelit otetaan huomioon.

**Taulukko 7.** Poisson-mallin tulokset, kun vain 5 viimeisintä ottelua on otettu huomioon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	427	595	41.78%	58.22%	1022
2003-2004	421	603	41.11%	58.89%	1024
2005-2006	413	617	40.10%	59.90%	1030
2006-2007	405	625	39.32%	60.68%	1030
2007-2008	384	643	37.39%	62.61%	1027
2008-2009	441	586	42.94%	57.06%	1027
2009-2010	435	592	42.36%	57.64%	1027
2010-2011	423	605	41.15%	58.85%	1028
2011-2012	408	618	39.77%	60.23%	1026
<b>Yhteensä:</b>	3757	5484	40.66%	59.34%	9241

**Taulukko 8.** Poisson-mallin tulokset, kun vain 6 viimeisintä ottelua on otettu huomioon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2002-2003	413	576	41.76%	58.24%	989
2003-2004	423	573	42.47%	57.53%	996
2005-2006	418	578	41.97%	58.03%	996
2006-2007	404	592	40.56%	59.44%	996
2007-2008	357	639	35.84%	64.16%	996
2008-2009	440	556	44.18%	55.82%	996
2009-2010	420	575	42.21%	57.79%	995
2010-2011	408	594	40.72%	59.28%	1002
2011-2012	394	301	39.60%	60.40%	995
<b>Yhteensä:</b>	3677	5284	41.03%	58.97%	8961

Vaikuttaisi siis siltä, että mitä useamman pelin maaliluvut keskiarvoon on laskettu mukaan, sitä paremmin osuvia Poissonin jakauman antamia todennäköisyysarvioita ja myös parempia ennustuksia saadaan. Tätä tukee myös huomio, että oikein menneiden otteluiden keskiarvo nousee taulukoissa 5-8 mitä useampi ottelu huomioidaan keskiarvoihin. Huomionarvoista on, että tässäkin mallissa kotijoukkueelle on huomioitu vain sen pelaamissa kotiotteluissa tekemät maalit ja vierasjoukkueelle vain sen vierasotteluissa tekemät maalit.

## 4.2. Elo-malli 1

Kolmas konstruktio hyödyntää Hvattumin ja Arntzenin (2010) tutkimuksessa esiteltyä Elo-lukuihin perustuvaa mallia. Mallin ennustamiskykyä ei testattu kaikilla tietokannasta löytyvillä kausilla, vain neljällä viimeisimmällä, eli kausilla 2008-2012. Kausia 2002-2008 käytettiin hyväksi tasapelitodennäköisyyksien kategorioiden muodostamiseksi.

Kuten taulukosta 9 huomataan, tällä menetelmällä tuotetut ennusteet ovat vain hiukan parempia kuin toisella konstruktiolla tuotetut ennusteet, ja aika samanlaisia kuin ensimmäisen konstruktion saavuttamat tulokset.

**Taulukko 9.** Perus-Elo-mallin tulokset.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	525	705	42.68%	57.32%	1230
2009-2010	498	732	40.49%	59.51%	1230
2010-2011	519	711	42.20%	57.80%	1230
2011-2012	502	728	40.81%	59.19%	1230
<b>Yhteensä:</b>	2044	2876	41.54%	58.46%	4920

Parhaiten algoritmi onnistui kaudella 2008-2009 jolloin onnistumisprosentti oli reilu 42%. Heikoin tulos oli kaudella 2009-2010, jolloin oikea lopputulos onnistuttiin ennustamaan vain 40% tapauksista. Keskimäärin algoritmi onnistui ennustamaan 41,1,% otteluista oikein. Huomattavaa on, että algoritmi ei antanut Elo-lukujen kehittyä kauden alussa esimerkiksi kolmea kierrosta, vaan ennustukset tehdään aivan kauden ensimmäisistä otteluista alkaen. Tämä varmasti vaikuttaa hiukan tulosten luotettavuuteen ja onnistumisprosentteihin.

**Taulukko 10.** Ottelut, joiden perus-Elo-malli ennusti päättyvän kotivoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	308	317	49.28%	50.72%	625
2009-2010	283	321	46.85%	53.15%	604
2010-2011	275	350	44.00%	56.00%	625
2011-2012	272	327	45.41%	54.59%	599
<b>Yhteensä:</b>	1138	1315	46.39%	53.61%	2453

**Taulukko 11.** Ottelut, joiden perus-Elo-malli ennusti päättyvän vierasvoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	217	388	35.87%	64.13%	605
2009-2010	215	411	34.35%	65.65%	626
2010-2011	244	361	40.33%	59.67%	605
2011-2012	230	401	36.45%	63.55%	631
<b>Yhteensä:</b>	906	1561	36.72%	63.28%	2467

Kun otteluita tarkastellaan niin, että ottelut jaetaan kahteen osaan sen perusteella, ennustiko algoritmi ottelun päättyvän kotivoittoon vai vierasvoittoon, saadaan myös aika mielenkiintoisia tuloksia. Ne on esitetty taulukoissa 10 ja 11. Voidaan huomata, että algoritmi onnistuu ennustamaan kotivoitoiksi ennustamiaan otteluita selkeästi paremmin oikein kuin vierasvoitoiksi ennustamiaan otteluita. Kotivoitoiksi ennustetuista oikein



menee keskimäärin noin 46,39% parhaan onnistumisprosentin ollessa lähes puolet, kaudella 2008-2009. Vierasvoitoiksi ennustetuista algoritmi sen sijaan onnistuu ennustamaan oikein keskimäärin vain noin 36,70% parhaan tuloksen ollessa 40,33% kaudella 2010-2011. Vaikka tällaiset tulokset eivät vaikuta kovin lupaavilta vedonlyönnin kannalta, kun pelataan aina ylikertoimia, on mahdollista, että voitokas pelistrategia pystytään luomaan.

Seuraavaksi tutkittiin, kuinka hyvin Elo-malli 1 pystyy ennustamaan otteluiden lopputuloksia, kun kaikille joukkueille lasketaan omat Elo-luvut sekä koti- että vierasotteluille. Ennustamiseen käytettiin jokaisessa ottelussa kotiotteluiden Elo-lukua kotijoukkueelle ja vierasotteluiden Elo-lukua vierasjoukkueelle. Tulokset on esitetty taulukossa 12. Tällä menetelmällä keskimääräinen onnistumisprosentti kasvaa hiukan, mutta ero perus-Elo-malliin, joka laskee vain yhden Elo-luvun kaikille joukkueille, on käytännössä merkityksetön. Kausien 2010-2011 ja 2011-2012 onnistumisprosentit ovat käytännössä samat, mutta mielenkiintoista on, että esimerkiksi kauden 2008-2009 onnistumisprosentti on noin neljä prosenttiyksikköä parempi kuin esimerkiksi kauden 2010-2011 vastaava. Tämän eron syytä on kuitenkin varsin vaikea tutkia ja selvittää.

**Taulukko 12.** Perus-Elo-malli – Koti- ja vierasotteluiden Elo-luvut laskettu erikseen.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	544	686	44.23%	55.77%	1230
2009-2010	522	708	42.44%	57.56%	1230
2010-2011	495	735	40.25%	59.76%	1230
2011-2012	496	734	40.34%	59.67%	1230
<b>Yhteensä:</b>	2057	2863	41.81%	58.19%	4920

**Taulukko 13.** Painotettu perus-Elo-malli.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	543	687	44.15%	55.85%	1230
2009-2010	515	715	41.87%	58.13%	1230
2010-2011	542	688	44.07%	55.93%	1230
2011-2012	527	703	42.85%	57.15%	1230
<b>Yhteensä:</b>	2127	2793	43.23%	56.77%	4920

Perus-Elo-mallia testattiin vielä käyttämällä painotettuja Elo-lukuja. Tämä tarkoittaa, että koko edellinen kausi simuloitiin aina ennakkoon ja tämän kauden Elo-luvut toimivat seuraavan kauden alussa Elo-lukuina joukkueille. Jokaiselle joukkueelle laskettiin vain yksi kattava Elo-luku, ei erikseen Elo-lukuja sekä koti- että vierasotteluille. Tulokset on esitetty taulukossa 13. Kuten voidaan huomata, painotettu perus-Elo-malli on toistaiseksi paras testattu malli tässä tutkimuksessa. Mikään muista malleista ei ole kyennyt keskimäärin yli 43% onnistumisosuuteen kaikista ennustetuista otteluista. Tämän perustella voitaisiin olettaa, että edellisen kauden suorituksella tai joukkueen tasolla näyttäisi olevan jonkinlainen yhteys seuraavan kauden suorituksiin, vaikka joukkueiden kokoonpanot ja valmentajatkin vaihtuvat usein kausien välillä.

### 4.3. Elo-malli 2

Seuraavaksi käsitellään Elo-malli 2 eli maaliperustaisen Elo-mallin suorituskykyä ja otteluiden lopputulosten ennustamiskykyä. Elo-malli 2:n ero perus-Elo-malliin eli Elo-malli 1:een on, että Elo-malli 2:ssa Elo-lukujen päivitysperusteena käytetään otteluiden toteutunutta maalieroa, ei odotettujen pisteiden ja toteutuneiden pisteiden erotusta. Aluksi testattiin maaliperustaista Elo-mallia ilman painotuksia tai erillisiä koti- ja vieras-Elo-lukuja. Saadut tulokset on esitetty taulukossa 14.

**Taulukko 14.** Maaliperustainen Elo-malli.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	548	682	44.55%	55.45%	1230
2009-2010	520	710	42.28%	57.72%	1230
2010-2011	496	734	40.33%	59.67%	1230
2011-2012	494	736	40.16%	59.84%	1230
<b>Yhteensä:</b>	2058	2862	41.83%	58.17%	4920

**Taulukko 15.** Maaliperustainen Elo-malli: ottelut, jotka algoritmi ennusti kotivoitoiksi.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	326	313	51.02%	48.98%	639
2009-2010	303	316	48.95%	51.05%	619
2010-2011	264	329	44.52%	55.48%	593
2011-2012	275	340	44.72%	55.28%	615
<b>Yhteensä:</b>	1168	1298	47.36%	52.64%	2466

Keskimääräinen ero onnistumisprosentilla perus-Elo-mallin ja maaliperustaisen Elo-mallin välillä on vain 0,3 prosenttiyksikköä maaliperustaisen Elo-mallin eduksi. Ero mallien välillä on siis merkityksetön. Seuraavaksi tutkittiin kuinka hyvin maaliperustainen Elo-malli toimii, kun tulokset jaetaan kahteen osaan sen perusteella ennustiko algoritmi otteluiden päättyvän koti- vai vierasvoittoon. Nämä tulokset on esitetty taulukoissa 15 ja 16. Kuten kaikki ottelut mukaan luettunakin, maaliperustainen Elo-malli onnistuu ennustuksissa paremmin kun huomioidaan vain ottelut, joiden algoritmi ennusti päättyvän kotivoittoon, verrattuna perus-Elo-malliin. Ja kuten aiempienkin mallien kohdalla, otteluissa, joiden algoritmi ennusti päättyvän vierasvoittoon, onnistumisprosentti on huomattavasti heikompi kuin otteluissa, jotka algoritmi ennusti päättyvän kotivoittoon. Vierasvoittojakin ennustaessaan maaliperustainen algoritmi toimii kuitenkin jonkin verran menestyksekkäämmin kuin perus-Elo-malli.

Tämän jälkeen tutkittiin, kuinka erillisten Elo-lukujen käyttö koti- ja vierasotteluissa muuttaa maaliperustaisen Elo-mallin onnistumisprosenttia. Kuten taulukon 17. tuloksista havaitaan, ero maaliperustaiseen Elo-malliin, joka laskee Elo-luvut kaikkien otteluiden perusteella, ero on erilliset Elo-luvut laskevan mallin eduksi noin puoli prosenttiyksikköä. Mielenkiintoista on, että vaihtelu parhaan (kausi 2008-2009) ja

huonoimman kauden (kausi 2010-2011) välillä on yli neljä prosenttiyksikköä. Vastaava vaihteluväli oli myös perus-Elo-mallissa, kun sekä koti- että vierasotteluille arvioitiin omat Elo-luvut.

**Taulukko 16.** Maaliperustainen Elo-malli: ottelut, jotka algoritmi ennusti vierasvoitoiksi.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	222	369	37.56	62.44	591
2009-2010	217	394	35.52	64.48	611
2010-2011	232	352	39.73	60.27	584
2011-2012	219	396	35.61	64.39	615
<b>Yhteensä:</b>	890	1511	37.07	62.93	2401

**Taulukko 17.** Maaliperustainen Elo-malli: erilliset Elo-luvut koti- ja vierasotteluille.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	547	683	44.47%	55.53%	1230
2009-2010	530	700	43.09%	56.91%	1230
2010-2011	495	735	40.24%	59.76%	1230
2011-2012	511	719	41.54%	58.46%	1230
<b>Yhteensä:</b>	2083	2837	42.34%	57.66%	4920

**Taulukko 18.** Maaliperustainen Elo-malli: painotetut Elo-luvut.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	550	680	44.72%	55.28%	1230
2009-2010	519	711	42.20%	57.80%	1230
2010-2011	536	694	43.58%	56.42%	1230
2011-2012	531	699	43.17%	56.83%	1230
<b>Yhteensä:</b>	2136	2784	43.41%	56.59%	4920

Lopuksi oli vuorossa maaliperustaisen Elo-mallin testaaminen painotetuilla Elo-luvuilla. Tuloksista taulukosta 18. huomataan, että painotettujen Elo-lukujen käyttö parantaa onnistumisprosenttia maaliperustaisella Elo-mallilla kuten kävi myös perus-Elo-mallin kanssa. Ero painotetun mallin eduksi suhteessa tavallisiin maaliperustaisiin Elo-lukuihin on noin 1,6 prosenttiyksikköä. Samoin kuin painotetussa perus-Elo-mallissa, algoritmi onnistuu parhaiten kaudella 2008-2009, jolloin onnistumisprosentti on 44,72%. Myös huonoin kausi on molemmilla malleilla sama, eli kausi 2009-2010.

#### 4.4. Pi-hockey -malli

Pi-hockey -malli on toteuttu suoraan Constantinoun ja Fentonin (2012) artikkelissa esitettyä pi-rating -mallia imitoiden. Artikkelissa ei kuitenkaan tuoda esille, kuinka pi-luvuista saadaan johdettua todennäköisyysarviot ottelun eri lopputuloksille, joten tähän on kehitetty oma menetelmä, joka selitettiin tarkemmin jo aiemmin. Taulukossa 19 on esitetty pi-hockey -mallin antamat tulokset, kun kaikki ottelut huomioidaan. Pi-hockey lähtee oletuksesta, että kaikki joukkueet ovat samantasoisia kauden alussa.

Kun pi-hockey -mallin saamia tuloksia verrataan muihin malleihin aivan perusmuodossaan, eli ilman painotuksia tai muita vastaavia, on sen saavuttama keskimääräinen onnistumisprosentti paras. Kun Poisson-mallista huomioidaan kaikki kaudet 2002-2012, on sen saavuttama onnistumiskeskisarvo 42,22%, eli vain 0,03 prosenttiyksikköä huonompi kuin pi-hockey -mallin vastaava. Mutta jos vertailukelpoisuuden parantamiseksi Poisson-mallistakin huomioidaan vain kaudet 2008-2012, saadaan sen onnistumisprosentiksi 41,97%, jolloin sen onnistumisprosentti on edelleen toiseksi paras malleista, mutta kuitenkin 0,28 prosenttiyksikköä heikompi.

**Taulukko 19.** Pi-hockey: kaikki ottelut.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	524	706	42.60%	57.40%	1230
2009-2010	526	704	42.76%	57.24%	1230
2010-2011	499	731	40.57%	59.43%	1230
2011-2012	531	699	43.17%	56.83%	1230
<b>Yhteensä:</b>	2080	2840	42.25%	57.75%	4881

Kuten muidenkin mallien kohdalla, seuraavaksi pi-hockey -mallin tuloksia tutkittiin niin, että laskettiin algoritmin kotivoitoiksi ennustamista otteluista oikein ja väärin menneiden otteluiden osuudet ja sen jälkeen samat osuudet myös algoritmin vierasvoitoiksi ennustamista otteluista. Nämä on esitetty taulukoissa 20 ja 21.

**Taulukko 20.** Pi-hockey: ottelut, joiden algoritmi ennusti päättyvän kotivoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	454	556	44.95	55.05	1010
2009-2010	463	570	44.82	55.18	1033
2010-2011	488	705	40.91	59.09	1193
2011-2012	531	699	43.17	56.83	1230
<b>Yhteensä:</b>	1936	2530	43.35	56.65	4466

Myös pi-hockey -malli onnistuu ennustamaan kotivoittoja paremmin kuin vierasvoittoja. Huomattavaa on kuitenkin, että pi-hockey ennustaa suurimman osan otteluista päättyvän kotivoittoon. Kaudella 2011-2012 se ennustaa kauden joka ottelun päättyvän kotivoittoon, mikä on hyvin poikkeuksellista. Tässä piilee myös selkeä parannuskohta koko sovellukselle tulevaisuutta varten. Esimerkiksi tätä sovellusta varten kootun tietokannan otteluista vajaa 43% päättyi kotivoittoon. Näin ollen pi-

hockeyn kotivoitoiksi ennustamien otteluiden määrä on todella epärealistinen. Tästä huolimatta pi-hockey kuitenkin siis on tähän mennessä tässä tutkimuksissa testatuista malleista menestyksekkäin ennustuksissaan.

**Taulukko 21.** Pi-hockey: ottelut, joiden algoritmi ennusti päättyvän vierasvoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	70	122	36.46%	63.55%	192
2009-2010	63	58	52.08%	47.93%	121
2010-2011	11	0	100.00%	0.00%	11
2011-2012	0	0	0.00%	0.00%	0
<b>Yhteensä:</b>	144	180	44.44%	55.56%	324

On hyvin mielenkiintoista, että esimerkiksi kaudella 2010-2011, niistä otteluista jotka pi-hockey -malli ennustaa vierasvoitoiksi, kaikki 11 otteluista menevät oikein. Samoin kaudella 2009-2010, yli puolet vierasvoitoiksi ennustetuista otteluista menee oikein. Kaudella 2008-2009 kuitenkin vain noin 36,5% vierasvoitoiksi ennustetuista osuu oikeaan.

#### 4.5. Aggregaatti-malli

Aggregaatti-malli on yksinkertaisesti yhdistelmä Poisson-mallista, perus-Elo-mallista sekä pi-hockey -mallista. Jokainen näistä laskee ensin omat todennäköisyysarvioinsa ottelun lopputuloksille, ja sen jälkeen varsinainen todennäköisyysarvio muodostetaan laskemalla keskiarvot kaikkien kolmen eri algoritmin todennäköisyysarvioista joka lopputulosvaihtoehdolle.

**Taulukko 22.** Aggregaatti-malli: kaikki ottelut.

<b>Taulukko 23: Aggregaatti-malli: kaikki ottelut</b>					
Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	544	686	44.23	55.77	1230
2009-2010	526	704	42.76	57.24	1230
2010-2011	526	704	42.76	57.24	1230
2011-2012	518	712	42.11	57.89	1230
<b>Yhteensä:</b>	2114	2806	42.97	57.03	4920

Kuten taulukosta 22 huomataan, aggregaatti-mallin keskimääräinen onnistumisprosentti on paras kaikista kehitetyistä malleista, kun huomioon otetaan jokaisesta tämän tutkimuksen mallista aivan perusversio. Aggregaatti-mallin voittomarginaali toiseksi parhaaseen, eli pi-hockey -malliin verrattuna on 0,72 prosenttiyksikköä. Lukuun ottamatta kauden 2008-2009, aggregaatti-mallin onnistumisprosentti on sangen stabiili, kahdella kaudella jopa täysin sama. Tämä seikka on mallin ennustettavuuden kannalta tärkeää.

**Taulukko 23.** Aggregaatti-malli: ottelut, joiden algoritmi ennusti päättyvän kotivoittoon.

Kausi	Oikein	Väärin	Oikein %	Väärin %	Ottelut yht.
2008-2009	422	464	47.63%	52.37%	886
2009-2010	420	503	45.50%	54.50%	923
2010-2011	403	536	42.92%	57.08%	939
2011-2012	416	528	44.07%	55.93%	944
<b>Yhteensä:</b>	1661	2031	44.99%	55.01%	3692

Kuten aiempienkin mallien kanssa, seuraavaksi tarkasteltiin algoritmin kotivoitoiksi ennustamien otteluiden onnistumisprosentteja ja sen jälkeen vierasotteluiksi ennustettujen vastaavia lukuja. Ne on esitety taulukoissa 23 ja 24. Aggregaatti-malli ei ole paras malli, kun otetaan huomioon vain kotivoitoiksi ennustetut ottelut. Sekä perus-Elo-malli että maaliperustainen Elo-malli suoritusvat tästä näkökulmasta tarkasteltuna paremmin kun aggregaatti-malli.

**Taulukko 24.** Aggregaatti-malli: ottelut, joiden algoritmi ennusti päättyvän kotivoittoon.

Kausi	Oikein	Vieras	Oikein %	Väärin %	Ottelut yht.
2008-2009	122	222	35.47%	64.53%	344
2009-2010	106	201	34.53%	65.47%	307
2010-2011	123	168	42.27%	57.73%	291
2011-2012	102	184	35.66%	64.34%	286
<b>Yhteensä:</b>	453	775	36.89%	63.11%	1228

Vierasvoitoiksi ennustetuissa otteluissa, aggregaatti-mallin onnistumisprosentti on myös kolmanneksi parhain. Parhaat ovat maaliperustainen Elo-malli ja pi-hockey -malli. Aggregaatti-mallin onnistumisprosentit myös vaihtelevat kausittain molemmissa kategorioissa. Kotivoitoiksi ennustetuista otteluista se kuitenkin siis pystyy ennustamaan oikein lähes 45% tapauksista.

#### 4.6. Yhteenveto tuloksista

Taulukossa 25 on koottuna kaikkien mallien onnistumisprosentit kausilla 2008-2012. Taulukosta nähdään, että kaikkein paras malli otteluiden lopputulosten ennustamiseen on kehitetyistä malleista painotettu maaliperustainen Elo-malli. Sen keskimääräinen onnistumisprosentti on 43,42%. Toiseksi paras malli on painotettu perus-Elo-malli, joka häviää painotettu maaliperustaiselle Elo-mallille kuitenkin vain 0,18 prosenttiyksikköä. Nämä ovat ainoat kehitetyistä malleista, joiden onnistumisprosentti on yli 43%.

Kun otetaan huomioon vain kotivoitoiksi ennustetut ottelut, paras malli on maaliperustainen Elo-malli. Se onnistuu ennustamaan 47,3% kotivoitoiksi ennustetuista otteluista oikein. Toiseksi paras malli on perus-Elo-malli, jonka onnistumisprosentti on 46,39%. Heikoin onnistumisprosentti oli Poisson- ja aggregaatti-malleilla, joilla se oli molemmilla keskimäärin 45,03%. Mallien onnistumisprosentit kotivoitoiksi ennustetuissa otteluissa listataan taulukossa 26.

**Taulukko 25.** Mallien vertailu: kaikki ottelut.

	<b>Kausi</b>				
<b>Malli</b>	2008-2009	2009-2010	2010-2011	2011-2012	Keskiarvot
Poisson	43.57%	42.39%	40.68%	42.25%	41.97%
Perus-Elo	42.68%	40.49%	42.20%	40.81%	41.55%
Perus-Elo: koti/vieras	44.23%	42.44%	40.24%	40.33%	41.81%
Perus-Elo: painotettu	44.15%	41.87%	44.07%	42.85%	43.24%
Maaliperus. Elo-malli	44.55%	42.28%	40.33%	40.16%	41.83%
Maaliperust. Elo: koti/vieras	44.47%	43.09%	40.24%	41.54%	42.34%
Maaliperust. Elo: painotettu	44.72%	42.20%	43.58%	43.17%	43.42%
Pi-hockey	42.60%	42.76%	40.57%	43.17%	42.28%
Aggregaatti-malli	44.23%	42.76%	42.76%	42.11%	42.97%

Taulukossa 27. on listattuna kaikkien mallien onnistumisprosentit otteluissa, jotka ennustettiin vierasvoitoiksi. Pi-hockey -malli on pudotettu pois tästä vertailusta, koska se ennusti niin harvan ottelun päättyvän vierasvoittoon, että sen tulokset eivät ole tässä suhteessa vertailukelpoisia muiden mallien kanssa. Vertailluista malleista paras ennustamaan vierasvoittoja oikein on maaliperustainen Elo-malli. Sen onnistumisprosentti on 37,11%. Heikoiten tässäkin suhteessa menestyy Poisson-malli, jonka onnistumisprosentti on 34,83%.

**Taulukko 26.** Mallien vertailu: ottelut, joiden algoritmi ennusti päättyvän kotivoittoon.

	<b>Kausi</b>				
<b>Malli</b>	2008-2009	2009-2010	2010-2011	2011-2012	Keskiarvot
Poisson-malli	47.96%	45.80%	42.47%	43.88%	45.03%
Perus-Elo-malli	49.28%	46.85%	44.00%	45.41%	46.39%
Maaliperustainen Elo-malli	51.02%	48.95%	44.52%	44.72%	47.30%
Pi-hockey	44.95%	44.82%	40.91%	43.17%	43.46%
Aggregaatti	47.63%	45.50%	42.92%	44.07%	45.03%

Tässä kappaleessa on käyty läpi kaikkien kehitettyjen mallien saamat onnistumisprosentit otteluiden lopputulosten ennustamisessa. Muista kuin Poisson-mallista ei ole huomioitu kuin kaudet 2008-2012, koska sekä Elo-malleissa että pi-hockey -mallissa, ja näin ollen myös aggregaatti-mallissa, kausia 2002-2008 on käytetty datana todennäköisyysarvioiden muodostamiseen, ja erityisesti niiden tarkkuuden maksimoimiseen.

Parhaiten toimivaksi malliksi pelkästään ennustamisen näkökulmasta havaittiin painotettu maalipohjainen Elo-malli. Se onnistui ennustamaan oikein noin 43,4% kaikista otteluista. Toiseksi paras malli oli painotettu perus-Elo-malli. Painotettujen mallien menestys on hyvä merkki siitä, että käytetyillä menetelmillä muodostetuilla Elo-luvuilla pystytään kuvaamaan sangen hyvin joukkueiden eroja seuraavalla kaudella.

**Taulukko 27.** Mallien vertailu: ottelut, joiden algoritmi ennusti päättyvän vierasvoittoon.

Malli	Kausi				
	2008-2009	2009-2010	2010-2011	2011-2012	Keskiarvot
Poisson-malli	33.33%	33.55%	36.86%	35.57%	34.83%
Perus-Elo-malli	35.87%	34.35%	40.33%	36.45%	36.75%
Maaliperustainen Elo-malli	37.56%	35.52%	39.73%	35.61%	37.11%
Aggregaatti	35.47%	34.53%	42.27%	35.66%	36.98%

Heikoiten malleista suoriutui perus-Elo-malli ilman painotuksia tai erillisiä Elo-lukuja koti- ja vierasotteluille: sen onnistumisprosentti oli noin 41,5%. Toiseksi heikoin malli oli perus-Elo-malli, joka arvioi Elo-luvut erikseen joukkueille sekä kotiotteluihin, että vierasotteluihin. Seuraavassa luvussa käsitellään mallien onnistumista vedonlyönnissä.

## 5. Vedonlyöntitulokset

Seuraavaksi käydään läpi ja esitellään malleittain vedonlyöntitulokset, jotka saatiin kun algoritmin muodostamien todennäköisyysarvioiden perusteella simuloitiin vedonlyöntiä. Vetoa lyötiin vain, kun jollekin ottelun lopputulokselle algoritmin antama todennäköisyys oli suurempi kuin vedonvälittäjän antama, eli löydettiin ylikerroin. Jos ottelussa oli useammalle lopputulokselle ylikerroin, lyötiin silloin vetoa järjestyksessä kotivoitto, vierasvoitto ja tasapeli. Joka ottelusta lyötiin siis vetoa vain kerran ja yhden lopputulosvaihtoehdon puolesta.

Simuloinnissa käytettiin todellisia, toteutuneita vedonvälittäjien kertoimia kyseisiin otteluihin. Kertoimia tallennettiin tietokantaan joka ottelusta kolmeltatoista eri vedonvälittäjältä, jos niiden kertoimet kyseiseen otteluun olivat tiedossa. Vedonvälittäjät joilta kertoimet kerättiin olivat 10Bet, 888Sport, Bet-at-Home, Bet365, Betclic, Betfair, Betsson, bwin, Expekt, Ladbrokes, PaddyPower, Tipico ja Unibet. Vedonvälittäjiin ei sinänsä simuloinnissa otettu kantaa, vaan kertoimeksi valittiin joka lopputulosvaihtoehdolle korkein mahdollinen kerroin. Simuloinneissa pelikassan kokona lähtötilanteessa käytettiin aina tuhatta (1000) rahayksikköä. Tuloslistauksissa rahayksikkönä käytetään euroa. Kellyn jakajana käytettiin aina lukua 5, joka vastaa kohtalaisen riskin ottoa pelaajalta. Asetettavalle panokselle ei asetettu mitään alarajaa, joten ne eivät aina vastaa todellisuutta, koska vedonvälittäjillä on aina jokin minimipanos. Asetettava panos laskettiin aina Kellyn kaavaa käyttämällä.

### 5.1. Poisson-malli – vedonlyöntitulokset

Poisson-mallilla lyötiin aluksi vetoa Kellyn jakajalla 5 niin, että kaikkia ylikertoimisia otteluita pelattiin. Pelikassan koko pelatun kauden jälkeen, asetettujen vetojen kappalemäärä ja niiden hinta euroissa, sekä voittojen kappalemäärä ja summa euroissa on esitetty taulukossa 28. Kuten voidaan havaita, vain kaudella 2011-2012, vedonlyönti kauden aikana on ollut voitokasta, mutta tuolloinkin voitto on vain 0,11 €.

Heikoimmallakin kaudella (kausi 2009-2010) tosin tappiolle jäätin kauden aikana vain 3,09 €, joka on todella pieni tappio, ainakin jos sitä verrataan keskivertopelaajan todennäköiseen tulokseen kun 1089 vetoa on lyöty kauden aikana. Luonnollisesti



tällaisia päätelmiä tehtäessä joudutaan luottamaan puhtaasti intuitioon ja maalaisjärkeen, koska on käytännössä mahdotonta mitata keskivertopelaajan keskimääräistä pelikassan kehitystä kokonaisen kauden aikana, tai edes lyhyemmälläkin aikavälillä.

**Taulukko 28.** Poisson-malli, Kellyn jakaja 5, kaikki lopputulokset huomioitu.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.81 €	793	26.23 €	316	25.04 €
2009-2010	1,000.00 €	996.91 €	842	28.06 €	329	24.97 €
2010-2011	1,000.00 €	997.91 €	827	25.84 €	315	23.74 €
2011-2012	1,000.00 €	1,000.11 €	854	26.86 €	346	26.97 €
<b>Keskiarvot:</b>	1,000.00 €	998.44 €	829	26.75 €	327	25.18 €

Seuraavaksi kokeiltiin, paraneeko pelikassan kehitys, jos lyödään vetoa vain otteluista, joiden algoritmi ennusti päättyvän kotivoittoon. Tulokset on esitetty taulukossa 29. Vain kotivoitoiksi ennustettuja ylikertoimia pelattaessa, pelikassan kehitys on hiukan parempi verrattuna tilanteeseen, kun vetoa lyötiin kaikista ylikertoimisista otteluista. Vain kotivoittoja pelaamalla pelikassa oli suurempi kuin kauden alussa kausilla 2008-2009. Tuolla kaudella pelikassa oli lopuksi suurempi kuin millään kaudella pelattaessa kaikkia ylikertoimisista otteluista. Keskimäärin, pelikassa oli kauden loppuessa 999,44 €. Keskimäärin kauden aikana lyötiin vetoa 544 kertaa joista keskimäärin 231 vetoa kaudessa oli voittavia vetoja, keskimääräisen koko kauden voittosumman ollessa 16,28 €.

**Taulukko 29.** Poisson-malli, Kellyn jakaja 5, vain kotivoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,000.90 €	547	18.14 €	242	19.00 €
2009-2010	1,000.00 €	998.85 €	568	17.26 €	240	16.11 €
2010-2011	1,000.00 €	998.38 €	504	15.05 €	199	13.43 €
2011-2012	1,000.00 €	999.61 €	557	16.96 €	241	16.57 €
<b>Keskiarvot:</b>	1,000.00 €	999.44 €	544	16.85 €	231	16.28 €

Seuraavaksi mitattiin ja analysoitiin pelikassan kehitystä, kun lyötiin vetoa vain ylikertoimisista otteluista, joiden algoritmi ennusti päättyvän vierasvoittoon. Tulokset on esitetty taulukossa 30. Vierasvoittoja pelattaessa pelikassan kehitys ei ollut aivan yhtä suotuisa kuin vain kotivoittoja pelatessa. Paras tulos saavutettiin kaudella 2011-2012, jolloin voittoa kertyi koko kauden aikana 0,50 €. Keskimääräinen vetojen yhteishinta kautta kohden oli 9,89 € ja vetoa lyötiin keskimäärin 285 kertaa kaudessa. Keskimääräinen voittosumma kautta kohden oli 8,89 € ja voittoja tuli keskimäärin 96 kappaletta kaudessa. Pelikassan koko kauden lopussa oli keskimäärin 999,00 euroa. Tämä oli 0,44 euroa pienempi kuin pelattaessa vain kotivoittoja.

Tämän jälkeen vetoa lyötiin niin, että Poisson-funktiolle syötettäviin maalikeskiarvoihin laskettiin vain kolmen viimeisimmän ottelun tehdyt maalit mukaan. Kotijoukkueelle laskettiin mukaan vain sen kotiotteluissa tekemät maalit ja vierasjoukkueelle sen vierasotteluissa tekemät maalit. Kuten tuloksista taulukossa 31 havaitaan, pelikassan kehitys muuttuu nyt heikommaksi verrattuna aiempien mallien tuloksiin. Kaudella 2011-2012, joka aiemmilla Poisson-mallin variaatioilla oli paras kausi, on pelikassa kauden lopussa nyt 995,84 €, joka on heikoiten kaikista analysoiduista kausista millään Poisson-mallin variaatiolla tähän mennessä. Keskimäärin pelikassan saldo oli kauden päättyessä 998,18 €.

**Taulukko 30.** Poisson-malli, Kellyn jakaja 5, vain vierasvoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	997.92 €	246	8.09 €	74	6.00 €
2009-2010	1,000.00 €	998.05 €	274	10.80 €	89	8.86 €
2010-2011	1,000.00 €	999.53 €	323	10.78 €	116	10.31 €
2011-2012	1,000.00 €	1,000.50 €	297	9.90 €	105	10.40 €
<b>Keskiarvot:</b>	1,000.00 €	999.00 €	285	9.89 €	96	8.89 €

Huomattavaa on, että keskimääräiset pelisummat euroissa mitattuna ovat lähes kolminkertaiset joka kaudella verrattuna Poisson-malliin (taulukko 28), joka huomioi maalikeskiarvojen laskemisessa kaikki siihen asti pelattujen otteluiden tehdyt maalit. Sama tosin pätee keskimääräisiin kokonaisvoittosummiinkin kaudessa. Pelattujen vetojen kappalemäärissä eroa on noin 100 vetoa kaudessa.

**Taulukko 31.** Poisson-malli, Kellyn jakaja 5, 3 viimeisintä ottelua huomioitu maalikeskiarvoon.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.77 €	940	66.28 €	381	65.06 €
2009-2010	1,000.00 €	999.26 €	976	63.41 €	400	62.67 €
2010-2011	1,000.00 €	998.84 €	974	64.54 €	381	63.39 €
2011-2012	1,000.00 €	995.84 €	954	62.86 €	360	58.69 €
<b>Keskiarvot:</b>	1,000.00 €	998.18 €	961	64.27 €	381	62.45 €

Seuraavaksi maalikeskiarvoihin laskettavien otteluiden määrä nostettiin kolmesta neljään. Tulokset on esitetty taulukossa 32. Neljän pelin huomiointi maalikeskiarvoissa ei juurikaan muuta tuloksia eikä tee vedonlyönnistä voitokasta. Pelikassan keskiarvosaldo on itse asiassa muutaman sentin pienempi kuin kun huomioitiin 3 viimeistä ottelua joukkueilta. Pelattujen vetojen määrä luonnollisesti vähenee, koska kaiken kaikkiaan käsiteltävien pelien määrä vähenee hiukan ja jokaisella joukkueella pitää olla vähintään neljä ottelua pelattuna ennen kuin niiden otteluita aletaan tarkastelemaan vedonlyönnin kannalta.

**Taulukko 32.** Poisson-malli, Kellyn jakaja 5, 4 viimeisintä ottelua huomioitu maalikeskiarvoon.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.72 €	910	57.31 €	376	57.04 €
2009-2010	1,000.00 €	999.13 €	904	52.61 €	367	51.74 €
2010-2011	1,000.00 €	997.77 €	930	56.66 €	359	54.43 €
2011-2012	1,000.00 €	995.77 €	927	54.10 €	356	49.87 €
<b>Keskiarvot:</b>	1,000.00 €	998.10 €	918	55.17 €	365	53.27 €

Kun joukkueiden maalikeskiarvoihin huomioitiin viidessä viimeisimmässä ottelussa tehdyt maalit, saatiin tuloksia, jotka esitellään taulukossa 33. Kun tuloksia verrataan Poisson-malliin, jossa joukkueiden kaikki aikaisemmat pelit lasketaan mukaan maalikeskiarvoihin, onnistuu viisi viimeisintä ottelua huomioiva malli paremmin, sen keskimääräisen pelikassan saldon kauden lopussa ollessa 999,15 € kun kaikki ottelut huomioivalla Poisson-mallilla vastaava luku on 998,44 €. Paras pelikassan saldo kauden loputtua tällä mallilla saavutettiin kaudella 2008-2009, jolloin saldo oli 1,08 € plussan puolella. Keskimäärin algoritmi löi vetoa 887 ottelusta kautta kohti ja voitti keskimäärin 356 vetoa kautta kohti keskimääräisen koko kauden voittosumman ollessa 47,70 €.

**Taulukko 33.** Poisson-malli, Kellyn jakaja 5, 5 viimeisintä ottelua huomioitu maalikeskiarvoon.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,001.08 €	866	50.03 €	352	51.11 €
2009-2010	1,000.00 €	1,000.02 €	898	47.99 €	367	48.01 €
2010-2011	1,000.00 €	998.28 €	897	49.50 €	362	47.79 €
2011-2012	1,000.00 €	997.20 €	886	46.69 €	341	43.89 €
<b>Keskiarvot:</b>	1,000.00 €	999.15 €	887	48.55 €	356	47.70 €

**Taulukko 34.** Poisson-malli, Kellyn jakaja 5, 6 viimeisintä ottelua huomioitu maalikeskiarvoon.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,002.55 €	836	45.74 €	346	48.29 €
2009-2010	1,000.00 €	998.80 €	852	43.56 €	354	42.36 €
2010-2011	1,000.00 €	998.16 €	859	43.70 €	340	41.86 €
2011-2012	1,000.00 €	996.51 €	841	41.50 €	314	38.00 €
<b>Keskiarvot:</b>	1,000.00 €	999.01 €	847	43.63 €	339	42.63 €

Lopuksi tarkasteltiin pelikassan kehitystä, kun kuusi viimeisintä ottelua joukkueilta huomioitiin maalikeskiarvon laskemisessa. Tulokset esitellään taulukossa 34. Kuten malleissa, joissa huomioitiin neljä ja viisi viimeisintä ottelua, myös kuusi ottelua huomioitaessa kaudella 2008-2009 saavutetaan pelikassan näkökulmasta paras tulos.

Tuolloin pelikassa oli 2,55 € plussan puolella kauden päätyttyä, kun keskimääräinen pelikassan saldo kautta kohden oli 999,01 €. Keskimäärin voitokkaita vetoja oli 339 kautta kohden.

Tässä kappaleessa on käsitelty Poisson-mallin eri variaatioiden menestystä vedonlyönnissä. Vaikka mikään variaatio ei menestynyt niin hyvin, että olisi tuottanut voittoa jokaisella analysoidulla kaudella, mikään ei myöskään onnistunut tuottamaan suuria tappioita.

Paras yksittäisen kauden pelikassan saldon kehitys saavutettiin Poisson-mallin variaatiolla, jossa maalikeskiarvoihin laskettiin kuusi viimeisintä ottelua joukkueille. Tällä variaatiolla pelikassa oli 1002,55 € kauden 2008-2009 päätteeksi. Tuolla kaudella vetoa lyötiin keskimäärin 836 kertaa kaudessa ja keskimäärin 43,70 eurolla. Voittoihin näistä vedoista johti keskimäärin 340 vetoa keskimääräisen voittosumman kautta kohti ollessa 41,86 €.

## 5.2. Perus-Elo-malli – vedonlyöntitulokset

Seuraavaksi käsitellään perus-Elo-mallin vedonlyöntisimulointien tuloksia. Perus-Elo-mallihan noudattelee täysin Hvattumin ja Arntzenin (2010) artikkelin Elo-mallia, joka taas noudattaa täysin alkuperäistä, shakkipelaajien arviointiin alunperin Arpad Elon kehittämää Elo-mallia. (Elo, 1978). Kaikissa simulaatioissa Kellyn jakajana on käytetty lukua 5 ja tuloksissa käytettynä rahayksikkönä on euro. Aluksi simuloitiin perus-Elo-mallia jossa vetoa lyötiin kaikista ylikertoimen tarjoavista otteluista ja Elo-lukuihin ei ole laskettu mukaan painotuksia, ja jokaiselle joukkueelle on määritelty vain yksi Elo-luku, ei erikseen koti- ja vierasotteluiden Elo-lukuja. Tulokset on esitelty taulukossa 35.

**Taulukko 35.** Perus-Elo-malli, Kellyn jakaja 5, kaikista lopputuloksista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.62 €	404	7.12 €	127	5.74 €
2009-2010	1,000.00 €	998.83 €	547	8.74 €	178	7.58 €
2010-2011	1,000.00 €	1,000.01 €	508	7.93 €	183	7.93 €
2011-2012	1,000.00 €	1,000.40 €	475	6.95 €	167	7.36 €
<b>Keskiarvot:</b>	1,000.00 €	999.47 €	484	7.69 €	164	7.15 €

Verrattuna esimerkiksi Poisson-malliin, joka huomioi kaikki aiemmat ottelut, perus-Elo-mallin saavuttamat tulokset keskimääräisenä pelikassan saldona kauden lopussa mitattuna, ovat hiukan parempia. Poisson-mallin keskimääräinen kauden lopussa mitattu pelikassan saldo oli 998,44 €, kun se perus-Elo-mallilla on 999,47 €, eli 1,03 € parempi. Parhaiten perus-Elo-malli suoriutui kaudella 2011-2012, jolloin pelikassa oli 0,40 € plussan puolella kauden päättyessä. Vaikka perus-Elo-malli ei anna Elo-lukujen kehittyä tiettyä ottelumäärää kauden alusta ennen kuin otteluille aletaan arvioimaan todennäköisyyksiä, niiden otteluiden määrä joista vetoa on lyöty, on huomattavasti pienempi kuin Poisson-mallilla. Keskimäärin perus-Elo-mallilla simuloidessa vetoa lyötiin 484 ottelussa kautta kohden, joista oikein meni keskimäärin 164 ottelua kaudessa. Huomattavaa on, että keskimääräinen vetosumma kautta kohden perus-Elo-

mallilla on alle puolet Poisson-mallin vastaavasta. Keskimäärin perus-Elo-malli löi vetoa 7,69 eurolla kautta kohden.

**Taulukko 36.** Perus-Elo-malli, Kellyn jakaja 5, vain kotivoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,000.33 €	33	1.39 €	12	1.71 €
2009-2010	1,000.00 €	1,000.09 €	35	0.37 €	13	0.46 €
2010-2011	1,000.00 €	999.84 €	52	0.50 €	13	0.34 €
2011-2012	1,000.00 €	1,000.06 €	49	0.38 €	17	0.44 €
<b>Keskiarvot:</b>	1,000.00 €	1,000.08 €	42	0.66 €	14	0.74 €

Seuraavaksi perus-Elo-mallia käytettiin simulointiin niin, että algoritmi löi vetoa ylikertoimia tarjoavista otteluista, vain jos niiden arvioitiin päättyvän kotivoittoon. Kuten taulukosta 36 nähdään, tämä variaatio perus-Elo-mallista toimii tähän asti testatuista malleista parhaiten, kun toimivuutta mitataan pelikassan keskimääräisenä kokona kauden lopussa. Keskimäärin pelikassan saldo oli kauden lopussa 1000,08 euroa. Perus-Elo-malli, jolla lyötiin vetoa vain algoritmin kotivoitoiksi arvioimia pelejä, tuotti voittoa kolmella kaudella neljästä. Lyödyt vetosummat ovat kuitenkin varsin pieniä verrattuna muihin tähän mennessä analysoituihin malleihin. Keskimäärin tämä perus-Elo-mallin variaatio löi vetoa vain 0,66 € kaudessa. Keskimääräinen koko kauden voittosumma oli 0,74 euroa.

**Taulukko 37.** Perus-Elo-malli, Kellyn jakaja 5, vain vierasvoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.29 €	371	5.74 €	115	4.03 €
2009-2010	1,000.00 €	998.75 €	512	8.37 €	165	7.12 €
2010-2011	1,000.00 €	1,000.17 €	456	7.42 €	170	7.59 €
2011-2012	1,000.00 €	1,000.34 €	426	6.57 €	150	6.91 €
<b>Keskiarvot:</b>	1,000.00 €	999.39 €	441	7.03 €	150	6.41 €

Taulukossa 37 on esitettynä tulokset, jotka saatiin kun perus-Elo-mallilla käytettiin simulointiin niin, että vetoa lyötiin vain niistä otteluista, joiden algoritmi ennusti päättyvän vierasvoittoon. Nyt vetoa on lyöty huomattavasti useammin verrattuna perus-Elo-mallin -variaatioon, jossa vetoa lyötiin vain ylikertoimisista otteluista, joiden algoritmi ennusti päättyvän kotivoittoon. Kun lyötiin vetoa vain vierasvoitoista, pelikassa oli suurempi kuin kauden alussa kahdella kaudella, kausilla 2010-2011 ja 2011-2012. Tosin noilla kausilla voitto oli vain muutamia kymmeniä senttejä. Keskimäärin pelikassan koko kauden lopussa oli 999,39 €. Keskimääräinen koko kauden vetojen hinta oli 7,03 € ja keskimääräinen koko kauden voittosumma oli 6,41 €. Vetoa lyötiin keskimäärin 441 kertaa kaudessa joista voittoa tuotti keskimäärin 150 vetoa kaudessa.

Taulukossa 38 on esitetty tulokset, kun käytettiin perus-Elo-mallin variaatiota, jossa sekä kotiotteluille että vierasotteluille lasketaan omat erilliset Elo-lukunsa. Tämä variaatio on hiukan parempi vedonlyönnin kannalta kuin perus-Elo-malli. Keskimäärin pelikassan koko kauden lopussa on 999,70 €. Kausi 2011-2012 on ainoa kausi, jolloin pelikassa on plussan puolella kauden lopussa. Tuolloin voittoa koko kauden aikana kertyi 0,67 €. Keskimäärin vetoa lyötiin 493 kertaa kaudessa keskimäärin 7,90 eurolla. Oikein lyödyistä vedoista päättyi keskimäärin 174 vetoa kaudessa ja keskimääräinen koko kauden voittosumma oli 7,60 €. Suurin yksittäisen kauden voittosumma saavutettiin kaudella 2011-2012, jolloin voittosumma oli 8,70 euroa.

**Taulukko 38.** Perus-Elo-malli, Kellyn jakaja 5, koti- ja vieras-Elo-luvut erikseen laskettu.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.85 €	366	6.41 €	123	5.26 €
2009-2010	1,000.00 €	999.59 €	553	9.03 €	201	8.62 €
2010-2011	1,000.00 €	999.70 €	524	8.13 €	176	7.83 €
2011-2012	1,000.00 €	1,000.67 €	528	8.03 €	194	8.70 €
<b>Keskiarvot:</b>	1,000.00 €	999.70 €	493	7.90 €	174	7.60 €

Lopuksi vedonlyöntiä simuloitiin perus-Elo-mallin -variaatiolla, jossa käytössä olivat painotetut Elo-luvut. Elo-lukujen painottaminen toteutettiin niin, että Elo-luvut, jotka kullakin joukkueella oli edellisen kauden loppuessa, siirrettiin suoraan muuttamattomina seuraavan kauden alun Elo-luvuiksi joukkueille. Tällä pyrittiin tilanteeseen, jossa Elo-luvut kauden alussa vastaisivat paremmin joukkueiden todellisia tasoeroja.

**Taulukko 39.** Perus-Elo-malli, Kellyn jakaja 5, painotetut Elo-luvut.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.40 €	565	10.48 €	204	8.87 €
2009-2010	1,000.00 €	997.86 €	628	12.39 €	226	10.25 €
2010-2011	1,000.00 €	1,001.20 €	632	14.93 €	263	16.13 €
2011-2012	1,000.00 €	1,001.79 €	704	16.76 €	283	18.55 €
<b>Keskiarvot:</b>	1,000.00 €	999.81 €	632	13.64 €	244	13.45 €

Tähän mennessä analysoiduista malleista ja niiden variaatioista, tämä perus-Elo-mallin variaatio on toiminut parhaiten kun mittayksikkönä käytetään keskimääräistä pelikassan kokoa kauden loppuessa. Tuo keskiarvo oli painotetulla perus-Elo-mallilla 999,81 euroa. Kausilla 2010-2011 ja 2011-2012 pelikassa oli yli euron verran kauden alkua suurempi, kaudella 2011-2012 1,70 euroa. Keskimäärin vetoa lyötiin 632 kertaa kaudessa keskimäärin 13,64 eurolla. Voitokkaita näistä vedoista oli keskimäärin 244 kaudessa ja keskimääräinen koko kauden voittosumma oli 13,45 euroa.

### 5.3. Maaliperustainen Elo-malli – vedonlyöntitulokset

Maaliperustainen Elo-malli eroaa perus-Elo-mallista siinä, kuinka Elo-lukuja päivitetään otteluiden jälkeen. Maaliperustaisessa Elo-mallissa päivittäminen tehdään perustuen ottelun toteutuneeseen maalieroon kun taas perus-Elo-mallissa päivittäminen tehdään perustuen joukkueiden ennustetun ja toteutuneen pistemäärän erotukseen.

**Taulukko 40.** Maaliperustainen Elo-malli, Kellyn jakaja 5, kaikista otteluista lyöty vetoa.

Kausi	Pelikassa		Vedot		Voitot	
	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.57 €	624	15.75 €	238	15.32 €
2009-2010	1,000.00 €	999.19 €	687	15.97 €	248	15.16 €
2010-2011	1,000.00 €	1,002.16 €	892	36.74 €	380	38.90 €
2011-2012	1,000.00 €	1,000.30 €	864	24.14 €	342	24.44 €
<b>Keskiarvot:</b>	1,000.00 €	1,000.31 €	767	23.15 €	302	23.46 €

Taulukossa 40 on esitetty tulokset, jotka saatiin kun maaliperustaista Elo-mallia käytettiin vedonlyönnin simulointiin niin, että vetoa lyötiin kaikista lopputuloksista, joihin saatiin ylikerron sille lopputulokselle, jonka algoritmi ennusti olevan todennäköisin. Tämä malli on ensimmäinen, jolla tehtyjen simulaatioiden keskimääräinen pelikassan koko kauden jälkeen on suurempi kuin se oli kauden alussa. Keskimäärin pelikassa oli kauden jälkeen 1000,31 euroa. Parhaalla kaudella 2010-2011 voittoa kertyi 2,16 euroa koko kauden aikana. Keskimäärin kauden aikan lyötiin vetoa 767 kertaa keskimäärin 23,15 eurolla kaudessa. Voittoihin noista 767 keskimääräisestä vedosta päättyi keskimäärin 302 vetoa kaudessa. Keskimääräinen koko kauden voittokertymä oli 23,46 €.

**Taulukko 41.** Maaliperustainen Elo-malli, Kellyn jakaja 5, vain kotivoitoista lyöty vetoa.

Kausi	Pelikassa		Vedot		Voitot	
	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,000.46 €	121	3.18 €	52	3.64 €
2009-2010	1,000.00 €	1,000.38 €	123	2.02 €	51	2.40 €
2010-2011	1,000.00 €	1,000.90 €	295	10.40 €	161	11.30 €
2011-2012	1,000.00 €	999.70 €	290	5.06 €	134	4.77 €
<b>Keskiarvot:</b>	1,000.00 €	1,000.36 €	207	5.17 €	100	5.53 €

Seuraavaksi vetoa lyötiin vain sellaisista otteluista, jotka maaliperustaisen Elo-mallin laskelmien perusteella arvioitiin kotivoitoiksi, ja joissa todennäköisyys kotivoitolle oli suurempi kuin vedonlyöntitoimistojen vastaava todennäköisyysarvio. Kuten taulukosta 41 nähdään, tällä menetelmällä pelikassa on plussalla kolmella kaudella neljästä. Suurin pelikassan koko kauden jälkeen oli kaudella 2010-2011, jolloin voittoa koko kauden aikana kertyi 0,90 euroa. Keskimäärin pelikassan koko kauden jälkeen oli 1000,36 €.

**Taulukko 42.** Maaliperustainen Elo-malli, Kellyn jakaja 5, vain vierasvoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.11 €	503	12.57 €	186	11.68 €
2009-2010	1,000.00 €	998.81 €	564	13.95 €	197	12.76 €
2010-2011	1,000.00 €	1,001.26 €	528	23.33 €	203	24.59 €
2011-2012	1,000.00 €	1,000.60 €	574	19.08 €	208	19.68 €
<b>Keskiarvot:</b>	1,000.00 €	999.95 €	542	17.23 €	199	17.18 €

Tämän jälkeen vetoa lyötiin vain otteluista, joissa algoritmin vierasvoitoksi arvioimissa otteluissa oli ylikerroin. Tällä menetelmällä kausilla 2010-2011 ja 2011-2012, pelikassan koko oli suurempi kuin kauden alussa. Keskimäärin pelikassan koko kauden jälkeen oli 999,95 euroa eli 0,05 euroa vähemmän kuin kauden alussa. Vetoa lyötiin keskimäärin 17,23 eurolla kaudessa ja keskimääräinen voittokertymä yksittäisen kauden aikana oli 17,18 euroa.

**Taulukko 43.** Maaliperust. Elo-malli, Kellyn jakaja 5, koti- ja vieras-Elo-luvut laskettu erikseen.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.50 €	831	26.29 €	357	24.79 €
2009-2010	1,000.00 €	998.69 €	869	27.45 €	364	26.14 €
2010-2011	1,000.00 €	999.75 €	840	36.69 €	348	36.44 €
2011-2012	1,000.00 €	1,000.24 €	919	37.74 €	388	37.98 €
<b>Keskiarvot:</b>	1,000.00 €	999.30 €	865	32.04 €	364	31.34 €

**Taulukko 44.** Maaliperustainen Elo-malli, Kellyn jakaja 5, painotetut Elo-luvut.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	998.27 €	681	17.42 €	262	15,68 €
2009-2010	1,000.00 €	998.53 €	708	18.08 €	284	16.61 €
2010-2011	1,000.00 €	1,000.97 €	1142	58.26 €	495	59.23 €
2011-2012	1,000.00 €	1,001.32 €	1040	39.30 €	434	40.61 €
<b>Keskiarvot:</b>	1,000.00 €	999.77 €	893	33.27 €	369	33.03 €

Kun simulointiin käytettiin maaliperustaisen Elo-mallin variaatiota, jossa kullekin joukkueelle laskettiin erilliset Elo-luvut koti- ja vierasjoukkueille, saatiin hiukan heikompia tuloksia kuin maaliperustaisella Elo-mallilla, jossa käytettiin samaa Elo-lukua joukkueelle sekä koti- että vierasotteluissa. Kuten taulukosta 43 havaitaan, kausi 2011-2012 oli ainoa kausi, jolloin tällä menetelmällä saavutettiin positiivinen pelikassan kehitys koko kauden aikana. Tuolla kaudella pelikassa oli kauden päättyessä 1000,24 €.



Tämä summa on 0,94 € suurempi, kuin keskimääräinen pelikassan koko kauden päättyessä tällä simulointimenetelmällä.

Lopuksi simulointiin käytettiin maaliperustaisen Elo-mallin variaatiota, jossa käytettiin painotettuja Elo-lukuja. Tulokset on esitettynä taulukossa 44. Keskimääräinen pelikassan koko tällä simulointimenetelmällä on hiukan suurempi kuin käytettäessä erillisiä Elo-lukuja sekä koti- että vierasotteluissa. Verrattuna maaliperustaiseen Elo-malliin, jossa käytetään vain yhtä Elo-lukua joukkuetta kohti, tulos on kuitenkin jonkin verran heikompi. Tällä menetelmällä pelikassan kehitys on kauden aikana positiivinen kausilla 2010-2011 ja 2011-2012. Näillä molemmilla kausilla pelattujen vetojen määrä kauden aikana on huomattavasti suurempi kuin kausilla 2008-2009 ja 2009-2010. Luonnollisesti myös sekä vetojen että voittojen yhteissumma kaudella on huomattavasti suurempi.

#### 5.4. Pi-hockey -malli – vedonlyöntitulokset

Pi-hockey -malli on jääkiekkoon sovellettu versio Constantinoun ja Fentonin (2012) artikkelissa esittelystä pi-rating -mallista. Mallissa käytetyt pi-luvut joukkueille lasketaan juuri samoin kuin mainitussa artikkelissa tehtiin, mutta todennäköisyys-arvioiden muodostaminen lopputuloksille toteutaa eri tavalla.

**Taulukko 45.** Pi-hockey -malli, Kellyn jakaja 5, kaikista lopputuloksista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.19 €	535	12.86 €	179	12.05 €
2009-2010	1,000.00 €	1,000.97 €	563	12.12 €	196	13.09 €
2010-2011	1,000.00 €	999.53 €	494	9.15 €	153	8.69 €
2011-2012	1,000.00 €	1,001.10 €	503	9.57 €	197	10.67 €
<b>Keskiarvot:</b>	1,000.00 €	1,000.20 €	524	10.93 €	181	11.13 €

Aluksi pi-hockey -mallia käytettiin vedonlyönnin simulointiin niin, että vetoa lyötiin kaikista lopputuloksista, joissa algoritmin antama todennäköisyysarvio oli suurempi kuin vedonvälittäjien vastaava. Taulukossa 45 on esitettynä tulokset tästä simulaatiosta. Kuten taulukosta 45 havaitaan, kausilla 2009-2010 ja 2011-2012 algoritmi onnistuu saavuttamaan positiivisen tuloksen koko kauden aikana, eli pelikassan saldo on suurempi kauden lopussa verrattuna kauden alkuun. Myös keskimäärin, pelikassan koko oli 0,20 € plussan puolella kun kaikki neljä kautta otettiin huomioon. Keskimäärin vetoa lyötiin 524 kertaa kaudessa keskimääräisen koko kauden vetosumman ollessa 10,93 €. Voittoon vedoista päättyi keskimäärin 181 vetoa ja kokonaisvoittosumma keskimäärin kaudessa oli 11,13 €.

Kun vedonlyöntiä simuloitiin pi-hockey -mallilla niin, että lyötiin vetoa vain algoritmin kotivoitoiksi ennustamista ylikertoimisista otteluista, tulokset heikentyivät hiukan. Keskimääräinen koko kauden pelikassan saldo oli 999,88 €, eli se oli keskimäärin 0,12 € miinuksella. Parhaalla kaudella 2011-2012, voittoa koko kauden aikana kertyi 1,10 €. Tuolla kyseisellä kaudella myös vetoja oli kappalemäärältään eniten. Myös vetojen kokonaishinta oli kaudella 2011-2012 suurin. Samoin voittojen yhteismäärä ja yhteissumma olivat suurimmat tuolla kaudella.

**Taulukko 46.** Pi-hockey -malli, Kellyn jakaja 5, vain kotivoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.02 €	392	8.63 €	139	7.65 €
2009-2010	1,000.00 €	1,000.48 €	392	7.09 €	144	7.57 €
2010-2011	1,000.00 €	998.92 €	457	7.53 €	142	6.45 €
2011-2012	1,000.00 €	1,001.10 €	503	9.57 €	197	10.67 €
<b>Keskiarvot:</b>	1,000.00 €	999.88 €	436	8.21 €	156	8.09 €

Lopuksi pi-hockey -mallia käytettiin vedonlyönnin simulointiin niin, että vetoa lyötiin vain algoritmin vierasvoitoksi arvioimista ylikertoimisista otteluista. Nämä tulokset on esitetty taulukossa 47. Kuten voidaan huomata, taulukossa esitetään tulokset vain kausilta 2008-2010. Tähän on syynä se, että kausilla 2010-2012, algoritmi ei lyönyt vetoa yhdestäkään vierasvoitoksi ennustetusta ottelusta. Toisin sanoen millekään vierasvoitoksi ennustetulle ottelulle ei saatu ylikerronta. Myös taulukossa esitetyillä kahdella kaudella, lyötyjen vetojen määrät ovat varsin pieniä verrattuna muihin malleihin. Keskimäärin vetoa lyötiin 88 kertaa kaudessa keskimäärin 1,87 euron edestä kaudessa. Voitokkaita noista vedoista oli keskimäärin 28 kappaletta keskimääräisen kokokauden voittokertymän ollessa 1,90 euroa. Mielenkiintoista on myös, että molempien taulukossa listattujen kausien kauden jälkeisten pelikassojen saldojen ero on vain 0,01 euroa.

**Taulukko 47.** Pi-hockey -malli, Kellyn jakaja 5, vain vierasvoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	1,000.03 €	108	2.43 €	31	2.46 €
2009-2010	1,000.00 €	1,000.04 €	68	1.30 €	25	1.34 €
<b>Keskiarvot:</b>	1,000.00 €	1,000.04 €	88	1.87 €	28	1.90 €

Pi-hockey -malli oli toinen malli analysoiduista malleista ja niiden variaatioista, jonka keskimääräinen pelikassan kehitys analysoitujen kausien aikana oli positiivinen. Se on myös ainoa malli, jossa kotivoitoiksi ja vierasvoitoiksi arvioitujen ylikertoimisten otteluiden suhde on hyvin epätasainen. Tämä epätasaisuus ilmenee niin, että kotivoitoksi ennustettuja ylikertoimisia otteluita on moninkertainen määrä verrattuna vierasvoitoiksi ennuistettuihin.

## 5.5. Aggregaatti-malli – vedonlyöntitulokset

Aggregaatti-malli on yhdistelmämalli, joka hyödyntää Poisson-mallia, perus-Elo-mallia sekä pi-hockey -mallia. Jokainen näistä mainituista malleista tuottaa omat todennäköisyysarvionsa kustakin ottelun lopputulosvaihtoehdosta ja näistä arvioista lasketaan sen jälkeen keskiarvot, joita käytetään ottelun ennustamiseen.

Kuten muidenkin mallien kohdalla, aluksi lyötiin vetoa kaikista niistä otteluista, joihin saatiin ylikerron sille lopputulosvaihtoehdelle, jonka malli ennusti todennäköi-

simmäksi. Tulokset on esitetty taulukossa 48. Tällä menetelmällä pelikassan kehitys kauden aikana oli positiivinen kausilla 2010-2011 ja 2011-2012. Molemmilla kausilla koko kauden aikana kertynyt voitto oli kuitenkin alle euron verran.

**Taulukko 48.** Aggregaatti-malli, Kellyn jakaja 5, kaikista lopputuloksista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.22 €	139	2.45 €	37	1.67 €
2009-2010	1,000.00 €	999.58 €	198	3.39 €	61	2.97 €
2010-2011	1,000.00 €	1,000.34 €	172	2.58 €	66	2.92 €
2011-2012	1,000.00 €	1,000.66 €	139	2.43 €	53	3.09 €
<b>Keskiarvot:</b>	1,000.00 €	999.95 €	162	2.71 €	54	2.66 €

Heikoin tulos saavutettiin kaudella 2008-2009, jolloin pelikassa oli kauden lopussa miinuksella 0,78 euroa. Keskimääräinen pelikassan saldo kauden jälkeen oli 999,95 €. Kuten taulukosta nähdään, kokonaisvetosummat kauden aikana olivat varsin pieniä, keskimäärin vain 2,71 euroa. Keskimäärin vetoa lyötiin 162 kertaa kauden aikana joista voittoon päättyi keskimäärin 54 ottelua. Keskimääräinen koko kauden voittosumma oli 2,66 €.

**Taulukko 49.** Aggregaatti-malli, Kellyn jakaja 5, vain kotivoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.79 €	239	4.12 €	86	3.91 €
2009-2010	1,000.00 €	999.75 €	303	3.77 €	104	3.52 €
2010-2011	1,000.00 €	999.53 €	264	3.09 €	92	2.61 €
2011-2012	1,000.00 €	999.74 €	319	4.06 €	122	3.80 €
<b>Keskiarvot:</b>	1,000.00 €	999.70 €	281	3.76 €	101	3.46 €

**Taulukko 50.** Aggregaatti-malli, Kellyn jakaja 5, vain vierasvoitoista lyöty vetoa.

	Pelikassa		Vedot		Voitot	
Kausi	Aluksi	Lopuksi	kpl	€	kpl	€
2008-2009	1,000.00 €	999.22 €	139	2.45 €	37	1.67 €
2009-2010	1,000.00 €	999.58 €	198	3.39 €	61	2.97 €
2010-2011	1,000.00 €	1,000.34 €	172	2.58 €	66	2.92 €
2011-2012	1,000.00 €	1,000.66 €	139	2.43 €	53	3.09 €
<b>Keskiarvot:</b>	1,000.00 €	999.95 €	162	2.71 €	54	2.66 €

Kun vetoa lyötiin vain sellaisista ylikertoimisista otteluista, joiden algoritmi oli ennustanut päättävän kotivoittoon, pelikassan kehitys oli joka kaudella negatiivinen.

Keskimäärin pelikassan saldo oli kauden lopussa 999,70 €. Parhaalla kaudella, eli kaudella 2008-2009, pelikassan saldo oli kauden päätteeksi 999,70 €. Heikoimmalla kaudella, eli kaudella 2010-2011, pelikassan saldo oli 999,53 € kauden lopussa. Keskimääräinen koko kauden vetoihin käytetty summa oli 3,76 €. Keskimääräinen koko kauden yhteisvoittosumma oli 3,46 €.

Kun lyötiin vetoa ylikertoimisista otteluista, joiden algoritmi ennakoiti päättyvän vierasvoittoon, saavutettiin hiukan parempia tuloksia kuin kun lyötiin vetoa vain kotivoitoista. Tällä menetelmällä keskimääräinen pelikassan koko kauden päätteeksi oli 999,95 €. Kausilla 2010-2011 ja 2011-2012 pelikassan kehitys oli positiivinen voiton kuitenkin jäädessä molemmilla kausilla alle yhteen euroon.

## 5.6. Yhteen veto vedonlyöntituloksista

Tässä kappaleessa on käyty läpi kaikkien mallien ja mallivariaatioiden vedonlyöntisimulointien tulokset. Olennaisimmat osat näistä tuloksista on koottu myös taulukkoon 51. Kuten taulukosta havaitaan, vain kahdessa tapauksessa, kaikkien simuloitujen kausien keskimääräinen pelikassan saldo kauden jälkeen on suurempi, kuin saldo oli kauden alussa. Kaikissa simuloinneissahan pelikassan saldo ennen jokaista kautta oli 1000,00 €. Maaliperustaista Elo-mallia ja pi-hockey -mallia simuloitiin käyttäessä päästiin keskimäärin aina voiton puolelle. Ensinmainitulla mallilla keskimääräinen voitto kautta kohti oli 0,31 euroa ja jälkimmäisellä mallilla 0,20 euroa.

**Taulukko 51.** Yhteen veto eri malleista.

Malli	Pelikassa keskimäärin	Pelikassa parhaalla kaudella
Poisson-malli	998.44 €	1,000.11 €
Poisson-malli, 3 ottelua	998.18 €	999.26 €
Poisson-malli, 4 ottelua	998.10 €	999.72 €
Poisson-malli, 5 ottelua	999.15 €	1,001.08 €
Poisson-malli, 6 ottelua	999.01 €	1,002.55 €
Perus-Elo-malli	999.47 €	1,000.40 €
Perus-Elo-malli, erilliset Elo-luvut	999.70 €	1,000.67 €
Perus-Elo-malli, painotettu	999.81 €	1,001.79 €
Maaliperust. Elo-malli	1,000.31 €	1,002.16 €
Maaliperust. Elo-malli, erilliset Elo-luvut	999.30 €	1,000.24 €
Maaliperust. Elo-malli, painotettu	999.77 €	1,001.32 €
Pi-hockey -malli	1,000.20 €	1,001.10 €
Aggregaatti-malli	999.95 €	1,000.66 €

Kun eri mallien ja niiden variaatioiden menestystä tarkastellaan kunkin mallin parhaan kauden näkökulmasta, tulokset ovat huomattavasti positiivisempia. Ainoastaan kahden mallivariaation tapauksessa ei parhaallakaan kaudella saavutettu positiivista pelikassan kehitystä. Nämä mallivariaatiot ovat Poisson-mallin variaatioita, joissa maalikeskiarvoja laskettaessa otetaan huomioon vain 3-4 viimeisintä ottelua, kaikkien jo pelattujen

otteluiden sijaan. Kaikkein paras tulos saavutetaan kuitenkin Poisson-mallin variaatiolla, jossa otetaan huomioon kuusi viimeisintä ottelua maalikeskiarvoa laskettaessa. Tällöin saavutettiin 2,55 euron voitto parhaalla kaudella. Toiseksi paras tulos saavutettiin maaliperustaisella Elo-mallilla, jonka parhaan kauden pelikassan loppusaldo oli 1002,16 €.

## 6. Yhteenveto ja loppusanat

Tässä tutkimuksessa on sovellettu erilaisia menetelmiä jääkiekko-ottelun lopputuloksen ennustamisen mahdollistamiseksi algoritmin avulla. Itse algoritmi on toteutettu Java-ohjelmointikielellä ja se hyödyntää MySQL-tietokantaa. Tilastolliset menetelmät, joita algoritmi hyödyntää ja soveltaa, ovat riippumaton Poissonin jakauma, Elo-järjestelmä ja pi-rating -järjestelmä, joka perustuu Elo-järjestelmään. Itse algoritmin kehityksen tarkoituksena oli selvittää, onko mahdollista kehittää sovellus, joka pystyy ennustamaan jääkiekko-otteluiden lopputuloksia, ja voiko tuon järjestelmän avulla muodostaa voitokkaan urheiluvedonlyöntistrategian.

Teorian kannalta tämä tutkimus ei juuri tuota mitään uusia ideoita jo olemassa oleviin menetelmiin. Ainoastaan todennäköisyyksien johtamiseen Elo- ja pi-luvuista tässä tutkimuksessa kehitettiin uusia menetelmiä. Kuitenkaan mikään näistä menetelmistä ei sinänsä perustunut mihinkään olemassa olevaan teoriaan, vaan vain itse tutkimuksen tekijän intuitioon sekä aiempiin kokeiluihin tämän aihepiirin parissa. Tämän tutkimuksen hyödyt teorialle ovat sen sovellusalueessa: tähän mennessä vain muutamassa tapauksessa on tässä tutkimuksessa esitettyjä ja esiteltyjä menetelmiä sovellettu jääkiekkoon. Yleensä sovellusalueena on käytetty jalkapalloa. Toisaalta itse algoritmi on suunniteltu siten, että sitä voidaan hyödyntää myös esimerkiksi jalkapallo-otteluiden tulosten ennustamiseen. Tällöin vain tietokanta täytyy vaihtaa, ja uuden sovellusalueen tietokannan täytyy olla rakenteeltaan samanlainen kuin nykyinen, jääkiekkoon perustuva tietokanta. Mitään erityisesti vain jääkiekon soveltamiseen liittyvää algoritmista tai tietokannasta ei ole.

Ohjelmistotekniikan kannalta tämä tutkimus osoittaa, että Java-ohjelmointikieltä voidaan käyttää erilaisten tilastollisten sovellusten toteutuskielenä. Ainoastaan erittäin suuria lukuja käsitellessä Java-kielen tietorakenteiden rajallisuus voi aiheuttaa ongelmia, ainakin jos erittäin monimutkaisia matemaattisia kaavoja täytyy soveltaa.

Tutkimuksessa saadut tulokset eivät olleet sellaisia, joita ennen tuloksia odoteltiin ja myös toivottiin saatavan. Verrattuna esimerkiksi niihin tutkimuksiin, joista käytetyt mallit oli alunperin esitelty, tulokset olivat paljon vaatimattomampia sekä ennustamisen kannalta että vedonlyönnin näkökulmasta. Vain kaksi käytetyistä menetelmistä onnistui saamaan aikaan positiivisen pelikassan kehityksen keskimäärin jokaisella simuloitulla kaudella. Nämä menetelmät olivat pi-hockey sekä maaliperustainen Elo-malli. Kun tuloksia tarkastellaan vaan ennustamisen kannalta, eli kuinka usein onnistuttiin ennustamaan ottelun lopputulos oikein, onnistumisprosentit kaikilla malleilla vaihtelivat 41 prosentin ja 43 prosentin välillä. Jos huomioitiin vain kotivoitoiksi ennustetut ottelut, onnistumisprosentti parani hiukan, mutta oli silloinkin kuitenkin aina reilusti alle 50 %.

Tämän tutkimuksen tavoitteena oli selvittää, onko mahdollista kehittää algoritmi, joka pystyy ennustamaan jääkiekko-otteluiden lopputuloksia, ja voidaanko tätä algoritmiä hyödyntämältä kehittää voitokas vedonlyöntistrategia. Koska on käytännössä mahdotonta mitata esimerkiksi keskivertopelaajan onnistumisprosenttia otteluiden lopputulosten ennustamisessa, on myös hyvin hankalaa laittaa perspektiiviin sitä, kuinka

hyvin tutkimuksessa onnistuttiin, ja kuinka hyviä kehitetyn algoritmin saamat tulokset ovat. Jos ajatellaan, että pelattaisiin niin, että luotetaan puhtaaseen sattumaan, voitaisiin ajatella, että onnistumisprosentti pitkällä aikavälillä olisi noin 33 %. Tähän verrattuna luotu algoritmi onnistuu paremmin, koska sen saavuttamat tulokset otteluiden lopputulosten ennustamisessa vaihtelivat 41 prosentin ja 43 prosentin välillä.

Joiltain osin voitokas vedonlyöntistrategia onnistuttiin kehittämään, mutta saadut voitot rahallisesti yksittäisen kauden aikana olivat todella pieniä, parhaimmillaankin vain noin kaksi euroa. Toisaalta, mikään käytetty malli ja sen soveltamiseksi kehitetty algoritmi ei onnistunut viemään pelikassaa myöskään miinukselle: huonoimmillaankin pelikassa oli vain reilu neljä euroa pienempi kuin lähtötilanteessa, mitä voidaan pitää kuitenkin varsin hyvänä tuloksena.

Tällä tutkimuksella on jotain rajoituksia. Näistä ensimmäinen on, että tilastollisena materiaalina simuloinneille ja algoritmin testaukselle käytettiin vain Yhdysvaltojen NHL-liigan tilastoja ja tuloksia. Tälle valinnalle syynä oli lähinnä se, että tilastoja oli hyvin ja helposti saatavilla, sekä se, että verrattuna maailman muihin jääkiekkosarjoihin, NHL:ssä pelataan enemmän otteluita joukkuetta kohti kauden aikana, ja joukkueita on enemmän kuin muissa sarjoissa. Toinen rajoitus on käytettyjen mallien määrä ja monipuolisuus. Tässä tutkimuksessa käytettiin vain Poissonin jakaumaan perustuvaa mallia sekä kahta Elo-luku-järjestelmään perustuvaa mallia. Lisäksi käytettiin näiden kaikkien yhdistelmää. Tutkimuksen suppeuden vuoksi useita muita hyviä ja lupaavia malleja jouduttiin sulkemaan testauksen ulkopuolelle. Eräs rajoitus eri matemaattisten mallien käytölle oli myös tutkimuksen tekijän sangen rajoittunut matemaattinen ajattelykyky, joka jonkin verran haittasi käytettyjen mallien valintaa ja niiden parasta mahdollista soveltamista.

Tämä tutkimus avaa paljon uusia mahdollisuuksia jatkotutkimukselle. Tässä tutkimuksessa lyötiin vetoa vain niin, että pyrittiin ennustamaan päättyykö ottelu kotivoittoon, tasapeliin vai vierasvoittoon. Näin ollen vaihtoehtoisia lopputuloksia oli aina kolme. Jatkossa algoritmiä voitaisiin kehittää esimerkiksi niin, että vedonlyöntiä harrastettaisiin vain sellaisissa pelimuodoissa, joissa vaihtoehtoisia lopputuloksia on vain kaksi. Tällaisia ovat esimerkiksi niin sanottu handicap-vedonlyönti. Handicap-vedonlyönnissä vedonvälittäjä antaa ennakolta huonommalle joukkueelle 0,5 maalista 1,5 maaliin pohjalle ottelun alusta. Näin ollen ottelun alussa tilanne voi olla esimerkiksi 1,5-0 heikomman joukkueen eduksi. Toinen vaihtoehto on vastaavasti vähentää paremmaksi ennakoidulta joukkueelta maaleja, jolloin sillä on miinus-maaleja ottelun alkaessa. Huomioitavaa on myös, että annettu tasointus ei koskaan ole tasaluku, vaan esimerkiksi 1,5. Tällä varmistetaan, ettei ottelu voi päättyä tasan, koska jääkiekossa ei voi tehdä puolikasta maalia. Näin ollen myös lopputulosvaihtoehtoja on vain kaksi. Toinen menetelmä, jossa vaihtoehtoisia lopputuloksia on vain kaksi, on yhteismaaliveikkaus. Siinä lyödään vetoa, tuleeko ottelussa enemmän tai vähemmän maaleja kuin raja-arvo on. Yleensä raja-arvoksi on määritetty 5,5 tai 6,5, joissain tapauksissa jopa 7,5. Jälleen, koska ottelussa ei voi tehdä puolikkaita maaleja, jompi kumpi vaihtoehto on aina oikea.

Luonnollinen jatkotutkimuksen mahdollisuus muodostuu myös erilaisten tilastollisten ja todennäköisyyslaskennan menetelmien käytöstä. Esimerkiksi maximum likelihood -menetelmää on käytetty usein yhdessä muiden mallien kanssa parhaiden mahdollisten todennäköisyysarvioiden muodostamiseksi. Usein kuitenkin maximum likelihood -mallit ovat hyvin monimutkaisia, joten esimerkiksi Java-kielellä sellaisen algoritmin kehittäminen on todennäköisesti erittäin haastavaa. Myös ordered probit ja logit -mallit ovat varsin käytettyjä todennäköisyysarvioiden muodostamisessa. On olemassa myös erilaisia ohjelmistopaketteja- ja ympäristöjä, jotka ovat kehitetty pelkästään tilastollista laskentaa varten. Tällaisten pakettien hyväksikäyttö osana tässä tutkimuksessa

kehitettyä algoritmia mitä todennäköisimmin parantaisi saatuja todennäköisyysarvioita ja tarjoaisi myös hyvän jatkotutkimusaiheen.

## Lähteet

- Agnew, G. A. & Carron, A. V. (1994). *Crowd effects and the home advantage*. International Journal of Sports Psychology, 25. pp. 53-62,
- Albert, J., Bennett, J. & Cochran J. J. (2005). *Anthology of Statistics in Sports*. United States, American Statistical Association and the Society for Industrial and Applied Mathematics. pp. 183-185.
- Barnett, V. & Hilditch, S. (1993). *The Effect of an Artificial Pitch Surface on Home Team Performance in Football (Soccer)*. Journal of the Royal Statistical Society. Series A (Statistics in Society), 156(1). pp. 39-50.
- Cattelan, M., Varin, C. & Firth, D. (2013). *Dynamic Bradley-Terry modelling of sports tournaments*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 62(1). pp. 135-150.
- Chu, S. (2003). *Using Soccer Goals to Motivate the Poisson Process*. INFORMS Transaction on Education, 3(2). pp. 62-68.
- Clarke, S. R. & Norman, J. M. (1995). *Home Ground Advantage of Individual Clubs in English Soccer*. Journal of the Royal Statistical Society. Series D (The Statistician), 44(4). pp. 509-521.
- Constantinou, A. C. & Fenton, N. E. (2012). *Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries*. Journal of Quantitative Analysis in Sports, julkaisujankohda vielä tuntematon.
- Constantinou, A. C., Fenton, N. E. & Neil, M. (2012). *pi-football: A Bayesian network model for forecasting Association Football match outcomes*. Knowledge-Based Systems, 36. pp. 322-339.
- Dixon, M. J. & Coles, S. G. (1997). *Modelling Association Football Scores and Inefficiencies in the Football Betting Market*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 46(2). pp. 265-280.
- Elo, A. E. (1978). *The rating of chessplayers, past and present*. New York: Arco Publishing.
- Goddard, J. (2005). *Regression models for forecasting goals and match results in association football*. International Journal of Forecasting, 21. pp. 331-340.
- Greer, D. L. (1983). *Spectator Booing and the Home Advantage: A Study of Social Influence in the Basketball Arena*. Social Psychology Quarterly, 46(3). pp. 252-261.
- Hvattum, L. M. & Arntzen, H. (2010). *Using ELO ratings for match result prediction in association football*. International Journal of Forecasting, 26. pp. 460-470.

- Lock, R. H. & Danehy, T. J. (1997). *Using a Poisson Model to Rate Teams and Predict Scores in Ice Hockey*. Julkaisematon artikkeli.
- Maher, M. J. (1982). *Modelling association football scores*. Statistica Neederlandica, 36(3). pp. 109-118.
- Min, B., Kim, J., Choe, C., Eom, H. & McKay, R. I. (2008). *A compound framework for sports results prediction: A football case study*. Knowledge-Based Systems, 21. pp 551-562.
- Moroney, M. (1951). *Facts from figures*. London, Pelican.
- Nevill, A. M., Balmer, N. J. & Williams, M. A. (2002). *The influence of crowd noise and experience upon refereeing decisions in football*. Psychology of Sport and Exercise, 3. pp. 261-272.
- Oberhofer, H., Philippovich, T. & Winner, H. (2010). Distance matters in away games: Evidence from the German football league. Journal of Economic Psychology, 31. pp. 200-211.
- Reep, C., Pollard, R. & Benjamin, B. (1971). *Skill and chance in ball games*. Journal of the Royal Statistical Society. Series A (General), 134(4). pp. 623-629.
- Smith, R. D., Ciacciarelli, A., Serzan, J. & Lambert D. (2000). *Travel and the Home Advantage in Professional Sports*. Sociology of Sport Journal, 17. pp 364-385.
- Suzuki, A. K., Salaras, L. E. B., Leite, J. G. & Louzada-Neto, F. (2010). *A Bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup*. Journal of the Operational Research Society, 61. pp. 1530-1539.
- Vasama, P. & Vartia, Y. (1971). *Johdatus tilastotieteseen I*. Helsinki, Ylioppilastuki.
- Vuoksenmaa, J., Kuronen, A. & Nâls, J. (1999). *Urheiluviedonlyönti – Voittajan opas*. Helsinki, ESBC