



Ett analytiskt system för prognostisering av resultat av innebandymatcher

Robert Montell

Examensarbete
Informationsteknik
2023

EXAMENSARBETE	
Yrkeshögskolan Arcada	
Utbildningsprogram:	Informationsteknik
Identifikationsnummer:	8934
Författare:	Robert Montell
Arbetets namn:	Ett analytiskt system för prognostisering av resultat av innebandymatcher
Handledare (Arcada):	Andrey Shcherbakov
Uppdragsgivare:	
<p>Sammandrag:</p> <p>Examensarbetets mål var att bygga ett analytiskt system för att kunna förutspå framtida matcher i Finlands högsta innebandyliga som heter F-Liiga. Poisson distributionen är den huvudsakliga metoden som används för att förutsäga det kommande resultatet. Poisson distributionen är uppbyggd med hjälp av Excel och VBA. Poisson modellen kombineras med en dummymetod och med det får man det slutliga resultatet. En Dummy-spelare fungerar som en ersättning för en normalspelare som fattas från en specifik match. Med hjälp av dummyspelaren kan man förutspå ett exaktare resultat för en match.</p> <p>Oberoende av den lilla mängden data som samlas in från innebandy har ett fungerande slutresultat nåtts. Slutresultatet innehåller vinstsannolikheten för båda lagen på en match i framtiden. Resultatet tar i beaktandet vilka spelare som fattas från laguppställningen och hur en spelares frånvaro påverkar slutresultatet. Även om riktiga resultat har uppnåtts måste man komma ihåg att de finns hundratals variabler som påverkar slutresultatet på en innebandymatch och i verkligheten är det omöjligt att förutspå ett exakt resultat på en match.</p> <p>Modellen är avgränsad till endast en innebandyliga och till endast enkla statistik för att hålla projektet inom realistiska ramar. Statistikerna som tas i beaktandet är hur många mål en spelare gjort i genomsnitt per match.</p>	
Nyckelord:	Expected goals, Expected Result, Poisson modellen, Innebandy, F-Liiga
Sidantal:	29
Språk:	Svenska
Datum för godkännande:	

DEGREE THESIS	
Arcada University of Applied Sciences	
Degree Programme:	Information Technology
Identification number:	8934
Author:	Robert Montell
Title:	An analytical system for predicting results for Floorball games
Supervisor (Arcada):	Andrey Shcherbakov
Commissioned by:	
<p>Abstract:</p> <p>The aim of the thesis was to build an analytical system to predict future matches in Finland's highest floorball league called F-Liiga. The Poisson distribution is the main method used to predict the upcoming result. The Poisson model is built using Excel and VBA. The Poisson model is combined with a dummy method to get the final result. A dummy player acts as a replacement for a regular player who is missing from a specific match. Using the dummy player helps to predict a more accurate result for a match.</p> <p>Despite the small amount of data collected from floorball, a functioning result has been achieved. The result contains the probability of winning for both teams in a future match. The result considers which players are missing from the lineup and how a player's absence affects the final result. Even though real results have been achieved, it should be noted that there are hundreds of variables that affect the result of a floorball match and in reality, it is impossible to predict an exact result for a match.</p> <p>The model is limited to only one floorball league and only simple statistics to keep the project within realistic boundaries. The statistics considered are the average number of goals a player scores per match.</p>	
Keywords:	Expected goals, Expected Result, Poisson distribution, Floorball, F-Liiga
Number of pages:	29
Language:	Swedish
Date of acceptance:	

INNEHÅLL

1	Inledning.....	5
1.1	Bakgrund.....	5
1.2	Syfte & mål.....	5
1.3	Avgränsning.....	5
2	Teori	6
3	Metoder	7
3.1	Poisson distribution & VBA	7
3.2	Visual Studio Code & Python.....	7
3.3	Dummy metoden.....	8
4	Datainsamling.....	8
5	Byggandet av systemet.....	12
5.1	Laguppställningen.....	12
5.2	Spelarnas statistik	14
5.3	Poisson modellen	14
6	Resultat.....	19
7	Diskussion	25
	Källor.....	26
	Bilagor	27

FIGURER

Figur 1. Ett exempel på hur matcherna ser ut i tilastopalvelu.fi.....	8
Figur 2. Exempel på slutresultat av matcher från GameData fliken.	9
Figur 3. Exempel på Classics spelare från F-liiga sidan.....	10
Figur 4. Webbskrapning för namnen för spelaren.....	10
Figur 5. Slutresultatet i Excel på en del av spelarna i Classic.	11
Figur 6. Exempel på spelarstatistik	11
Figur 7. Slutresultatet på tilastot.py filen, visar statistik för Classics spelare.	12
Figur 8. En bild från tilastopalvelu.fi på laguppställningen mellan Classic och Nokian KRP.	13
Figur 9. Skärmbild från TeamData fliken med genomsnittet för målen.....	15
Figur 10. Bild på medeltalen av mål gjorda och släppta.	16
Figur 11. Ett exempel på förväntade resultatet mellan TPS och Happee.....	17
Figur 12. Poisson ekvationen.....	17
Figur 13. Exempel på hur POISSON.DIST funktionen fungerar.....	18
Figur 14. Exempel på en poisson distribution.	18
Figur 15. Exempel på olika slutresultat för innebandymatchen TPS – Happee.....	19
Figur 16. Flowchart för hur informationen rör sig i systemet	19
Figur 17. Alla matcher.....	20
Figur 18. Matchlänken i rätt format.	20
Figur 19. Laguppställningen från Classic – KRP.....	20
Figur 20. Spelarna för Nokian KRP med gult som fattas.....	21
Figur 21. Spelarna för Classic, spelarna med gult som fattas.....	22
Figur 22. Prognos på slutresultat för Classic – Nokian KRP.	23
Figur 23. Expected goals för Dummy och Lastikka.....	23
Figur 24. Nya resultat för matchen efter Lastikka fattas.	24
Figur 25. En confusion matrix på KRPs sista 6 matcher 22-23	25

1 INLEDNING

1.1 Bakgrund

Innebandy är en relativt liten sport i världen om man jämför med andra populära sporter. Men sanningen är den att innebandyn blir bara populärare hela tiden men tycks inte få tillräcklig uppmärksamhet och resurser för att kunna växa. Det leder till att sporten saknar viktiga egenskaper som redan har tagits i bruk i andra bollsporter, såsom fotboll.

Mängden data som samlas in från innebandymatcher är verkligen minimal. Det finns till exempel ingen statistik på hur länge ett visst lag har hållit bollen, vilket är mycket normalt i andra sporter. Egentligen nästan det enda data som samlas in är, skotten, målen och målvaktens parader. Att så lite data samlas in är mycket märkvärdigt eftersom sporten ändå har mer än två miljoner spelare.

Problemet är en ond cirkel eftersom det lilla professionella intresset för sporten innebär att det också finns begränsad mängd data som samlas in. Detta leder till att tillväxten av sporten stagnerar och inga framsteg sker.

1.2 Syfte & mål

Syftet med detta arbete är att bygga ett system som kan prognostisera resultat av framtida innebandymatcher på basen av förflutna matcher och den tillgängliga laguppställningen.

Målet med systemet är att få ut ett konkret svar (i siffror) på hur mycket en innebandymatch kommer att sluta och vinstsannolikheten för båda lagen.

1.3 Avgränsning

För att hålla systemet inom realistiska räckhåll är arbetet begränsat till bara en innebandyliga och till endast matcher från denna säsong (2022–2023). Den valda ligan är F-liiga det vill säga den högsta ligan i männens serie i Finland.

2 TEORI

Den nya omgående trenden i idrottsvärlden är ”expected goals”. Förväntade mål (eller xG) mäter kvaliteten på en målchans genom att beräkna sannolikheten för att den kommer att göras från en viss position på planen under en viss tid av matchen. Detta värde baseras på flera faktorer från innan skottet togs. xG mäts på en skala mellan noll och ett, där noll representerar en chans som är omöjlig att göra mål och ett representerar en chans som en spelare förväntas göra mål varje gång. (Withmore 2021).

Hur xG skiljer sig från detta examensarbete är att xG beräknar de förväntade målen under en match, medan examensarbetet försöker räkna ut de förväntade målen redan före matchen och beaktar inte alls matchens händelser.

I en innebandymatch spelar två lag motvarandra och en match räcker 60 minuter. Båda lagen har samtidigt fem utespelare på plan och en målvakt. Allt som allt får man ha 20 spelare med i laguppställningen. Detta består oftast av tre kedjor med totalt nio anfallare och sex backar och av tre bytesspelare som är på bänken (får oftast lite eller ingen speltid alls). Sedan har man en bytesmålvakt som hoppar in om första målvakten får en skada (eller spelar mycket dåligt och släpper in flera mål).

Ett innebandylag är uppbyggt av spelare som har olika nivåer av talang. Hur bra ett innebandylag är beror på flera faktorer som hur bra spelarna är i laget och hur bra de spelar ihop. Att spela en hemmamatch ger större sannolikhet att vinna matchen eftersom man inte behöver resa långa vägar till en match, publiken hejar på hemmalaget och kanske till och med buar till motståndarlaget. Alla dessa är bara några få exempel som påverkar ett framtida matchresultat, men alla faktorer är viktiga att ta i beaktandet.

För att förutspå resultatet för en match i framtiden finns det flera faktorer som påverkar slutresultatet. Det två viktigaste är förflutna matchers slutresultat och vilka spelare som är med i laggruppen för inkommande match. Om de tre bästa spelarna i laget som gör 90% av målen för laget är alla borta, kommer det uppenbart att påverka slutresultatet på ett negativt sätt.

3 METODER

För att nå en lösning till problemet används poisson distributionsmodellen för att förutspå framtida resultat och den kombineras med dummy metoden som tar i beaktande vilka spelare som fattas från laguppställningen. Dummy metoden baserar sig på poisson modellen och följer samma principer.

3.1 Poisson distribution & VBA

Den väsentligaste metoden som används i projektet är poisson distributions modellen, eller poissonfördelningen. Poissonfördelningen utvecklades av den franske 1800-talsmatematikern Siméon Denis Poisson. Det är en sannolikhetsteori som använder historiska data för att mäta sannolikheten för hur många gånger en händelse kommer att inträffa under en viss period. För att uttrycka det enkelt kan poissonfördelningen hjälpa en att förutsäga hur många mål det görs i matchen per lag. (Ebejer 2023).

Poissonfördelningen verkade lämplig för detta projekt eftersom den är ganska enkel att konstruera och är väldigt effektiv. Distributionen används ofta i andra sporter som analyserar samma saker som detta examensarbete.

Visual Basic for Application, alltså VBA, är ett programmeringsspråk som är utvecklad av Microsoft och är integrerad i de viktigaste Office Programmen som Word och Excel. (CFI Team 2023). VBA är ett utmärkt sätt för att lösa svårare problem i Excel som de vanliga Excel funktionerna inte klarar av. Med hjälp av VBA är det enkelt att automatisera enkla kalkyler i Excel. Excel i sig är redan ett mycket komplicerat och effektivt verktyg men om man vill lösa ännu aningen svårare problem är kombinationen av Excel och VBA det bästa alternativet.

VBA används i projektet för att automatisera beräkningen av medeltal i Excelfilen.

3.2 Visual Studio Code & Python

Visual Studio Code är en öppen källkods utvecklingsmiljö skapad av Microsoft. Flera miljoner människor använder VS Code och är mycket populärt och lätt att använda för programmerare. (Heller 2022). Programmet är gratis att använda och stöder flera hundra

programmeringsspråk med hjälp av en kompilator. En av dessa språk är uppenbart programmeringsspråket Python.

Huvud programmeringsspråket i detta arbete är Python. Python är ett av de lättaste programmeringsspråken och allra populäraste bland datavetenskaper. (Gallinelli 2021). Språkets enkelhet och den stora mängden olika bibliotek som NumPy och Pandas, gör Python till det bästa alternativet för val av programmeringsspråk för detta examensarbete.

Python används i projektet för att webskrapa spelarna från de olika lagen och spelarnas statistik.

3.3 Dummy metoden

Sista metoden i projektet är dummy metoden. Dummy, alltså något som fungerar som ersättning för det riktiga objektet. I detta fall används en dummispielare för att ersätta ut en spelare som är borta från en match.

4 DATAINSAMLING

Det finns tre olika data som samlas in: hur alla lagens matcher har slutat under säsongen, spelarnas namn och övriga data från alla spelare, som mål, passningar och poäng. Det behövs tre olika Excel filer för detta. Först samlas in data från alla lagens matcher från denna säsong (22–23). Data samlas in från tilastopalvelu.fi och läggs in till den första Excelfilen in i GameData fliken.

Ottelu	Päivä	Aika	Paikka	Koti	Vieras	Tulos	
3468	15.09.2022	18:30	Energia Areena 1	EräViikingit	- TPS	3-9	Seuranta
3271	16.09.2022	18:30	Kupittaaan ph 1	FBC Turku	- EräViikingit	2-8	Seuranta
3272	16.09.2022	18:30	Nurmohalli 1	Jymy	- OLS	10-2	Seuranta
3273	16.09.2022	18:30	Raholan lk 1	Nokian KrP	- Indians	4-7	Seuranta
3274	16.09.2022	18:30	Lahti Energia Areena 1	LASB	- Happee	5-10	Seuranta
3275	17.09.2022	17:00	Ideapark Bläk Boks 1	Classic	- OLS	9-8	Seuranta

Figur 1. Ett exempel på hur matcherna ser ut i tilastopalvelu.fi.

Sedan måste all onödiga data som alla länkar och hemmaarena tas bort. Ända data som är viktigt är datumet på matchen, hemmalaget, borta laget och slutresultatet. Efter data klistras in blir det problem med slutresultatet, Excel tror det är ett datum. Annat problem som uppstår är att slutresultatet måste delas i två skilda delar; målen för hemmalaget och målen för bortalaget. Eftersom Excel tror att slutresultatet är ett datum kan det utnyttjas till att separera slutresultatet till två skilda siffror, med metoderna *Month()* och *Day()*. Sedan bytts siffrorna till formen *General* så fungerar allt.

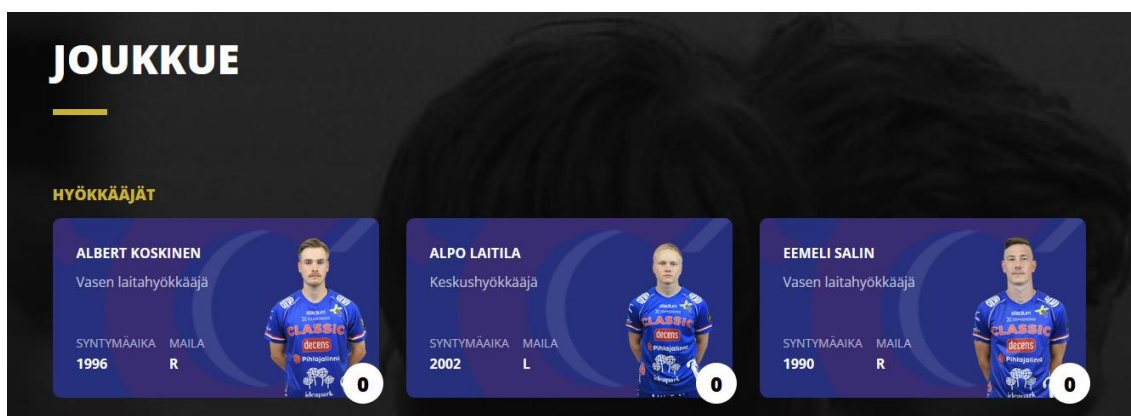
	A	B	C	D	E	F
1	Päivä	Koti	HG	Vieras	AG	
2	30.12.2022	Happee		5 Classic		7
3	30.12.2022	Jymy		2 SPV		4
4	30.11.2022	TPS		3 Classic		2
5	30.09.2022	Indians		5 LASB		5
6	30.09.2022	EräViikingit		5 Nokian KrP		9
7	30.09.2022	Classic		9 Oilers		1
8	30.01.2023	Nokian KrP		10 Jymy		4
9	29.01.2023	LASB		2 Indians		11
10	29.01.2023	TPS		3 OLS		3
11	29.01.2023	Oilers		9 SPV		3

Figur 2. Exempel på slutresultat av matcher från GameData filen.

Figur 2 är en skärmdump av GameData filen. Det finns fem kolumner med: Päivä; datumet på när matchen spelades, Koti; hemmalaget, HG; målen för hemmalaget, Vieras; bortalaget och AG; målen för bortalaget.

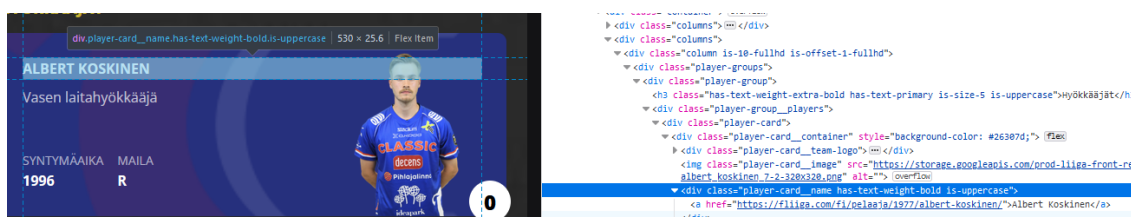
Andra delen av datainsamlingen är insamlingen av lagen och alla dess spelare. Följande del är gjord i Visual Studio Code med Python programmeringsspråket.

Spelarnas namn samlas in från F-ligans officiella sidor från “fliiga.com/joukkueet/pe-laajat”.



Figur 3. Exempel på Classics spelare från F-liiga sidan

Från figur 3 ser man exemplet på F-ligans sidor på hur de olika spelarna ser ut. Från inspectorn (F12) kan spelarna webbskrapas. Därifrån söks "div" som har klass namnet "player-card__name has-text-weight-bold is-uppercase". Då fås namnet på spelaren och sedan loopar man igenom alla spelaren för alla lag.



Figur 4. Webbskrapning för namnen för spelaren

Till slut läggs spelarna i en Excel sheet för att läsa dom lätt. Detta görs med en `xlsx.writer` där vi definierar Excelsn namn och worksheetens namn. Sedan loopas spelarna igenom från varje lag och skrivs in i Exceln.

	A	B	C	D	E	F	G	H	I
1	Players								
2	Albert Koskinen								
3	Alpo Laitila								
4	Eemeli Akola								
5	Eemeli Huhtakallio								
6	Eemeli Salin								
7	Eetu Sikkinen								
8	Janne Lamminen								
9	Jere Pulkkinen								
10	Juuso Woivalin								
11	Konsta Tykkyläinen								
12	Lassi Toriseva								

Figur 5. Slutresultatet i Excel på en del av spelarna i Classic.

Sista delen av datainsamlingen är statistiken på spelarna. Samma logik används som när spelarnas namn webskrapades men denna gång används en annan länk (fliiga.com/player-stats). Då fås spelarnas statistik.

PELAAJAN TILASTOT

Kausi: 2023 Valhe: PUDOTUSPELIT

Pelaaja	O	M	S	P	+	-	+/-	YVM	AVM	L	L%
Ville Lastikka	4	6	4	10	15	10	5	0	0	52	11.54
Oskari Heikkilä	4	3	3	6	14	7	7	0	0	23	13.04
Eetu Sikkinen	4	3	1	4	11	15	-4	0	0	11	27.27
Konsta Tykkyläinen	4	2	1	3	7	7	0	0	0	24	8.33
Nico Salo	4	1	2	3	4	14	-10	0	0	18	5.56
Alpo Laitila	4	1	2	3	8	7	1	0	0	7	14.29

Figur 6. Exempel på spelarstatistik

Denna gång samlas in hela kolumnen. Excelwriter används igen för att få data in i Excel. En ny Excel fil definieras med namnet stats.xlsx. Samma webskrapningsstrategi används för att få all data in i Excel.

Som kan ses från figur 6 finns det en hel del statistik. Den viktigaste statistiken för detta projekt är O = Ottelut, alltså hur många matcher en spelare har spelat under säsongen, M = Maalit, målen spelaren gjort och GPG = Goals per Game, alltså medeltalet för hur

många mål en spelare gör per match. I framtida utvecklingar kunde flera av statistiken användas till godo.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Pelaaja	O	M	S	P	+	-	+/-	YVM	AVM	L	L%	GPG
2	Sami Johansson	32	38	22	60	80	56	24	5	1	292	13.01	1.1875
3	Eemeli Salin	23	26	18	44	55	32	23	4	1	175	14.86	1.130435
4	Ville Lastikka	24	20	24	44	60	34	26	3	1	158	12.66	0.833333
5	Albert Koskinen	33	25	15	40	63	40	23	6	0	190	13.16	0.757576
6	Oskari Heikkilä	33	15	23	38	78	44	34	1	1	136	11.03	0.454545
7	Alpo Laitila	33	6	24	30	64	47	17	2	0	54	11.11	0.181818
8	Konsta Tykkyläinen	33	19	7	26	74	42	32	0	1	129	14.73	0.575758
9	Toni Salminen	33	10	10	20	73	50	23	0	0	67	14.93	0.30303
10	Eetu Sikkinen	23	8	12	20	61	29	32	0	0	56	14.29	0.347826
11	Ville Hietaranta	31	14	5	19	41	26	15	0	0	71	19.72	0.451613
12	Matias Vaajala	32	10	5	15	44	22	22	0	0	66	15.15	0.3125
13	Nico Salo	19	7	8	15	45	27	18	2	0	61	11.48	0.368421

Figur 7. Slutresultatet på tilastot.py filen, visar statistik för Classics spelare.

5 BYGGANDET AV SYSTEMET

Projektet består för i huvudsak av tre olika delar; laguppställningen, spelarnas statistik och poisson modellen. Alla dessa delar är viktiga för att skapa en stark helhet. I följande kapitel förklaras hur de olika delarna är uppbyggda.

5.1 Laguppställningen

I den första delen i utvecklingen av systemet behöver laguppställningen kännas till för en specifik match. Det viktigaste är inte vem som är med, utan vem som fattas, eftersom när ”xR” (Expected Result) räknas för en kommande match bör det vara känt vilken spelare som minskar på den med sin frånvaro.

	35 TORISEVA Lassi (AM) 1994 (189/90)		
4. kenttä	40 VÄNTTINEN Oliver 2003 (0/0)	14 KIMPANPÄÄ Miika 2003 (0/0)	18 WOIVALIN Juuso 1996 (178/80)
3. kenttä	68 SALMINEN Toni 1997 (0/0)	30 HELIN Vilho 2005 (0/0)	33 VAAJALA Matias 1999 (176/59)
2. kenttä	2 AKOLA Eemeli "A" 1999 (176/72)	4 TYKKYLÄINEN Konsta 1998 (194/89)	7 KOSKINEN Albert 1996 (186/79)
1. kenttä	10 HEIKKILÄ Oskari 1997 (177/80)	88 LAITILA Alpo 2002 (0/0)	20 SALO Nico "C" 1994 (182/78)
	11 JOHANSSON Sami 1993 (183/81)	12 SIKKINEN Eetu "A" 1997 (189/78)	66 LEIKKANEN Mikko 1988 (189/81)
1. kenttä	9 KOVANEN Joonatan 1999 (187/74)	70 RANTALA Joonas "C" 1999 (186/66)	44 LAAKSO Mikko 1998 (181/77)
2. kenttä	33 MATTILA Jere 2000 (180/75)	77 MÄKINEN Miska 1996 (187/76)	72 JOHANSSON Henri 1985 (174/73)
3. kenttä	4 KRUMINS Morics 1992 (188/80)	16 HYRKKÖNEN Kim 2000 (170/60)	79 AHOLA Juuso 2003 (0/0)
4. kenttä	71 PAULAHARJU Simo 1999 (187/76)	61 ALKKI Lauri 1996 (0/0)	88 JONAESON Nico 1994 (171/66)
	76 UKKONEN Eemil 1996 (0/0)	42 ARKKILA Olli 1998 (190/110)	39 VIITAKOSKI Valtteri "A" 1996 (188/80)
	64 KOISTINEN Jere 2002 (0/0)	13 EKLUND Emil 1996 (187/66)	10 SIMONEN Topias 2002 (0/0)
	22 TUOMALA Miro "A" (AM) 1996 (0/0)		

Figur 8. En bild från tilastopalvelu.fi på laguppställningen mellan Classic och Nokian KRP.

Första delen i analyserandet av laguppställningen gjordes redan i datainsamlings kapitlet då alltså namnen på spelarna i de olika lagen lades in i Excel. Följande del är då att ta spelarna från lagets Excel flik och jämföra dem med dagens matchens laguppställning och sedan se vem som fattas.

Först görs en variabel för att webbskrapa in dagens match. Länken som används är från tilastopalvelu.fi.

Som man kan se från figur 8 innehåller laguppställningen mycket onödig information, som spelarnas födelseår, längd, vikt, spelnummer och så vidare. Den nödvändiga informationen är endast förnamn och efternamn.

Sidan innehåller mycket information som är gömd så först rensas ut allt extra. Sedan webbskrapas alla spelaren och läggs in i `home_players[]` (alla hemmalagets spelare) och i `away_players[]` (alla borta lagets spelare). Efter det bör alla spelarens namn vara skrivna i gemener, utom förstås första bokstaven. Som ses från figur 8, är alla spelarnas hela efternamn skrivna med versaler.

Till sist ändras ännu ordningen från förnamn till efternamn så att det är i samma format som i Exceln så att allt fungerar tillsammans. Med det samma rensas alla extra bokstäver efter vissa spelare, "C" och "A", som betyder kapten och vice-kapten, se bilaga 2.

Då har slutresultatet nåtts och man kan printa ut hemma- och bortalaget i konsolen, men det viktigaste som redan nämndes i början är då att se vem som fattas.

5.2 Spelarnas statistik

Följande del är mestadels gjord redan i datainsamlingskapitlet men vissa saker måste ännu slutföras. Den viktigaste är Goals per Game kolumnen, eller GPG. Alla lagens spelares GPG räknas ihop och med detta fås information till dummy metoden, se kapitel 3.3. Dummy spelarens GPG blir då alltså medeltalet av alla lagens GPG. Då man vet dummy spelarens medelvärde kan modellen användas och se vilka spelare som fattas som har större värde än dummy spelaren, resten av spelarna som är då borta är irrelevanta. En spelare är irrelevant om den är sämre än dummy spelaren.

5.3 Poisson modellen

Den slutliga men också den viktigaste delen i systemet är poisson distributionen. Poisson distributionen är grunden för hela systemet. Modellen är byggt i Excel med hjälp av VBA. Till en början behövs tre olika flikar: PModel, TeamData och GameData. Första delen i

produktionen av modellen är redan gjord tidigare i datainsamlingskapitlet, då alla lagens tidigare matcher samlades till filen GameData.

Följande del är TeamData, dit samlas målgenomsnittet av föregående matcher. För varje lag har det individuellt räknas ett medeltal för hur många mål det släpper och hur många mål lagen gör per match, både hemma och borta. PAE: Points Away Earned, PAA: Points Away Allowed, PHE: Points Home Earned, PHA: Points Home Allowed. Dessa är alltså hur många mål som gjorts på hemmaplan (PHE), hur många mål som släppts hemma (PHA) och samma för bortamatcherna, släppts borta (PAA), gjorts borta (PAE). Denna del är gjord med VBA.

Kort sagt läser koden matchresultaten från GameData filen och sedan räknar den ihop medelvärden för målen som läggs in i TeamData filen.

	A	B	C	D	E	F
1	Team	PAE	PAA	PHE	PHA	
2	Classic	6.9375	3.9375	6.9375	4.75	
3	EräViikingit	4.529412	6.823529	4.133333	6.733333	
4	FBC Turku	3.6875	8.3125	3.1875	7.9375	
5	Happee	6.1875	4.1875	6.1875	4.4375	
6	Indians	5.733333	5.333333	5.235294	4.705882	
7	Jymy	4.176471	7.705882	5.133333	5.4	
8	LASB	3.6875	5.9375	4.25	6.5	
9	Nokian KrP	6.8	5	7.294118	4.411765	
10	Oilers	5.6875	4.6875	7	3.625	
11	OLS	4.647059	6.117647	5.933333	6.933333	
12	SPV	4.75	5.875	6.4375	4	
13	TPS	6.666667	3.4	5.823529	4.058824	
14						

Figur 9. Skärmdump från TeamData filen med genomsnittet för målen.

För att få den sista delen av distributionen fungera måste själva modellen göras. Till den användes Excels egna POISSON.DIST() funktionen.

Före POISSON.DIST() funktionen kan användas behövs först flera variabler. Första variablerna är "League Home Earned Average", "League Home Allowed Average", "League Away Earned Average" och "League Away Allowed Average." Då räknas alltså medeltalen för målen som gjorts och släppts hemma och borta för alla lag i hela serien.

Figur 10 visar för tillfället de olika medeltalen men efter det spelas flera matcher ändras medeltalen. Orsaken till att talen upprepas beror på att de är motsatser till varandra. När ett hemma lag gör mål till borta laget så läggs till det en siffra i Home Earned och alltså då också i Away Allowed och vice versa.

League Home Earned Avg	5.645833
League Home Allowed Avg	5.260417
League Away Earned Avg	5.260417
League Away Allowed Avg	5.645833

Figur 10. Bild på medeltalen av mål gjorda och släppta.

Efter medeltalen räknats ut så räknas ut så kallade maktsiffror för laget. Med maktsiffrorna ser vi hur laget jämför sig till genomsnittet av ligan. Först väljs ut två lag från ligan och sedan läggs vem som är hemma och vem borta. I figur 11 har TPS valts som hemma-lag och Happee som borta lag. TPS målgenomsnitt hemma är 5.82 (Home Earned) och hur många mål de släpper i genomsnitt är 4.05 (Home Allowed). För Happee är dessa tal 6.18 (Away Earned) och 4.18 (Away Allowed). På basen av detta kan styrkevärden (maktsiffra) för lagen räknas ut. Det första som vill jämföras är hur bra hemmalaget är på att göra mål jämfört med ligans genomsnitt. Med detta får vi Home Offense Strength. Formeln för ekvationen är alltså Home Earned Avg (5.82) dividerat med League Home Earned Average (5.64). Slutresultatet för TPS är alltså 1.03, se figur 11. Sedan görs samma ekvation för att räkna ut TPS Defence Strength och sedan samma för borta laget. Till slut fås båda lagens styrkevärden och med hjälp av dessa kan lagen jämföras med varandra och sedan kan matchresultatet förutspås.

För att räkna ut TPS expected goals (4.31, D2), tas först TPS Home Strength (1.03) gånger Happees Defence Strength (0.74) och sedan gånger League Home Avg (5.64) med varandra och sedan nås slutresultatet. Då kan jämföras hur bra TPS är på att göra mål med hur bra Happee är på att vara utan att släppa mål. Samma metod används sedan för att räkna expected goals för Happee.

	A	B	C	D
1	Home Team	Home Earned	Home Allowed	Expected
2	TPS	5.823529412	4.058823529	4.319297
3		Home Offense Strength	Home Defense Strength	
4		1.031473844	0.771578334	
5				
6	Away Team	Away Earned	Away Allowed	Expected
7	Happee	6.1875	4.1875	4.774141
8		Away Offense Strength	Away Defense Strength	
9		1.176237624	0.741697417	

Figur 11. Ett exempel på förväntade resultatet mellan TPS och Happee.

Efter man vet båda lagens hemma och borta defence och attack strength kan Poisson Ekvationen användas för att räkna ut sannolikheten för olika matchresultat. Från figur 12 ses ekvationen för poisson modellen, men som tur behövs inte hela modellen användas eftersom Excel har sin egen funktion för det.

The Poisson Formula is:

$$P(k \text{ events in interval}) = (\lambda^k e^{-\lambda}) / k!$$

In this formula:

- **P** is the probability
- **k** is the number of occurrences in the interval (number of goals)
- **λ** is the expected number of goals
- **e** is Euler's number (e = 2.71828...)
- **k!** is the factorial of k

Figur 12. Poisson ekvationen

För att få ett snyggt rutnät för de olika resultaten skrivs ut siffrorna från 0–25 i Exceln. Dessa tal representerar mål som kan förekomma i en innebandymatch. Sällan kommer det 25 mål men sådana matcher har också setts. Detta görs två gånger, målen för borta och för hemmalagen.

=POISSON.DIST(G\$11,\$D\$7,FALSE)*POISSON.DIST(\$F12,\$D\$2,FALSE)

	POISSON.DIST(x, mean, cumulative)	E	F	G	H	I	J	
	Home Allowed	Expected	League Home Earned Avg	5.645833				
12	4.058823529	4.319297	League Home Allowed Avg	5.260417				
th	Home Defense Strength		League Away Earned Avg	5.260417				
44	0.771578334		League Away Allowed Avg	5.645833				
	Away Allowed	Expected						
75	4.1875	4.774141						
th	Away Defense Strength							
24	0.741697417							
							Away	
				0	1	2	3	
				0	G\$11,	0.0537%	0.1281%	0.2038%
				1	0.0485%	0.2318%	0.5533%	0.8805%
66				2	0.1048%	0.5006%	1.1949%	1.9015%

Figur 13. Exempel på hur POISSON.DIST funktionen fungerar.

Till sist läggs in första alternativet av mål för bortalaget (G11) alltså 0 mål. Sedan medeltalet för bortalaget (4.77) och sedan om värden ska kumulera (FALSE) och detta tas gånger med samma data men för borta laget. Detta upprepas för hur många mål har valts (25) och sedan fås sannolikheterna för de olika målen. Slutresultatet ses från figur 14. Som ses från figur 14 är den sannolika slutresultaten för matchen TPS – Happee, 4-4.

	Away															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.0112%	0.0537%	0.1281%	0.2038%	0.2433%	0.2323%	0.1848%	0.1261%	0.0752%	0.0399%	0.0191%	0.0083%	0.0033%	0.0012%	0.0004%	0.0001%
1	0.0485%	0.2318%	0.5533%	0.8805%	1.0509%	1.0034%	0.7984%	0.5445%	0.3250%	0.1724%	0.0823%	0.0357%	0.0142%	0.0052%	0.0018%	0.0006%
2	0.1048%	0.5006%	1.1949%	1.9015%	2.2695%	2.1670%	1.7243%	1.1760%	0.7018%	0.3723%	0.1777%	0.0771%	0.0307%	0.0113%	0.0038%	0.0012%
3	0.1510%	0.7207%	1.7204%	2.7377%	3.2676%	3.1200%	2.4825%	1.6931%	1.0104%	0.5360%	0.2559%	0.1111%	0.0442%	0.0162%	0.0055%	0.0018%
4	0.1630%	0.7782%	1.8577%	2.9563%	3.5284%	3.3690%	2.6807%	1.8283%	1.0911%	0.5788%	0.2763%	0.1199%	0.0477%	0.0175%	0.0060%	0.0019%
5	0.1408%	0.6723%	1.6048%	2.5538%	3.0481%	2.9104%	2.3158%	1.5794%	0.9425%	0.5000%	0.2387%	0.1036%	0.0412%	0.0151%	0.0052%	0.0016%
6	0.1014%	0.4840%	1.1553%	1.8384%	2.1942%	2.0951%	1.6671%	1.1370%	0.6785%	0.3599%	0.1718%	0.0746%	0.0297%	0.0109%	0.0037%	0.0012%
7	0.0626%	0.2986%	0.7128%	1.1344%	1.3539%	1.2928%	1.0287%	0.7016%	0.4187%	0.2221%	0.1060%	0.0460%	0.0183%	0.0067%	0.0023%	0.0007%
8	0.0338%	0.1612%	0.3849%	0.6125%	0.7310%	0.6980%	0.5554%	0.3788%	0.2260%	0.1199%	0.0572%	0.0248%	0.0099%	0.0036%	0.0012%	0.0004%
9	0.0162%	0.0774%	0.1847%	0.2939%	0.3508%	0.3350%	0.2665%	0.1818%	0.1085%	0.0575%	0.0275%	0.0119%	0.0047%	0.0017%	0.0006%	0.0002%
10	0.0070%	0.0334%	0.0798%	0.1270%	0.1515%	0.1447%	0.1151%	0.0785%	0.0469%	0.0249%	0.0119%	0.0052%	0.0020%	0.0008%	0.0003%	0.0001%
11	0.0027%	0.0131%	0.0313%	0.0499%	0.0595%	0.0568%	0.0452%	0.0308%	0.0184%	0.0098%	0.0047%	0.0020%	0.0008%	0.0003%	0.0001%	0.0000%
12	0.0010%	0.0047%	0.0113%	0.0179%	0.0214%	0.0204%	0.0163%	0.0111%	0.0066%	0.0035%	0.0017%	0.0007%	0.0003%	0.0001%	0.0000%	0.0000%
13	0.0003%	0.0016%	0.0037%	0.0060%	0.0071%	0.0068%	0.0054%	0.0037%	0.0022%	0.0012%	0.0006%	0.0002%	0.0001%	0.0000%	0.0000%	0.0000%
14	0.0001%	0.0005%	0.0012%	0.0018%	0.0022%	0.0021%	0.0017%	0.0011%	0.0007%	0.0004%	0.0002%	0.0001%	0.0000%	0.0000%	0.0000%	0.0000%
15	0.0000%	0.0001%	0.0003%	0.0005%	0.0006%	0.0006%	0.0005%	0.0003%	0.0002%	0.0001%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
16	0.0000%	0.0000%	0.0001%	0.0001%	0.0002%	0.0002%	0.0001%	0.0001%	0.0001%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
17	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
18	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
19	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
20	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
21	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
22	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
23	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
24	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
25	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%

Figur 14. Exempel på en poisson distribution.

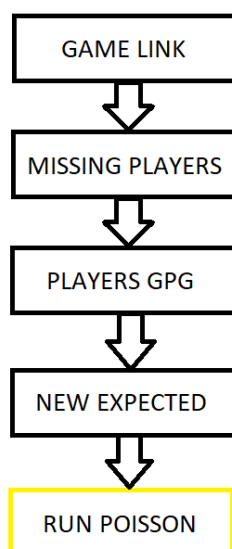
Efter poisson distributionen räknats ut kan vinstsannolikheterna för båda lagen räknas ut. Sannolikheten för att TPS eller hemmalaget i detta exempel vinner räknas med att summera ihop alla tal i distributionen där hemma målen (på vänster) är högre än målen för borta laget (uppe). Till denna används Excel funktionen `SUM()` och sedan målas ut alla värden som uppfyller kravet. Sedan görs samma sak men för att spelet ska sluta jämnt, alltså alla slutresultat där båda lagen gör lika antal mål. Till sist räknas ut vinstsannolikheten för borta laget och eftersom sannolikheten för hemmavinst och jämt spel redan är känt, kan ekvationen $1 - (\text{hemmavinst} + \text{jämt spel})$ användas för att få ut vinsten för borta laget.

39	1 Predicted	37.4468%
40	X Predicted	13.2809%
41	2 Predicted	49.2724%

Figur 15. Exempel på olika slutresultat för innebandymatchen TPS – Happee.

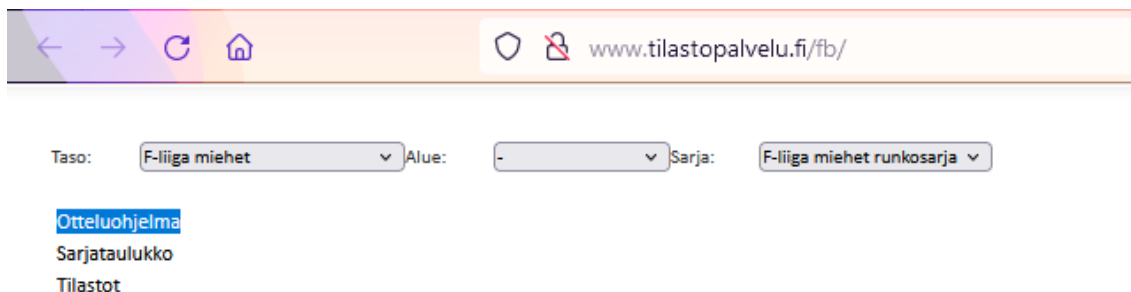
6 RESULTAT

I detta kapitel går igenom hur systemet hänger ihop som en helhet och hurdana resultat systemet producerar.



Figur 16. Flowchart för hur informationen rör sig i systemet

Första delen i systemets användning vill en match väljas som kommer att spelas i framtiden. Allra bäst är det att använda systemet två timmar före matchens start då laguppställningen för båda lagen måste vara publicerade. I detta exempel används matchen mellan Classic och Nokian KRP som spelades den 11.3.



Figur 17. Alla matcher

Det första som görs är sökningen av matchlänken. Alla matcher hittas från tilastopalvelu.fi under "F-liiga miehet", "F-liiga miehet runkosarja" och "Otteluohjelma". Sedan väljs matchen från "seuranta" och "kokoonpanot". Sedan klistras matchlänken in i teams.py filen in i kokoonpanot variabeln.



Figur 18. Matchlänken i rätt format.

Efter det körs systemet och konsolen visar vilka spelare som är med och vem som fattas från laguppställningen.

```
Kotijoukkue tänään: ['Oskar Sillanpää', 'Lassi Toriseva', 'Eemeli Akola', 'Otto Kilpi', 'Konsta Tykkyläinen', 'Albert Koskinen', 'Eemeli Salin', 'Oskari Heikkilä', 'Sami Johansson', 'Eetu Sikkinen', 'Miika Kimp anpää', 'Juuso Woivalin', 'Nico Salo', 'Vilho Helin', 'Matias Vaajala', 'Oliver Vanttinen', 'Ville Hietar anta', 'Mikko Leikkanen', 'Toni Salminen', 'Alpo Laitila']
Vierasjoukkue tänään: ['Miro Tuomala', 'Kaapo Kettunen', 'Morics Krumins', 'Joonatan Kovanen', 'Topias Si monen', 'Emil Eklund', 'Kim Hyrkkönen', 'Jere Mattila', 'Valtteri Viitakoski', 'Olli Arkkila', 'Mikko Laa kso', 'Lauri Alkki', 'Jere Koistinen', 'Joon Rantala', 'Simo Paulaharju', 'Henri Johansson', 'Eemil Ukko nen', 'Miska Mäkinen', 'Juuso Ahola', 'Nico Jonaeson']
Missing home: ['Eemeli Huhtakallio', 'Janne Lamminen', 'Jere Pulkkinen', 'Markus Vainionpää', 'Niilo Hel in', 'Otso Auvinen', 'Teemu Karppanen', 'Valtteri Molkka', 'Viljami Virtanen', 'Ville Lastikka']
Missing away: ['Elias Partanen', 'Jarno Laukkanen', 'Jorens Malkavs', 'Luka Ainoa', 'Mikael Ollikainen', 'Miki Saarinen', 'Oskari Tenho', 'Sami Saarikko', 'Vili Liponen']
```

Figur 19. Laguppställningen från Classic – KRP.

Följande del är att kolla vem av spelarna som påverkar framtida resultatet i matchen. Detta kollas med hjälp av dummy metoden. I dummy metoden har medeltalet för alla lags

alla spelares Goals per Game (GPG) räknats ut, alltså hur många mål en spelare gör i genomsnitt per match. Eftersom man bara får ha endast 20 spelare med i laguppställningen betyder det att fast fem spelare skulle fattas från laguppställningen betyder det inte att de påverkar slutresultatet på något sätt. De kan till exempel vara spelare som inte gjort ett enda mål under säsongen.

Då man vill se om en spelare som fattas påverkar laguppställningen jämförs spelarens GPG med Dummy spelarens GPG. Dummy spelarens GPG är 0,2318. Så om en spelare som fattas har mindre GPG än 0,2318 tas spelaren inte i beaktandet.

Oskari Tenho	17	1	1	2	4	3	1	0	0	8	12.5	0.058824
Jere Koistinen	28	1	1	2	4	12	-8	0	0	21	4.76	0.035714
Luka Ainoa	13	0	0	0	4	6	-2	0	0	3	0	0
Miro Tuomala	11	0	0	0	0	0	0	0	0	0	0	0
Jarno Laukkanen	20	0	0	0	10	7	3	0	0	7	0	0
Sami Saarikko	4	0	0	0	0	0	0	0	0	0	0	0
Miki Saarinen	27	0	0	0	0	0	0	0	0	1	0	0
Jorens Malkavs	1	0	0	0	0	1	-1	0	0	0	0	0
Mikael Ollikainen	1	0	0	0	0	0	0	0	0	0	0	0

Figur 20. Spelarna för Nokian KRP med gult som fattas.

Först öppnas stats.xlsx för att se statistiken för spelarna. För Nokian KRP har spelarna målats med gult som fattas. Spelaren Vili Liponen har inte spelat en enda match under säsongen så han syns inte med i stats.xlsx. GPG syns helt i högra sidan av bilden, som det märks uppfyller ingen av spelarna minimikravet för 0,2318 GPG. Detta betyder att ingen av Nokian KRPs spelare som fattas från laguppställningen är väsentlig för laget. I praktiken betyder detta att KRPs xG eller xR (Expected Result) kommer inte att ändras i poisson distributionen senare.

Pelaaja	O	M	S	P	+	-	+/-	YVM	AVM	L	L%	GPG
Sami Johansson	32	38	22	60	80	56	24	5	1	292	13.01	1.1875
Eemeli Salin	23	26	18	44	55	32	23	4	1	175	14.86	1.130435
Ville Lastikka	24	20	24	44	60	34	26	3	1	158	12.66	0.833333
Albert Koskinen	33	25	15	40	63	40	23	6	0	190	13.16	0.757576
Oskari Heikkilä	33	15	23	38	78	44	34	1	1	136	11.03	0.454545
Alpo Laitila	33	6	24	30	64	47	17	2	0	54	11.11	0.181818
Konsta Tykkyläinen	33	19	7	26	74	42	32	0	1	129	14.73	0.575758
Toni Salminen	33	10	10	20	73	50	23	0	0	67	14.93	0.30303
Eetu Sikkinen	23	8	12	20	61	29	32	0	0	56	14.29	0.347826
Ville Hietaranta	31	14	5	19	41	26	15	0	0	71	19.72	0.451613
Matias Vaajala	32	10	5	15	44	22	22	0	0	66	15.15	0.3125
Nico Salo	19	7	8	15	45	27	18	2	0	61	11.48	0.368421
Eemeli Akola	33	6	7	13	72	43	29	0	0	37	16.22	0.181818
Otto Kilpi	26	7	4	11	33	21	12	0	0	51	13.73	0.269231
Mikko Leikkanen	12	4	5	9	28	16	12	2	0	28	14.29	0.333333
Valtteri Molka	32	3	4	7	11	11	0	0	0	19	15.79	0.09375
Jere Pulkkinen	20	2	5	7	20	16	4	0	0	13	15.38	0.1
Viljami Virtanen	17	2	5	7	20	14	6	0	0	11	18.18	0.117647
Janne Lamminen	22	3	3	6	34	40	-6	0	0	38	7.89	0.136364
Oliver Vanttinen	7	1	0	1	2	3	-1	0	0	3	33.33	0.142857
Lassi Toriseva	32	0	1	1	0	0	0	0	0	1	0	0
Vilho Helin	10	0	1	1	8	3	5	0	0	14	0	0
Miika Kimpanpää	18	0	1	1	8	11	-3	0	0	5	0	0
Niilo Helin	18	0	1	1	8	14	-6	0	0	6	0	0
Juuso Woivalin	19	0	0	0	2	5	-3	0	0	1	0	0
Oskar Sillanpää	33	0	0	0	0	0	0	0	0	0	0	0
Eemeli Huhtakallio	1	0	0	0	0	0	0	0	0	0	0	0
Otso Auvinen	1	0	0	0	0	0	0	0	0	0	0	0

Figur 21. Spelarna för Classic, spelarna med gult som fattas.

Sedan ses på Classics spelare. Då märks att det finns en spelare högt på listan som fattas, Ville Lastikka. Hans GPG är 0,83, vilket är högre än minimikravet. Detta betyder att när Lastikka är borta påverkar de Classics xG för inkommande matchen på ett negativt sätt. Inga andra spelare påverkar förväntade målen. (Figur 21).

Den sista, men den viktigaste delen i arbetet är att plugga in spelarna och lagen in i poisson modellen för att få ut konkreta svar. Efter Classics och Nokian Krps data läggs in i modellen kan man se att det förutspådda slutresultatet är 6,14 till Classic och precis samma för KRP alltså 6,14. Vinst sannolikheten är också den samma med 44% och en chans på jämnt spel på 11%. (Figur 22).

Home Team	Home Earned	Home Allowed	Expected
Classic	6.9375	4.75	6.143911
	Home Offense Strength	Home Defense Strength	
	1.228782288	0.902970297	
Away Team	Away Earned	Away Allowed	Expected
Nokian KrP	6.8	5	6.140198
	Away Offense Strength	Away Defense Strength	
	1.292673267	0.885608856	
1 Predicted	44.2899%	Hemma vinst	
X Predicted	11.5042%	Jämnt spel	
2 Predicted	44.2060%	Borta vinst	

Figur 22. Prognos på slutresultat för Classic – Nokian KRP.

Denna data tar alltså inte i beaktandet att en av Classics bästa spelare är borta, Ville Lastikka. Det är det sista steget så får man ut det slutgiltiga svaret.

	Dummy	Player (Lastikka)
Dummy	0.231825	0.833333333
Expected	0.252276712	0.906850397
xR loss	0.654573685	
new Expected	5.489337754	

Figur 23. Expected goals för Dummy och Lastikka

För att räkna ut Lastikkas xG, används samma formel som för att räkna ut ett lags xG, denna gång tas bara Lastikkas GPG i beaktandet som redan räknades tidigare. Lastikkas GPG (mål per match) var då alltså 0,83. Formeln för xG blir då: $0,83 * 1,22 * 0,88$, alltså $GPG * \text{Home offense strength} * \text{Away defense strength}$. Som ses från figur 23, är Lastikkas xG, för matchen 0,906 mål (Expected).

Laget behöver ändå inte spela med en spelare mindre fast Lastikka är borta, utan en annan spelare kommer i stället för honom, denna är då Dummy Spelaren. Dummy spelarens GPG är 0,23 och Dummy spelarens xG är 0,25 (räknat med samma formel som för Lastikka). Då blir Classics xR (expected result) förlust $0,9 - 0,25$, alltså Lastikkas xG minus Dummyns xG, som blir 0.65. Detta tas sedan minus från den ursprungliga xR som

var 6,14, och lagets nya xR blir då $6,14 - 0,65 = 5,489$. Efter den nya xR pluggas in i poisson modellen ser svaret lite annorlunda ut.

Home Team	Home Earned	Home Allowed	Expected
Classic	6.9375	4.75	5.489338
	Home Offense Strength	Home Defense Strength	
	1.228782288	0.902970297	
Away Team	Away Earned	Away Allowed	Expected
Nokian KrP	6.8	5	6.140198
	Away Offense Strength	Away Defense Strength	
	1.292673267	0.885608856	
1 Predicted	36.6976%	Hemma vinst	
X Predicted	11.6266%	Jämmt spel	
2 Predicted	51.6758%	Borta vinst	

Figur 24. Nya resultat för matchen efter Lastikka fattas.

Då märks att Nokian KRPs vinstchans har ökat från 44% till 51,6% och Classics vinst sannolikhet har minskat från 44% till 36%. Då har lösningen på den matchen gjorts färdigt. Eftersom matchen redan spelades, så vet man slutresultatet vilket var då 3–4 för Nokian KRP. Kanske det var just Lastikkas frånvaro som var den avgörande faktorn för Classics förlust.

Målet med projektet har nåtts med att få ut konkreta vinstsannolikheter. Föregående exempel innehåller mycket varians eftersom de finns tusentals faktorer som påverkar matchens slutresultat. Att få ett fungerande system till 100% är omöjligt. Endast goda gissningar kan göras. Som ses från poisson-tabellen med de olika sannolikheterna för exakta slutresultat (figur 14) är sannolikaste slutresultatet 4–4, men den träffar endast 3,5% av tiden eller en av 28 matcher. Vinstsannolikheten är lättare att gissa, och blir lättare ju större skillnad de är på lagens nivå.

Nedan en matris (figur 25) på Nokian KRPs sista sex matcher av säsongen. Alla sex matcher förutspåddes på att KRP skulle vinna dem men i verkligheten vann de endast tre, spelade två matcher jämnt och förlorade en. Detta ger en liten blick på variansen av systemet och sporten.

		Predicted		
		Vinst	Jämnt	Förlust
Result	Nokian KRP			
	Vinst	3		
	Jämnt	2		
Förlust	1			

Figur 25. En confusion matrix på KRPs sista 6 matcher 22-23

7 DISKUSSION

Problem som förekommer med denna typ av förenklad modell är att till exempel Jere Niemelä (TPS) som är back inte gör så många mål men är ändå minst lika viktig spelare som Miko Kailiala som gör flera mål för TPS. Niemeläs betydelse kunde man se när han blev skadad för åtta matcher under säsongen och då förlorade TPS tre av de åtta matcherna, när TPS hade endast förlorat två matcher före det på det första 20 matcherna i säsongen. Modellen tar inte i beaktandet backar och andra som till exempel Peter Kotilainen från SPV som gjorde 40% av lagets poäng men hade för det mesta passningar och inte så många mål. Framtida utvecklingar kunde ta i beaktandet backar och ge deras mål till exempel dubbelt så stor betydelse, eller ge poäng för assist i systemet och på så sätt göra ut en modell som skulle beskriva den verkliga situationen bättre.

Annat problem är målvakternas betydelse, finns inget som helst system och ta i beaktandet om första målvakten är borta ut laguppställningen, men dessa är alla saker som kan göras i framtiden.

Dummy spelarens GPG borde säkert räknas individuellt för alla lag, för vissa lag kan få mera betydelse av att än spelare är borta med hjälp av dummy spelaren, än vad de skulle få i verkligheten. Till exempel laget LASBs målmedeltal per match var i genomsnitt 0,19 medan dummyns genomsnitt var 0,23. Också lag som Classic eller Nokian KRP får en större förlust med dummy spelaren när deras genomsnitt var 0,29 för deras spelare.

Huvudsakligen är jag nöjd med slutresultatet på projektet men i framtiden kunde man göra flera förbättringar. Förbättringarna skulle dock ändå göra så att man skulle få lite exaktare resultat ur programmet, men som i verkligheten inte har en stor betydelse på slutresultatet.

KÄLLOR

Corporate Finance Institute. (19 april 2023). *Excel VBA*. <https://corporatefinanceinstitute.com/resources/excel/excel-vba/>

Flatiron school. (4 mars 2021). *10 Best Data Science Programming Languages*. <https://flatironschool.com/blog/data-science-programming-languages/>

Fliiga. (2021). *2023 F-Liiga Miehet joukkueet*. <https://fliiga.com/miehet/tulospalvelu/2023/runkosarja/joukkuetilastot/>

Infoworld. (8 juli 2022). *What is Visual Studio Code? Microsoft's extensible code editor*. <https://www.infoworld.com/article/3666488/what-is-visual-studio-code-microsofts-extensible-code-editor.html>

Nhlfinns. (28 april 2023). *Salibandytilastot Miehet 1990–2023*. <https://nhlfinns.net/salibandy>

The Analyst. (24 juli 2021). *What Are Expected Goals (xG)?* <https://theanalyst.com/eu/2021/07/what-are-expected-goals-xg/>

The Punters Page. (26 april 2023). *Poisson Distribution in Sports Betting – A Step by Step Guide*. <https://www.thepunterspage.com/poisson-distribution-betting/>

Tilastopalvelu. (2013). *2023 F-liiga Miesten kauden ottelut*. http://www.tilastopalvelu.fi/fb/index.php?option=com_content&view=article&id=10&stgid=7000

Wagered On Tilt. (3 december 2021). *Poisson Distribution Model Microsoft Excel for Sports Betting* [Video]. Youtube. <https://youtu.be/SjD4NdHjF6w>

BILAGOR

```
def searchUrl(url):
    options = Options()
    options.headless = True
    driver = webdriver.Firefox(options=options, service=Service('D:\Geckodriver\geckodriver.exe'))
    driver.get(url)
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    driver.close()
    return soup
```

Kod 1. Funktionen för webbskrapning

```
# Reverse order of names to etunimi - sukunimi for compatibility,
and remove "C and A marks"

y = " ".join(missing_home)
x = y.split()
wordlist = []
for item in x:
    if len(item) > 1:
        wordlist.append(item)

y = " ".join(missing_away)
x = y.split()
wordlist2 = []
for item in x:
    if len(item) > 1:
        wordlist2.append(item)

z1 = [' '.join(x) for x in zip(wordlist[1::2], wordlist[0::2])]
#etunimi-sukunimi, kotijoukkueen kkp
z2 = [' '.join(x) for x in zip(wordlist2[1::2], wordlist2[0::2])]
#etunimi-sukunimi, vierasjoukkueen kkp
```

Kod 2. Omvänd ordning för namnen och rensning av extra tecken.

```

Sub findHomeAway()
Application.ScreenUpdating = False

Dim pModel As Worksheet
Dim teamData As Worksheet
Dim schedule As Worksheet

Set pModel = Worksheets("PModel")
Set teamData = Worksheets("TeamData")
Set schedule = Worksheets("GameData")

lRow = schedule.Range("A1").End(xlDown).Row
teamCount = teamData.Range("A1").End(xlDown).Row

'Sort by schedule home
schedule.Activate
schedule.Range("A1:E" & lRow).Sort Key1:=Range("B1"),
Key2:=Range("C1"), Header:=xlYes, _
Order1:=xlAscending, Order2:=xlAscending

For x = 2 To teamCount
    team = teamData.Range("A" & x).Value2
    teamStartRow = Application.WorksheetFunction.Match(team,
schedule.Range("B:B"), 0)
    teamEndRow =
(Application.WorksheetFunction.CountIf(schedule.Range("B:B"), team)
+ teamStartRow) - 1
    PHE = Application.WorksheetFunction.Average(schedule.Range("C"
& teamStartRow & ":C" & teamEndRow))
    PHA = Application.WorksheetFunction.Average(schedule.Range("E"
& teamStartRow & ":E" & teamEndRow))
    teamData.Range("D" & x).Value2 = PHE
    teamData.Range("E" & x).Value2 = PHA
Next

```

Kod 3. Exempel på VBA kod