

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS  
**MATEMATINĖS STATISTIKOS KATEDRA**

Ugnius Paukštys

(parašas)

**Krepšinio varžybų laimėtojo prognozavimas**

Magistro baigiamasis darbas

Vilnius, 2012

Darbo vadovas:  
doc., dr. Julius Kruopis

(parašas)

Recenzentas:  
prof.. Petras Vaitkus

(parašas)

Registracijos Nr.: .....  
Darbo gynimo data: .....

## TURINYS

<b>ĮVADAS</b> .....	<b>4</b>
<b>1. KREPŠINIO KOMANDOS CHARAKTERISTIKOS</b> .....	<b>5</b>
1.1. KREPŠINIO KOMANDOS ŽAIDĖJŲ CHARAKTERISTIKOS .....	5
1.2. KOMANDŲ STATISTINĖS CHARAKTERISTIKOS .....	6
<b>2. TURIMI DUOMENYS</b> .....	<b>7</b>
2.1. PRADINĖ DUOMENŲ LENTELE .....	7
2.2. TRIJŲ ANKSTESNIŲ VARŽYBŲ VIDUTINĖS CHARAKTERISTIKOS .....	9
<b>3. TEORINĖ DALIS</b> .....	<b>9</b>
3.1. LOGISTINĖ REGRESIJA.....	10
3.1.1. <i>Klasifikavimas</i> .....	12
3.2. NEURONINIAI TINKLAI .....	14
3.2.1. <i>Dirbtinio neurono modelis</i> .....	14
3.2.2. <i>Neuronų tinklo modelis</i> .....	15
3.2.3. <i>Neuroninio tinklo mokymas</i> .....	15
3.2.4. <i>Parceptronas</i> .....	16
<b>4. RUNGTYNIŲ LAIMĖTOJO PROGNOZAVIMAS</b> .....	<b>16</b>
4.1. LAIMĖTOJO PROGNOZAVIMAS SU LOGISTINE REGRESIJA (EUROLYGA).....	16
4.2. LAIMĖTOJO PROGNOZAVIMAS SU LOGISTINE REGRESIJA (VTB LYGA).....	21
4.3. LAIMĖTOJO PROGNOZAVIMAS NAUDOJANT NEURONINIUS TINKLUS (EUROLYGA) .....	23
4.3. LAIMĖTOJO PROGNOZAVIMAS NAUDOJANT NEURONINIUS TINKLUS (VTB LYGA).....	26
<b>5. LAŽYBŲ BENDROVĖS</b> .....	<b>26</b>
5.1. KOEFICIENTŲ NUSTATYMAS .....	27
5.2. TIKIMYBĖS .....	29
5.3. LAŽYBŲ PUNKTO KOEFICIENTAI .....	30
5.4. LAŽYBŲ KONTOROS TIKIMYBIŲ PALYGINIMAS SU TIK. ĮVERČIAIS BEI ŽAIDĖJO PELNO PASKAIČIAVIMAS .....	34
<b>IŠVADOS</b> .....	<b>40</b>
<b>SANTRAUKA</b> .....	<b>41</b>
<b>SUMMARY</b> .....	<b>42</b>
<b>LITERATŪRA IR ŠALTINIAI</b> .....	<b>43</b>

## Ivadas

Lietuvoje krepšinis visada buvo sporto šaka numeris vienas. Retas galėtų tai paneigti. Šis magistrinis darbas taip pat skirtas krepšiniui. Bakalauriniame darbe prognozavau krepšinio 2008-2009 metų stipriausios Europos krepšinio lygos varžybų baigtis.

Visi duomenys buvo surinkti oficialiose Eurolygos bei VTB lygos internetinėse svetainėse. Darbe analizuojami 2009-2010 m. Eurolygos sezono [1] bei 2011-2012 VTB (VTB bankas - pagrindinis lygos rėmėjas) lygos sezono [7] duomenys. Kaip ir baigiamajame bakalauro darbe prognozuojama buvo pažingsninės logistinės regresijos pagalba. Taip pat, papildomai, išbandyti neuroniniai tinklai prognozuojant rungtynių baigtį.

Likusioje darbo dalyje aprašyta lažybų bendrovių veikla, pateikti vienos lažybų kontoros siūlomi lažybiniai koeficientai už krepšinio lygos Eurolygos varžybų baigtis. Remiantis koeficientais bus paskaičiuotos lažybų kontoros siūlomos tikimybės laimėti vienai ar kitai komandai. Taip pat bus paskaičiuoti tikimybių įverčiai remiantis prognozuotais rezultatais.

Darbo paskutiniame skyrelyje bus bandoma statyti už vieną ar kitą komandą lyginant lažybų kontoros tikimybes ir paskaičiuotus tikimybių įverčius.

Atlikti tyrimai ir gautos išvados gali būti įdomios visiems Lietuvos gyventojams. Gautus rezultatus galėtų panaudoti netgi lažybų bendrovės tikslindamos savo bukmekerių nustatytas tikimybes įvairiems statymams.

Darbo pabaigoje yra pateiktos išvados, santrauka lietuvių ir anglų kalbomis, bei literatūros sąrašas.

## 1. Krepšinio komandos charakteristikos

### 1.1. Krepšinio komandos žaidėjų charakteristikos

Krepšinio varžybų metu yra vedamas protokolas. Jame yra užfiksuojama visa žaidėjų statistika bei visos komandos statistika. Protokole yra pažymami šie atskirų žaidėjų statistiniai duomenys: žaistų minučių skaičius; surinktų taškų skaičius; pataikytų ir pramestų metimų skaičių santykiai iš dviejų taškų zonos, iš trijų taškų zonos ir nuo baudų metimo linijos; atkovotų kamuolių skaičius puolime, gynyboje ir viso atkovotų kamuolių skaičius (atkovotų kamuolių puolime ir gynyboje suma); rezultatyvių perdavimų skaičius; perimtų kamuolių skaičius; padarytų klaidų skaičius; blokuotų metimų ir gautų blokų skaičiai; gautų ir uždirbtų pražangų skaičiai; naudingumo balų skaičius (Naudingumo balai = pelnytų taškų skaičius + atkovotų kamuolių skaičius + rezultatyvių perdavimų skaičius + perimtų kamuolių skaičius + blokuotų metimų skaičius + išprovokuotų baudų skaičius – pramestų metimų skaičius – klaidų skaičius – gautų blokų skaičius – gautų pražangų skaičius. (Taip jis skaičiuojamas Eurolygoje)).

Naudingumo balas geriausiai atspindi žaidėjo naudą komandai, nors dažniausiai visur yra skelbiama žaidėjų įmestų taškų statistika. Štai paprastas pavyzdys, parodantis, kad nebūtinai naudingiausias žaidėjas yra tas, kuris pelnė daugiausia taškų. Tarkime, vienas žaidėjas pelnė 21 tašką, pataikęs 8 metimus iš 20, atkovojo 2 kamuolius, atliko 1 rezultatyvų perdavimą, bet suklydo 5 kartus ir prasižengė 4 kartus. Jo naudingumo balas būtų lygus  $21+2+1-12-5-4=3$ . Kitas žaidėjas pelnė 15 taškų, pataikęs 6 metimus iš 8, atkovojo ir perėmė po 4 kamuolius, išprovokavo 2 varžovų pražangas ir vienas jo mestas metimas buvo blokuotas. Šio žaidėjo naudingumo balas lygus  $15+4+4+2-2-1=22$ . Taigi, nors antrasis žaidėjas ir pelnė mažiau taškų, tačiau taikliau mesdamas į krepšį ir su geresniais kitais statistiniais rodikliais jis buvo daug naudingesnis komandai už pirmąjį žaidėją.

Žaidėjo efektyvumo koeficientas – tai dar viena charakteristika, nusakanti žaidėjo naudą komandai. Šis koeficientas yra ne kas kita, kaip žaidėjo naudingumo balų skaičius padalintas iš jo žaistų minučių skaičiaus. Kuo koeficientas didesnis, tuo žaidėjas buvo naudingesnis komandai savo rungtyniavimo metu.

Ūgis, svoris ir amžius taip pat gali turėti lemiamos įtakos rungtynių pabaigoje. Vyresnis žaidėjas yra labiau patyręs ir mažiau jaudinasi rungtynių pabaigose. Tuo tarpu jaunesnis žaidėjas gali rungtyniauti ilgiau nepavargdamas. Pajėgiausių komandų sudėtyse bandoma suderinti perspektyvių jaunų žaidėjų energiją su patyrusių žaidėjų patirtimi. Pagal ūgį žaidėjai yra

skirstomi į pozicijas. Įžaidėjai – iki 195 cm. ūgio, gynėjai – 190 – 200 cm., lengvieji puolėjai – 195 - 205 cm., sunkieji puolėjai 198-210, centro puolėjai nuo 203cm. Tai tik apytiksliai ūgių vertinimai, nes kuo toliau, tuo mažiau pozicijos dedamos į kokius nors ūgių rėmus. Šių dienų krepšinyje yra ir įžaidėjų virš 200 cm. ūgio, ir tokių žaidėjų, kurie yra labai universalūs, kad gali žaisti vos ne visose pozicijose.

Yra ir įvairių kitokių charakteristikų, bet aš išvardinau šiuo metu labiausiai naudojamą krepšinio pasaulyje, be kurių neapsieina jokios aukščiausio rango varžybos.

## 1.2. Komandų statistinės charakteristikos

Kiekviena komanda, kaip ir žaidėjai, po rungtynių protokole turi savo statistiką. Ją labai nesunku apskaičiuoti. Tai tiesiog visų žaidėjų statistikų sumos. Pvz.: komanda pelnė tiek taškų, kiek jos visi žaidėjai pelnė kartu sudėjus; komandos atkovotų kamuolių skaičius lygus jos visų žaidėjų atkovotų kamuolių sumai ir t.t. Dažnai būna, kad keli vienos komandos žaidėjai besigrumdami atkovoja kamuolį. Tada atkovotas kamuolys neužrašomas jokiam žaidėjui, o tiesiog įrašomas, kaip komandos atkovotas kamuolys. Dėl to, kai kuriais atvejais, net ir susumavus kurią nors visų žaidėjų statistiką negausime, kad tai yra ir visos komandos statistika. Yra ir kitų labai svarbių komandoms charakteristikų. Apie komandos pajėgumą galima spręsti iš pergalių ir pralaimėjimų skaičių. Taip pat svarbus yra įmestų ir praleistų taškų skirtumas, kuris dažnai sezono pabaigoje turnyrinėje lentelėje tarp komandų, turinčių po vienodai pergalių ir pralaimėjimų, turi lemiamos įtakos komandų išsidėstymui. Komanda, kurios tas skirtumas didesnis, lentelėje „stovės“ aukščiau už mažesnę skirtumą turinčią komandą. Tai yra svarbu, nes atkrantamosiose varžybose tektų žaisti su sąlyginai silpnesne komanda. Atkreipti dėmesį verta ir ne tik į įmestų ir praleistų taškų skirtumą, bet atskirai ir į įmestus ir praleistus taškus viso sezono metu. Jie parodo komandos pajėgumus puolime ir gynyboje. Istorija rodo, kad dažniau čempionatus laimi ne geriausią puolimą demonstruojančios komandos, o komandos, kurios nepamiršta ir gynybos.

Komandos pergalės priklauso ne tik nuo žaidėjų meistriškumo, bet ir nuo gero trenerio. Geras treneris gali ir su silpnesne komanda nuveikti daugiau, nei silpnesnis treneris su pajėgesne komanda. Labai retai treneris pradeda treniruoti stiprią komandą prieš tai neturėjęs jokios patirties. Dažniausiai prieš tai jie treniruoja silpnesnes komandas apie 20 metų ir pasiekdamas daug pergalių gauna pasiūlymą iš pajėgesnių klubų.

Prie komandinių charakteristikų galime pridėti ir vidutinį komandos žaidėjų ūgį bei vidutinį komandos žaidėjų amžių. Visgi šios charakteristikos retai būna akcentuojamos, kaip turinčios

lemiamą įtaką pergalėms. Šiuolaikiniame krepšinyje aukštu lygiu žaidžiančių komandų žaidėjų ūgiai yra panašūs. Skiriasi tik jų meistriškumas. [5]

## 2. Turimi duomenys

### 2.1. Pradinė duomenų lentelė

Turime visus 2009-2010 metų Eurolygos sezono bei 2011-2012 VTB lygos sezono varžybų protokolus. Viso – 188 protokolai iš Eurolygos ir 157 iš VTB. Protokolai yra paimti iš oficialių šių lygų internetinių svetainių. Iš jų yra išrinkti visų žaidėjų naudingumo balai pagal pozicijas ir vietas startiniame penketuke bei komandos statistikos. Toliau aprašysiu visus duomenis ir kaip juos pažymėjau:

Kom1 – namuose žaidžiančios komandos pavadinimas;

Kom2 – svečiuose žaidžiančios komandos pavadinimas;

x1 – namų komandos starto penketo įžaidėjo naudingumo balų skaičius ;

x2 – namų komandos starto penketo gynėjo naudingumo balų skaičius ;

x3 – namų komandos starto penketo lengvojo puolėjo naudingumo balų skaičius ;

x4 – namų komandos starto penketo sunkiojo puolėjo naudingumo balų skaičius ;

x5 – namų komandos starto penketo centro puolėjo naudingumo balų skaičius ;

x6 – namų komandos atsarginio įžaidėjo naudingumo balų skaičius ;

x7 – namų komandos atsarginio gynėjo naudingumo balų skaičius ;

x8 – namų komandos atsarginio lengvojo puolėjo naudingumo balų skaičius ;

x9 – namų komandos atsarginio sunkiojo puolėjo naudingumo balų skaičius ;

x10 – namų komandos atsarginio centro puolėjo naudingumo balų skaičius ;

xx1 - namų komandos žaidėjų atkovotų kamuolių skaičius;

xx2 – namų komandos žaidėjų rezultatyvių perdavimų skaičius;

xx3 – namų komandos žaidėjų perimtų kamuolių skaičius;

xx4 – namų komandos žaidėjų padarytų klaidų skaičius;

xx5 - namų komandos žaidėjų blokuotų metimų skaičius;

xx6 - namų komandos žaidėjų išprovokuotų pražangų skaičius;

xx7 – namų komandos žaidėjų gautų pražangų skaičius;

y1 – svečių komandos starto penketo įžaidėjo naudingumo balų skaičius ;

y2 – svečių komandos starto penketo gynėjo naudingumo balų skaičius ;

y3 – svečių komandos starto penketo lengvojo puolėjo naudingumo balų skaičius ;

y4 – svečių komandos starto penketo sunkiojo puolėjo naudingumo balų skaičius ;  
 y5 – svečių komandos starto penketo centro puolėjo naudingumo balų skaičius ;  
 y6 – svečių komandos atsarginio įžaidėjo naudingumo balų skaičius ;  
 y7 – svečių komandos atsarginio gynėjo naudingumo balų skaičius ;  
 y8 – svečių komandos atsarginio lengvojo puolėjo naudingumo balų skaičius ;  
 y9 – svečių komandos atsarginio sunkiojo puolėjo naudingumo balų skaičius ;  
 y10 – svečių komandos atsarginio centro puolėjo naudingumo balų skaičius ;  
 yy1 - svečių komandos žaidėjų atkovotų kamuolių skaičius;  
 yy2 – svečių komandos žaidėjų rezultatyvių perdavimų skaičius;  
 yy3 – svečių komandos žaidėjų perimtų kamuolių skaičius;  
 yy4 – svečių komandos žaidėjų padarytų klaidų skaičius;  
 yy5 - svečių komandos žaidėjų blokuotų metimų skaičius;  
 yy6 - svečių komandos žaidėjų išprovokuotų pražangų skaičius;  
 yy7 – svečių komandos žaidėjų gautų pražangų skaičius;  
 l – rungtynių laimėtojas. 1 žymi, kad laimėtoja yra namų komanda, 0 – svečių komanda;  
 tsk – taškų skirtumas, kuriuo laimėjo ar pralaimėjo namuos žaidžiusi komanda.

xx7 ir yy7 nėra tarp VTB duomenų, nes šie duomenys nepateikti lygos protokoluose.

1 lentelė. Pradiniai statistiniai duomenys.

Kom1	Kom2	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	xx1	xx2	xx3	xx4	xx5	xx6	xx7
Cibona	MdP	6	12	-4	-2	6	-1	-8	0	3	-3	37	4	9	27	1	22	17

y1	Y2	y3	y4	y5	y6	y7	y8	y9	y10	yy1	yy2	yy3	yy4	yy5	yy6	yy7	l	tsk
23	10	15	4	7	5	4	1	34	12	32	21	19	14	2	17	22	0	-45

Pilna lentelė pateikta CD esančioje direktorijoje "Euroleague" Microsoft Excel faile duomenys.xls. Tokia pati duomenų lentelė yra ir VTB lygos (tik be xx7 ir yy7). Ją rasite CD esančioje direktorijoje "VTB lyga" Microsoft Excel faile duomenys.xls.

Jeigu kuris nors žaidėjas nežaidžia, tai jo naudingumo balas įrašomas, kaip lygus 0. Kartais būna, kad rungtyniauja 12 žaidėjų. Tada tų žaidėjų naudingumo balų skaičius pridedamas prie atitinkamai tos komandos atsarginio žaidėjo naudingumo balų, kuris žaidžia toje pačioje pozicijoje, kaip ir 11 ar 12 žaidėjas.

Pasinaudojus šiais duomenimis buvo sukurtos kita lentelė, kuri bus reikalinga taikant logistinės regresijos modelius.



## 2.2. Trijų ankstesnių varžybų vidutinės charakteristikos

Ši duomenų lentelė panaši į pirmąją savo struktūra ir elementų skaičiumi. Pagrindinis skirtumas – joje vietoj žaidėjo naudingumo balo įrašomas prieš tai jo žaistų trejų rungtynių naudingumo balų vidurkis. Pvz: jei žaidėjas pirmose trejose varžybose surinko atitinkamai 5, 7 ir 15 naudingumo balų, tai ketvirtų varžybų duomenų eilutėje jo naudingumo balas bus įrašytas kaip 9. Jeigu prieš tai buvo žaista mažiau trijų rungtynių, tai ir vidurkis imamas atitinkamai iš 2 ar 1 varžybų. Dėl to, šioje duomenų lentelėje nėra pirmo turo duomenų, nes prieš pirmą turą nebuvo varžybų. Taip pat ir su komandine statistika – yra imamos prieš tai buvusių trijų rungtynių statistikų vidurkiai. Šioje lentelėje yra du nauji duomenys l2 ir l3. Jie parodo kiek pergalių iškovoję komanda per prieš tai žaistas trejas rungtynes.

2 lentelė. Suvidurkinti trijų varžybų duomenys.

Kom1	Kom2	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	xx1	xx2	xx3	xx4	xx5	xx6	xx7
ASVEL	Ulker	8	-4	1	7	28	0	-4	0	5	3	32	7	9	11	3	16	14

l2	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	yy1	yy2	yy3	yy4	yy5	yy6	yy7	l3	l
0	8	-3	-2	12	-2	5	-2	-3	4	3	25	8	4	12	1	26	21	0	0

Pilna lentelė pateikta CD direktorijoje "Euroleague" Microsoft Excel faile 1.xls.

VTB lentelė atitinkamai VTB direktorijoje tuo pačiu pavadinimu.

Buvo sudaryta ir daugiau įvairių duomenų lentelių, bet apsiribosiu tik suvidurkintų trijų varžybų duomenų lentele, nes būtent šiuos duomenis naudojant su logistine regresija buvo gauti geriausi prognozavimo rezultatai. Neaprašytų duomenų lentelių pavyzdžių galite rasti mano bakalauriniame darbe.

## 3. Teorinė dalis

Duomenų analizei naudosiu logistinės regresijos metodus ir neuroninius tinklus. Pateikiu pagrindinius logistinės regresijos ir neuroninių tinklų faktus. ([2];[3];[6])

### 3.1. Logistinė regresija

Tarkime, kad stebimas įvykis  $A$ , kurio tikimybė priklauso nuo nepriklausomų kintamųjų (kovariančių)  $x_1, \dots, x_m$  reikšmių. Apibrėžkime atsitiktinį dydį  $Y$ , įgyjantį reikšmes 0 ir 1, tokį, kad  $Y \equiv 1$ , kai įvyksta įvykis  $A$ . Taigi galime sakyti, kad eksperimento metu stebimos a.d.  $Y$  reikšmės. Apibrėžkime tikimybę:

$$\pi(\mathbf{x}) = P\{Y = 1 \mid \mathbf{x}\} = P\{A \mid \mathbf{x}\}, \quad (1.1)$$

t. y. įvykio  $A$  sąlyginę tikimybę, kai kovariantės reikšmė yra  $\mathbf{x}$ ; čia  $\mathbf{x} = (x_0, x_1, \dots, x_m)^T$  kovariančių vektorius;  $x_0 \equiv 1$ . Modelio (1.1) atveju priklausomo kintamojo vidurkis, kai duotos kovariančių vektoriaus  $\mathbf{x}$  reikšmės, yra aprašomas  $E(Y \mid \mathbf{x}) = \pi(\mathbf{x})$ , kuris įgyja reikšmes iš intervalo (0, 1). Logistinės regresijos atveju vertinama įvykio  $A$  tikimybė  $\pi(\mathbf{x})$ , kai duotos  $\mathbf{x}$  reikšmės.

Logistinės regresijos (angl. logistic regression) modelis:

$$\log \text{it}(\mathbf{x}) = \ln \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m = \boldsymbol{\beta}^T \mathbf{x}. \quad (1.2)$$

Lygybę (1.2) galime užrašyti:

$$\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m) = e^{\boldsymbol{\beta}^T \mathbf{x}}. \quad (1.3)$$

Gauname

$$\pi(\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}} = \frac{e^{\boldsymbol{\beta}^T \mathbf{x}}}{1 + e^{\boldsymbol{\beta}^T \mathbf{x}}}.$$

Funkcijos  $\log \text{it}(\mathbf{x})$  apibrėžimo sritis yra  $\mathbf{R}^m$  ir su bet kuria realia  $\boldsymbol{\beta}$  reikšme funkcija  $\pi(\mathbf{x})$  įgyja reikšmes iš intervalo (0, 1).

Įvykio  $A$  pasirodymo ir nepasirodymo tikimybių santykis  $\pi(x)/(1 - \pi(x))$  vadinamas šansu (angl. odds). Tarkime, kad  $j$ -oji kovariantė yra tolydi. Imkime du kovariančių vektorius  $x^{(1)}$  ir  $x^{(2)}$ , kurių visos koordinatės, išskyrus  $j$ -ąją, lygios, o  $x_j^{(2)} = x_j^{(1)} + 1$ .

Naudodamiesi (1.3) gauname:

$$\frac{\pi(x^{(2)})/(1 - \pi(x^{(2)}))}{\pi(x^{(1)})/(1 - \pi(x^{(1)}))} = e^{\beta_j}.$$

Parametras  $e^{\beta_j}$  yra šansų santykis (angl. odds ratio). Jis parodo, kiek kartų pasikeičia įvykio  $A$  šansas, kai  $j$ -oji kovariantė padidėja vienetu, kitoms kovariantėms nepakitus.

Santykį  $\pi(x)/(1-\pi(x))$  galime naudoti objektams klasifikuoti. Jei šis santykis didesnis už vieną, tai reiškia, kad įvykio  $\{Y = 1\}$  tikimybė yra didesnė už įvykio  $\{Y = 0\}$ , kai kovariantės  $\mathbf{X}$  reikšmė lygi  $\mathbf{x}$ .

Tegu  $\eta = 1$ , jeigu objektas priklauso I klasei ir  $\eta = 0$ , jei objektas priklauso II klasei. Gauname klasifikavimo taisyklę;

$$\text{kai } \frac{\pi(x)}{1-\pi(x)} > 1, \text{ objektas priklauso I klasei.},$$

$$\text{kai } \frac{\pi(x)}{1-\pi(x)} < 1, \text{ objektas priklauso II klasei.}$$

Logaritmuodami gauname, kad ši taisyklė ekvivalenti tokiai:

$$\boldsymbol{\beta}^T \mathbf{x} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m > 0 \rightarrow \text{I klasė.}, \quad (1.4)$$

$$\boldsymbol{\beta}^T \mathbf{x} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m < 0 \rightarrow \text{II klasė.}$$

Kartais yra tikslinga vieną iš klaidų fiksuoti. Tada taisyklėje (1.4) vietoje 0 gali būti parinktas kitas slenkstis, t.y. objektas priskiriamas I ar II klasei priklausomai nuo to ar  $l(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{x} - c > 0$  ar  $l(\mathbf{x}) < 0$ ;  $l(\mathbf{x})$  vadinama diskriminantine funkcija.

Tarkime, kad turime tokius statistinius duomenis:

$$(Y_i; x_{0i}, x_{1i}, \dots, x_{mi}) = (Y_i; \mathbf{x}^{(i)}), i = 1, \dots, n,$$

čia  $Y_i$  įgyja dvi reikšmes 0 arba 1;  $x_{0i} \equiv 1$  (atitinka laisvąjį narį), o  $(x_{1i}, \dots, x_{mi})$  yra kovariančių matavimai i-ajam objektui.

Modelio parametrų  $\boldsymbol{\beta}$  įvertiniai randami didžiausiojo tikėtimumo metodu, t. y. minimizuojant tikėtimumo funkciją:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [\pi(\mathbf{x}^{(i)})]^{Y_i} [1 - \pi(\mathbf{x}^{(i)})]^{1-Y_i}. \quad (1.5)$$

Tarkime, kad  $\hat{\boldsymbol{\beta}}$  yra didžiausio tikėtimumo metodu gautas parametro  $\boldsymbol{\beta}$  įvertinys. Tada įvykio  $A$  sąlyginė tikimybė, žinant kovariantę  $\mathbf{x} = (1, x_1, \dots, x_m)^T$ , vertinama statistika

$$\hat{\pi}(\mathbf{x}) = \frac{e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}}}{1 + e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}}}.$$

Analogiškai įvertinsime šansų santykius  $e^{\beta_j}$ ,  $j = 0, 1, \dots, m$ , bei diskriminantinę funkciją:  $\hat{l}(\mathbf{x}) = \hat{\boldsymbol{\beta}}^T \mathbf{x} - c$ , kuria ir naudojame klasifikuodami objektus.

Tarkime, kad norime patikrinti hipotezę:

$$H : \beta_1 = \dots = \beta_m = 0, \quad (1.6)$$

t. y. regresijos modelis netinka (reikšmės  $\mathbf{x}$  žinojimas nepagerina tikimybės  $\pi(\mathbf{x})$  prognozės).

Tikėtumo santykio statistikos

$$D_E = -2 \ln \frac{L_0}{L(\hat{\beta})} = -2 \ln \frac{L(\tilde{\beta}_0, 0, \dots, 0)}{L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)} \quad (1.7)$$

skirstinys, kai hipotezė  $H$  teisinga, aproksimuojamas chi kvadrato skirstiniu su  $m$  laisvės laipsniais; vardiklyje – didžiausiojo tikėtumo funkcija, gauta įstačius didžiausiojo tikėtumo įvertinį  $\hat{\beta}$  į (1.5); skaitiklyje – modelio, kuriame  $\beta_1 = \dots = \beta_m = 0$ , didžiausio tikėtumo funkcija (1.5). Hipotezė (1.6) atmetama esant reikšmingumo lygmeniui  $\alpha$ , jei  $D_E > \chi_{1-\alpha}^2(m)$ .

Tarkime, kad norime patikrinti hipotezę:

$$H : \beta_j = 0,$$

kai alternatyva  $\bar{H} : \beta_j \neq 0$ . Kriterijaus statistikos

$$W = (\hat{\beta}_j)^2 / \hat{\sigma}_{jj},$$

kai hipotezė  $H$  teisinga, skirstinys aproksimuojamas chi kvadrato skirstiniu su 1 laisvės laipsniu; čia  $\hat{\sigma}_{jj}$  yra  $\hat{\beta}_j$  dispersijos įvertinys.

Parametrų  $\beta_1, \dots, \beta_m$ , tikimybės  $\pi(\mathbf{x})$  ir šansų santykių  $e^{\beta_j}$ ,  $j = 1, \dots, m$  pasiklovimo intervalai randami aproksimavimą normaliuoju skirstiniu.

Logistinės regresijos atveju determinacijos koeficientas

$$R^2 = 1 - \left( \frac{L_0}{L(\hat{\beta})} \right)^{2/n};$$

čia žymėjimai kaip (1.7). Koeficiento  $R^2$  maksimumas mažesnis už 1, t.y.  $R_{\max}^2 = 1 - (L_0)^{2/n}$ .

Koreguotasis determinacijos koeficientas

$$\tilde{R}^2 = \frac{R^2}{R_{\max}^2},$$

kurio maksimali reikšmė 1. Kuo determinacijos koeficiento reikšmė didesnė, tuo modelis geriau suderintas su duomenimis.

### 3.1.1. Klasifikavimas

Modelio tinkamumą įvykiams klasifikuoti galima apibūdinti klasifikavimo lentele, kurioje surašomos teisingų ir klaidingų sprendimų tikimybės

$$\alpha_{ij} = \mathbf{P}\{\eta = i | Y = j\}, \quad i, j = 0,1.$$

Praktiškai svarbesnės atvirkštinės tikimybės

$$\beta_{ji} = \mathbf{P}\{Y = j | \eta = i\} = \frac{\alpha_{ij}\omega_j}{\alpha_{i1}\omega_1 + \alpha_{i0}\omega_0}, \quad i, j = 0,1; \quad (1.8)$$

čia  $\omega_0$  ir  $\omega_1$  - klasių apriorinės tikimybės:  $\omega_0 = \mathbf{P}\{Y = 0\}$ ,  $\omega_1 = \mathbf{P}\{Y = 1\}$ ,  $\omega_0 + \omega_1 = 1$ .

Ne įvykių srauto dydis

$$Q = \mathbf{P}\{\eta = 1\} = \alpha_{10}\omega_0 + \alpha_{11}\omega_1. \quad (1.9)$$

Jeigu ne įvykių dalis  $\omega_1$  yra maža, tai įrašę formulėje (1.8) ir formulių (1.9) vardikliuose  $\omega_1 = 0$ ,  $\omega_0 = 1$  gausime apytiksles formules

$$Q \approx \alpha_{10}; \quad \beta_{10} \approx \omega_1 \frac{\alpha_{01}}{\alpha_{00}} = \omega_1 \gamma; \quad \beta_{11} \approx \omega_1 \frac{\alpha_{11}}{\alpha_{10}} = \omega_1 \delta, \quad (1.10)$$

kuriomis galime apibūdinti klasifikavimo tikslumą, net ir nežinodami tikslų tikimybių  $\omega_1$  ir  $\omega_0$  reikšmių.

Praktiškai tikimybės  $\alpha_{ij}$ ,  $i, j = 0,1$ , nežinomos, jas reikia vertinti iš statistinių duomenų. Pažymėkime  $V_{00}$  ir  $V_{10}$  kaip įvykių ir ne įvykių skaičių įvykių aibėje ( $V_{00} + V_{10} = N_0$ ); analogiškai  $V_{01}$  ir  $V_{11}$  - ne įvykių aibėje ( $V_{01} + V_{11} = N_1$ ). Naudodami šiuos duomenis gauname tikimybių  $\alpha_{ij}$  įvertinius

$$\hat{\alpha}_{ij} = \frac{V_{ij}}{N_j}, \quad i, j = 0,1.$$

Klasifikavimo rezultatus apibūdinantys skaičiai  $V_{ij}$  pateikti 1.1 lentelėje, o tikimybių  $\alpha_{ij}$  įvertiniai  $\hat{\alpha}_{ij}$  - 1.2 lentelėje.

Turėdami įvertinius  $\hat{\alpha}_{ij}$  gauname charakteristikų (1.10) įvertinius.

$$\hat{Q} = \hat{\alpha}_{10}; \quad \hat{\gamma} = \frac{\hat{\alpha}_{01}}{\hat{\alpha}_{00}}, \quad \hat{\delta} = \frac{\hat{\alpha}_{11}}{\hat{\alpha}_{10}}.$$

**1.1 lentelė.** Klasifikavimo rezultatai

	0	1	$\Sigma$
0	$V_{00}$	$V_{10}$	$N_0$
1	$V_{01}$	$V_{11}$	$N_1$

**1.2 lentelė.** Tikimybių  $\alpha_{ij}$  įvertiniai

	0	1	$\Sigma$
0	$\hat{\alpha}_{00} = \frac{V_{00}}{N_0}$	$\hat{\alpha}_{10} = \frac{V_{10}}{N_0}$	1
1	$\hat{\alpha}_{01} = \frac{V_{01}}{N_1}$	$\hat{\alpha}_{11} = \frac{V_{11}}{N_1}$	1

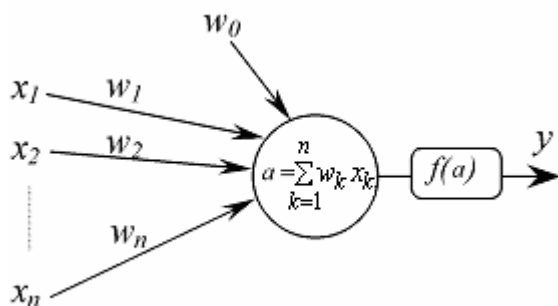
## 3.2. Neuroniniai tinklai

Dirbtiniai neuroniniai tinklai - tai informacijos apdorojimo struktūros, netiksliai imituojančios kai kuriuos gyvųjų organizmų smegenyse vykstančius informacijos apdorojimo procesus. Dirbtiniai neuroniniai tinklai sudaromi iš daugelio tarpusavyje sujungtų labai paprastų skaičiavimo elementų. Šie elementai, jungiami vieni su kitais įvairaus stiprumo jungtimis, yra apytikris biologinių neuronų modelis. Labiausiai viliojantis atrodo biologinių sistemų gebėjimas mokytis, prisitaikyti ir adaptuotis.

Mokymosi metu smegenyse keičiasi jungčių, siejančių neuronus, stiprumas. Toks jungčių stiprumo kitimas būdingas ir dirbtiniams neuroniniams tinklams. Jiems mokytis naudojami duomenų pavyzdžiai. Mokymo metu, duomenims veikiant dirbtinį neuroninį tinklą, specialiais algoritmais iteratyviai keičiami jungčių stiprumo koeficientai, vadinami svoriais. Informacija, reikalinga konkrečiam uždaviniui spręsti, sukaupiama svorių vertėse.

### 3.2.1. Dirbtinio neurono modelis

Neurono įėjime veikia keletas įėjimo signalų (verčių). Šie signalai gali būti viso neuroninio tinklo įėjimo signalai ar kitų tinklo neuronų išėjimo signalai. Kiekviena įėjimo jungtis turi savo perdavimo koeficientą (svorį). Kiekvienas neuronas turi jam priskirtą slenksčio vertę. Neurono sužadinimo signalas formuojamas skaičiuojant svorinę įėjimo signalų sumą ir atimant slenksčio vertę. Turint sužadinimo signalą ir naudojant neurono perdavimo funkciją, skaičiuojamas neurono išėjimo signalas (vertė). Neurono svoriai gali būti ir neigiami. Neigiamas svoris reiškia, jog jungtis turi slopinamąjį, bet ne žadinamąjį efektą.



1.1 pav

Labiausiai paplitusi dirbtinio neurono schema pavaizduota 1.1 pav. Čia  $x_1, \dots, x_n$  žymi neurono įėjimo signalus. Atitinkamai svoriai sužymėti  $w_1, \dots, w_n$ . Žymėjimas  $w_0$  reikškia slenksčio vertę,  $f()$  - perdavimo funkciją ir  $y$  - neurono išėjimą. Toliau labiau paplitusios neurono perdavimo funkcijos:

hiperbolinio tangento funkcija:  $f = \tanh(\text{net})$ ,

loginio sigmoido funkcija:  $f = 1/(1+\exp(-\text{net}))$

šiuolinė funkcija:  $f = 1$ , jei  $\text{net} \geq 0$  ir  $f = -1$ , jei  $\text{net} < 0$

tiesinė funkcija:  $f = \text{net}$ ;

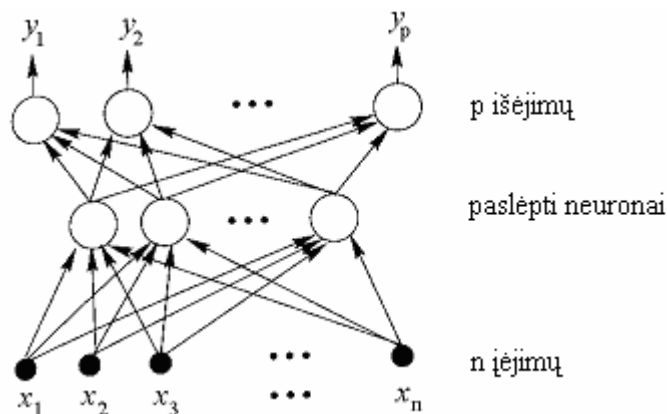
čia  $\text{net} = \sum_i w_i x_i$

Visos funkcijos būdingos tam tikro tipo neuroniniams tinklams, kitos priklauso nuo mokymo taisyklių arba parenkamos pagal sprendžiamą uždavinį.

### 3.2.2. Neuronų tinklo modelis

Atskiri neuronai jungiami į neuroninį tinklą, Bet kurios paskirties neuroninis tinklas turi įėjimus, perduodančius kintamųjų vertes iš išorės, ir išėjimus, formuojančius tinklo atsaką. Dažnai būna ir tarpinių (paslėptųjų) neuronų, skaičiuojančių vidinių tinklo funkcijų vertes. Įėjimo, paslėptieji ir išėjimo neuronai jungiami vieni su kitais.

Sakoma, kad dirbtinis neuroninis tinklas yra tiesioginio sklidimo, jeigu signalai sklinda iš įėjimų per visus paslėptus elementus ir pasiekia išėjimo neuronus.c



1.2 pav

### 3.2.3. Neuroninio tinklo mokymas

Įvairūs neuroninių tinklų tipai naudoja skirtingus mokymo metodus. Yra du pagrindiniai neuroninių tinklų apmokymo tipai: mokymas su mokytoju ir mokymas be mokytojo.

Neuroninis tinklas apmokomas naudojant mokymo imtį. Mokymo su mokytoju imtis sudaroma iš įėjimo verčių kartu su atitinkamomis išėjimo (užduoties) vertėmis. Neuroninis tinklas mokosi nežinomos priklausomybės tarp įėjimo ir išėjimo duomenų. Jei tinklas tinkamai apmokytas, jis gali modeliuoti funkciją, siejančią įėjimo ir išėjimo kintamuosius.

### 3.2.4. Parceptronas

Parceptronas yra skirtas klasifikavimo uždaviniams spręsti. Parceptrono išėjimo vertės yra 1 (TAIP) arba -1 (NE).

Parceptrono su fiksuotų funkcijų rinkiniu, pažymėtu  $\phi_j$ , po kurio eina adaptyviųjų svorių sluoksnis  $w_j$  ir slenkstinė išėjimo funkcija  $g()$ , perdavimo funkcija yra

$$y(x) = g\left(\sum_{j=0}^M w_j \phi_j(x)\right) = g(w^T \Phi(x)), \quad (1.11)$$

kur  $M$  yra fiksuotų funkcijų skaičius, o slenkstinė išėjimo funkcija

$$g(a) = -1, \text{ kai } a < 0 \text{ ir } g(a) = 1, \text{ kai } a \geq 0.$$

## 4. Rungtynių laimėtojo prognozavimas

Šiame skyriuje bus aprašyta tik dalis atliktos statistinės analizės. Aprašysiu tik tą logistinės regresijos modelį, kuriuo buvo gauti geriausi prognozavimo rezultatai. Daugiau modelių galite rasti mano baigiamajame bakalauriniame darbe. Analizė buvo atliekama SAS paketu. Apie SAS paketo kodus plačiau pasiskaityti galite [4] vadovėlyje. Toliau sekančiame skyriuje bus aprašytas teisingiausiai spėjantis logistinės regresijos modelis iš visų bandytų.

### 4.1. Laimėtojo prognozavimas su logistine regresija (Eurolyga)

Šis skyrelis skirtas Eurolygos komandų varžybų laimėtojo prognozavimui atsižvelgiant į prieš tai sužaistų varžybų statistines charakteristikas. Taigi, bus prognozuojama pagal 1.xls lentelės duomenis. Joje visi stulpeliai atitinka nepriklausomus kintamuosius, išskyrus komandų pavadinimus ir 1, kuris yra priklausomas kintamasis t.y., jis bus prognozuojamas. Rungtynių statistiniai duomenys čia surašyti nuo antrojo etapo, nes pirmojo etapo nebuvo galima prognozuoti dėl to, kad nebuvo jokių duomenų prieš tai, jog būtų galima prognozuoti. Iš visų bandytų logistinės regresijos modelių, daugiausia teisingų rungtynių baigčių suprognozavo pažingsninė logistine regresija. SAS yra realizuoti keli pažingsninės regresijos modeliai. Kokį



metodą naudoti, galime nurodyti procedūros LOGISTIC sakinyje MODEL su pasirinktimi SELECTION. Vienas jų - FORWARD metodas, kai yra pradama nuo nulio kintamųjų modelyje. Šiuo modeliu ir buvo gauti geriausi rezultatai. Pažingsninės regresijos pagalba labai sumažėjo kintamųjų kiekis. Su SAS paketo procedūra LOGISTIC, nurodžius priklausomą kintamąjį ir nepriklausomus kintamuosius, buvo gauti tokie rezultatai output lange:

### The LOGISTIC Procedure

#### Model Information

Data Set	WORK.KURSINIS
Response Variable	1
Number of Response Levels	2
Number of Observations	176
Model	binary logit
Optimization Technique	Fisher's scoring

Lentelėje *Model Information* pateikiama: *Response Variable* – priklausomo kintamojo vardas; Number of response levels – priklausomo kintamojo reikšmių skaičius. Jis yra 2 t.y., priklausomas kintamasis įgyja dvi reikšmes (0 – namuose žaidžiusi komanda pralaimėjo, 1 – namuose žaidžiusi komanda laimėjo); Number of observation – skaičiavimams naudotų stebėjimų skaičius.

#### Response Profile

Ordered Value	1	Total Frequency
1	0	65
2	1	111

Probability modeled is 1=0.

Lentelėje Response Profile pateikiamas kiekvienos priklausomo kintamojo reikšmės dažnis (stulpelis Total Frequency). Nurodoma priklausomo kintamojo reikšmė, kuri atitinka įvykį, kurio tikimybę norime įvertinti (eilutė Probability Modeled). Taigi įvykis bus pralaimėjimas, o ne įvykis - pergalė žaidžiant namie. Iš dažnių puikiai matosi, kad namie žaidžiančios komandos laimi daug dažniau negu pralaimi. Taip yra visose sporto varžybose. Savų sirgalių palaikomos komandos dažnai įveikia net akivaizdžiai stipresnes ekipas.

#### Model Fit Statistics

R-Square 0.2118 Max-rescaled R-Square 0.2893

Lentelėje Model Fit Statistics pateikiami determinacijos koeficientas (R-Square) ir koreguotasis determinacijos koeficientas (Max-rescaled R-Square). Taigi regresijos modelis paaiškina 21,18% duomenų sklaidos. Modelio suderinamumas su duomenimis yra lygus 0,2893.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.8907	5	<,0001
Score	37.0493	5	<,0001
Wald	29.5991	5	<,0001

Lentelėje Testing Global Null Hypothesis Beta = 0 pateikiamos statistikų, skirtų patikrinti hipotezei  $H : \beta_1 = \dots = \beta_m = 0$ , reikšmės ir atitinkamos p reikšmės. Gaunama, kad hipotezė atmetama. Voldo kriterijus rodo, kad H galime ir neatmesti.

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Wald Error	Chi-Square	Pr > ChiSq
Intercept	1	2.0808	0.9116	5.2099	0.0225
x6	1	-0.0904	0.0437	4.2741	0.0387
xx2	1	-0.1915	0.0618	9.5988	0.0019
xx5	1	-0.3917	0.1518	6.6620	0.0098
y4	1	0.0943	0.0317	8.8275	0.0030
y7	1	0.1060	0.0444	5.7120	0.0168

Lentelėje Analysis of Maximum Likelihood Estimates pateikiami parametų įverčiai (stulpelis Estimate; standartinės paklaidos (stulpelis Standart Error); statistikos, skirtos patikrinti hipotezei: parametras lygus nuliui, reikšmė (stulpelis Wald Chi - Square); atitinkama p reikšmė (stulpelis Pr > ChiSq). Matome, kad likę parametrai yra statistiškai reikšmingi, nes jų p reikšmės mažesnės už reikšmingumo lygmenį.

Gavome, kad iš 36 kovariančių liko tik 5. Visi šie parametrai yra reikšmingi modeliui, t.y. atmetama hipotezė, kad jie lygūs nuliui. Apžvelgsiu šiuos parametrus plačiau. Liko šios kovariantės: x6 (namų komandos atsarginio įžaidėjo naudingumo balų skaičius), xx2 namų

komandos žaidėjų rezultatyvių perdavimų skaičius), xx5 (namų komandos žaidėjų blokuotų metimų skaičius), y4 (svečių komandos starto penketo sunkiojo puolėjo naudingumo balų skaičius), y7 (svečių komandos atsarginio gynėjo naudingumo balų skaičius).

#### Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
x6	0.914	0.839	0.995
xx2	0.826	0.732	0.932
xx5	0.676	0.502	0.910
y4	1.099	1.033	1.169
y7	1.112	1.019	1.213

Lentelėje Odds Ratio pateikiami šansų santykių taškiniai įverčiai (Point Estimate) ir pasiklovimo intervalai (Confidence Limits). Pvz: x6 – namų komandos atsarginiui įžaidėjui surinkus vienu naudingumo balu daugiau, namų komandai sumažėja šansas pralaimėti, nes point estimate  $x1 < 1$ . Idomu, kad labai svarbiu modelis laiko namų komandos blokuotų metimų skaičių.

Klasifikavimo taisyklės įvertis gaunamas taip: objektas priskiriamas I klasei, jei  $\hat{l}(x) = \hat{\beta}^T x - c > 0$  ( $c$  – pasirinktas slenkstis). Lentelėje Classification Table pateikiamos klasifikavimo gerumui nustatyti naudojamos charakteristikos.

#### Classification Table

Prob Level	Correct		Incorrect		Percentages				
	Non-Event	Event	Non-Event	Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.100	65	18	93	0	47.2	100.0	16.2	58.9	0.0
0.200	58	45	66	7	58.5	89.2	40.5	53.2	13.5
0.300	54	65	46	11	67.6	83.1	58.6	46.0	14.5
0.400	41	79	32	24	68.2	63.1	71.2	43.8	23.3
0.500	31	94	17	34	71.0	47.7	84.7	35.4	26.6
0.600	20	101	10	45	68.8	30.8	91.0	33.3	30.8
0.700	12	105	6	53	66.5	18.5	94.6	33.3	33.5
0.800	4	107	4	61	63.1	6.2	96.4	50.0	36.3

Proc Level stulpelis pateikia slenksčių sąrašą. Jeigu prognozuota įvykio tikimybė viršija tam tikrą slenksčio reikšmę  $z \in (0,1)$ , tai stebėjimas prognozuojamas kaip įvykis, priešingu atveju – kaip ne įvykis.

Iš lentelės matyti, kad su šiuo modeliu SAS prognozuoja 71 % tikslumu esant slenksčiui 0.5. Su šiuo slenksčiu gauname tokius rezultatus:

31 atvejais programa teisingai klasifikavo 0 (pralaimėjimą) (t.y. 31-iais atvejais tikroji 1 reikšmė buvo 0 ir tie atvejai buvo klasifikuojami kaip 0); 34 atvejus programa klasifikavo kaip 1 (pergalę), nors buvo 0; taigi jautrumas lygus  $31/(31+34) \approx 47.7\%$ ;

94 atvejais programa teisingai klasifikavo 1 (t.y. 94-iais atvejais tikroji y reikšmė buvo 1 ir tie atvejai buvo klasifikuojami kaip 1); 17 atvejus programa klasifikavo kaip 0, nors buvo 1; taigi specifiškumas lygus  $94/(94 + 17) \approx 84.7\%$ . [3]

Šiuos rezultatus pateiksiu 3.4. skyrelio lentelėmis 1.1 ir 1.2.

	0	1	$\Sigma$
0	31	34	65
1	17	94	111

	0	1	$\Sigma$
0	0.477	0.523	1
1	0.153	0.847	1

Klasifikavimo rezultatai esant slenksčiui 0.5

Stulpelyje False POS paskaičiuojamas procentas neteisingai prognozuotų įvykių nuo tiesingai prognozuotų įvykių + neteisingai prognozuotų įvykių, t.y.  $17 / (31+17) = 35.4\%$ .

Stulpelyje False NEG paskaičiuojamas procentas neteisingai prognozuotų ne įvykių nuo teisingai prognozuotų ne įvykių + neteisingai prognozuotų ne įvykių, t.y.  $34 / (94 + 34) = 26.6\%$ .

Pasirinkus slenksstį 0.4, galime pagerinti jautrumą iki 63.1%, bet tada specifiškumas nukrenta iki 71.2%, o bendras prognozės tikslumas iki 68.2%.

Truputį geresnį rezultatą galime gauti išėmus duomenis paskutinių keturių varžybų, kurios vyko neutralioje aikštėje, dėl ko SAS prastai tų rungtynių baigtis prognozavo. Taigi, išėmus tas varžybas, prognozavimo tikslumas išauga iki 71.5% esant slenksčiui 0.6. Taip pat atsiranda naujas reikšmingas kintamasis - 13 (svečių komandos pergalių skaičius per paskutines trejas varžybas).

## 4.2. Laimėtojo prognozavimas su logistine regresija (VTB lyga)

Šis skyrelis skirtas VTB lygos komandų varžybų laimėtojo prognozavimui atsižvelgiant į prieš tai sužaistų varžybų statistines charakteristikas. Prognozavimo aprašymas identišką Eurolygos varžybų prognozavimui. Vėl geriausi rezultatai gauti su pažingsnine regresija. Toliau aprašysiu tik skirtumus tarp Eurolygos ir VTB lygos prognozių.

Response Profile

Ordered Value	1	Total Frequency
1	0	55
2	1	90

Probability modeled is 1=0.

Iš lentelės Response Profile matome, kad procentaliai VTB lygoje (55/145~36.9%) daugiau yra komandų namie pralaiminčių nei Eurolygoje (65/176~37.9%). Tai nesunku paaiškinti, nes VTB lygoje buvo tokių nestiprių komandų, kaip Talino Kalevas ar Minsko ekipa. Jos dažniausiai pralaimėdavo nepriklausomai nuo to kur žaidavo.

### The LOGISTIC Procedure

#### Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.7529	1.3780	7.4176	0.0065
l2	1	-0.5054	0.1921	6.9263	0.0085
yy6	1	-0.1588	0.0592	7.1890	0.0073

Gavome, kad iš 34 kovariančių liko tik 2:

l2 - namuose žaidžiančios komandos pergalių skaičius per paskutines trejas varžybas

yy6 - svečių komandos paskutinių trejų varžybų gautų pražangų vidurkis

### Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
12	0.603	0.414	0.879
yy6	0.853	0.760	0.958

Odds Ratio Estimates lentelėje matosi kokią didelę įtaka turi namų komandos pergalių skaičius prieš varžybas.

Toliau klasifikavimo lentelė:

### Classification Table

Prob Level	Correct		Incorrect		Percentages				
	Non-Event	Event	Non-Event	Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.100	54	0	90	1	37.2	98.2	0.0	62.5	100.0
0.200	49	16	74	6	44.8	89.1	17.8	60.2	27.3
0.300	43	37	53	12	55.2	78.2	41.1	55.2	24.5
0.400	30	59	31	25	61.4	54.5	65.6	50.8	29.8
0.500	17	77	13	38	64.8	30.9	85.6	43.3	33.0
0.600	12	88	2	43	69.0	21.8	97.8	14.3	32.8
0.700	3	90	0	52	64.1	5.5	100.0	0.0	36.6
0.800	1	90	0	54	62.8	1.8	100.0	0.0	37.5

Iš lentelės matyti, kad su šiuo modeliu SAS prognozuoja 69 % tikslumu esant slenksčiui 0.6. Su šiuo slenksčiu gauname tokius rezultatus:

12 atvejais programa teisingai klasifikavo 0 (pralaimėjimą), 43 atvejus programa klasifikavo kaip 1 (pergalę), nors buvo 0; taigi jautrumas lygus  $12/(12+43) \approx 21.8\%$ ;

88 atvejais programa teisingai klasifikavo 1 (pergalę), 2 atvejus programa klasifikavo kaip 0, nors buvo 1; taigi specifiškumas lygus  $88/(88 + 2) \approx 97.8\%$ . [3]

Šiuos rezultatus pateiksiu 3.4. skyrelio lentelėmis 1.1 ir 1.2.

	0	1	Σ
0	12	43	55
1	2	88	90

	0	1	Σ
0	0.218	0.782	1
1	0.022	0.978	1

Klasifikavimo rezultatai esant slenksčiui 0.5

Nors modelis ir prognozuoja geriausiu atveju 69% , visgi jis netinka dėl prasto jautrumo. Esant slenkščiui 0.4, jautrumas pagerėtų iki 54.5%, o specifiškumas kristų iki 65.6%. Bendras teisingų prognozių skaičius taip pat smuktų iki 61.4%. Tokį netikslų prognozavimą lyginant su Eurolyga galima aiškinti netvarkingu tvarkaraščiu VTB lygoje. Eurolygoje vykdavo varžybos tik kartą per savaitę. Taigi klubai budavo pasiruošę ir daugiau mažiau pailsėję prieš varžybas. Tuo tarpu VTB tvarkaraštis buvo sudarytas netvarkingai. Kartai būdavo sužaidžiama iki 3 rungtynių per savaitę. Favoritai VTB lygoje dažnai eksperimentuodavo komandų sudėtimi, leisdavo pailsėti savo lyderiams ir pan. Iš to buvo ne vienas neprognozuojamas pralaimėjimas. Nenuostabu, kad ir SAS sekėsi praščiau prognozuoti laimėtoją.

### 4.3. Laimėtojo prognozavimas naudojant neuroninius tinklus (Eurolyga)

Duomenys per neuroninį tinklą buvo paleisti naudojantis statistiniu paketu SPSS. Plačiau apie neuroninius tinklus pasiskaitykite 3.2 skyrelyje. Toliau pateiksiu output'o rezultatus:

		N	Percent
Sample	Training	79	44.9%
	Testing	97	55.1%
Valid		176	100.0%
Excluded		0	
Total		176	

Šiame žingsnelyje imtis suskirstoma į dvi imtis: apmokamą (44.9% visos imties) ir testinę (55.1%).

Iėjimo kintamieji yra standartizuoti  $(\frac{x-\mu}{s})$ , aktyvacijos funkcija paslėptam sluoksniui yra

hiperbolinio tangento:  $\gamma(c) = \tanh(c) = \frac{(e^c - e^{-c})}{(e^c + e^{-c})}$ . O išėjimo funkcija vadinama "softmax"

$$\gamma(c_k) = \frac{\exp(c_k)}{\sum_j \exp(c_j)}$$

Dar žemiau pateikta lentelė, kurioje surašyti kintamieji ir jų svarba modeliui. Svarbiausias šiam modelyje pasirodo yra yy6 – svečių komandos pražangų skaičius; Taip pat svarbus y9 – svečių komandos atsarginio sunkiojo puolėjo naudingumo balų skaičius. Toliau eina xx5 - namų komandos blokų vidurkis; x2 - namų komandos starto penketo atakuojančio gynėjo naudingumo vidurkis; y4 - svečių komandos startinio penketo sunkiojo krašto puolėjo naudingumo vidurkis. Idomu, kad visiškai nesvarbus faktorius pasirodė l2 - namų komandos pergalių skaičius per paskutines trejas varžybas.

Independent Variable Importance		
	Importance	Normalized Importance
x1	,029	39,4%
x2	,060	80,2%
x3	,037	49,3%
x4	,019	25,0%
x5	,043	57,3%
x6	,043	57,4%
x7	,028	37,8%
x8	,023	31,4%
x9	,047	63,4%
x10	,008	10,9%
xx1	,021	28,2%
xx2	,023	31,1%
xx3	,003	4,6%
xx4	,019	25,6%
xx5	,062	83,6%
xx6	,017	22,3%
xx7	,003	3,6%
l2	,007	9,7%
y1	,036	48,4%
y2	,024	32,5%
y3	,014	18,5%
y4	,055	74,2%
y5	,027	36,3%



y6	,034	46,3%
y7	,043	57,7%
y8	,004	5,4%
y9	,004	5,1%
y10	,011	15,4%
yy1	,003	3,6%
yy2	,036	48,0%
yy3	,048	64,0%
yy4	,027	36,9%
yy5	,024	32,4%
yy6	,074	100,0%
yy7	,007	9,1%
l3	,039	52,0%

Klasifikavimo lentelė:

Sample	Observed	Predicted		
		0	1	Percent Correct
Training	0	21	7	75.0%
	1	10	41	80.4%
	Overall Percent	39.2%	60.8%	<b>78.5%</b>
Testing	0	21	16	56.8%
	1	9	51	85.0%
	Overall Percent	30.9%	69.1%	<b>74.2%</b>

Dependent Variable: l

Raudonai pažymėjau kiek procentaliai teisingai buvo prognozuotas laimėtojas. Rezultatas lenkia pažingsninės logistinės regresijos rezultatą – 74.2%. Jautrumas ir specifiškumas taip pat geresni: atitinkamai 56.8% ir 85%.

### 4.3. Laimėtojo prognozavimas naudojant neuroninius tinklus (VTB lyga)

Praleidus pro neuroninius tinklus VTB lygos duomenis gavome kitus svarbius kintamuosius modeliui lyginant su pažingsnine logistine regresija:

Y6 - atsarginio svečių komandos įžaidėjo naudingumo balai

xx1 - namų komandos atkovoti kamuoliai

x2 - namų komandos starto penketo gynėjo naudingumo balai

y4 - svečių komandos starto penketo sunkiojo puolėjo naudingumo balai

Klasifikavimo lentelė:

Sample	Observed	Predicted		
		0	1	Percent Correct
Training	0	14	14	50.0%
	1	10	36	78.3%
	Overall Percent	32.4%	67.6%	67.6%
Testing	0	14	13	51.9%
	1	9	35	79.5%
	Overall Percent	32.4%	67.6%	69.0%

Dependent Variable: I

Matome, kad VTB lygos laimėtojų tiksliau prognozuoti nepavyko. Tačiau nors ir mažesnis specifiškumas, nei logistinėje regresijoje, bet pagerėjo jautrumas.

## 5. Lažybų bendrovės

Lažybų bendrovės uždirba siūlydamos koeficientus laimėti vienai ar kitai komandai. Koeficientai sudėliojami 5-3 dienos prieš varžybas. Artėjant rungtynėms jie gali būti koreguojami dėl padidėjusio pinigų srauto už vieną komandą ar susitraumavus vienam iš žaidėjų. Koeficientai yra užrašomi taip:  $k_1 : k_2$ . Tai reiškia, kad lošėjas pastatęs už pirmą komandą 100lt ir jai laimėjus išloštų 100  $k_1$  litų. Pralaimėjimo atveju pralaimėtų visus pastatytus pinigus. Analogiškai pastačius už antrą komandą, jos pergalės atvejų išloštų 100  $k_2$

Lažybų koeficientus skaičiuoja ir koreguoja bukmekeriai. Vienos lažybų kontoros bukmekeriai savo darbą apibūdina taip: "Pagrindinis mano darbas - sudėlioti lažybų koeficientus

ir valdyti riziką, kurią lažybų bendrovė prisiima". Į klausimą kaip nustatomi lažybų koeficientai atsakoma labai paprastai: "Kiekvienų lažybų įvykio tikimybė skaičiuojama remiantis įvairiais faktoriais: varžybos namuose/išvykoje, komandos forma, žaidėjų motyvacija, traumas ir pan. Kai įvertiname visus veiksnius, nustatome lažybų įvykio atsitikimo procentą ir iš jo padaliname 100".

Šitas atsakymas yra tipinis ieškant atsakymų kaip yra sudėliojami koeficientai. Visgi jis neatsako į klausimą kaip randamos tikimybės. Absoliuti dauguma tekstų, internetinių puslapių šia tema yra apie koeficientų sudarinėjimą, bet ne apie tai, kaip dar prieš koeficientus randamos tikimybės. Toliau, neatsižvelgdamas į tai, kaip sudaromos tikimybės, aprašysiu plačiau koeficientų sudarinėjimą lažybų bendrovėse.

## 5.1. Koeficientų nustatymas

Tegu  $p$  ir  $q=1-p$  yra tikrosios (teisingos) tikimybės, kur  $p$  - tikimybė laimėti I komandai,  $q$  - laimėti II komandai. Lažybų kontora dirba tokiu principu: jeigu tikimybės  $p$  ir  $q$  yra teisingos, tai jas pakoreguoja taip, kad lažybų kontoros pelnas būtų toks pat nepriklausomai nuo to už kurią komandą lošėjas stato.

Koeficientų išraiškos per tikimybes:

$$k_1 = \frac{1}{p + \varepsilon_1} \quad (1.12)$$

$$k_2 = \frac{1}{p + \varepsilon_2} \quad (1.13)$$

Pastačius 100lt už pirmąją komandą vidutinis išlošis būtų:

$$1) 100(k_1 - 1)p - 100q \quad (1.14)$$

Pastačius 100lt už antrąją komandą vidutinis išlošis būtų

$$2) 100(k_2 - 1)q - 100p \quad (1.15)$$

sulyginę (1.14) ir (1.15) išraiškas gauname:

$$(k_1 - 1)p - q = (k_2 - 1)q - p$$

$$p k_1 = q k_2 \quad (1.16)$$

Galiojant šiai lygybei vidutinis išlošis yra vienodas statant už bet kurią komandą.

Įsistatom (1.12) ir (1.13) :

$$\frac{p}{p + \varepsilon_1} = \frac{q}{q + \varepsilon_2} \Rightarrow pq + p\varepsilon_2 = pq + \varepsilon_1 q$$

$$pq + p\varepsilon_2 = pq + \varepsilon_1q$$

$$p\varepsilon_2 = q\varepsilon_1 \quad (1.17)$$

Žinome, kad lažybų kontora užsideda kažkokį pelną. Pelną pažymėkime  $c$ . Tada:

$$\varepsilon_1 + \varepsilon_2 = c \quad (1.18)$$

Išsireiškę  $\varepsilon_2$  iš (1.18) ir įsistatę į (1.17) gausime:

$$(1-p) \varepsilon_1 = p(c - \varepsilon_1) \Rightarrow \varepsilon_1 - p\varepsilon_1 = pc - p\varepsilon_1$$

$$\text{Taip } \varepsilon_1 = pc \quad (1.19)$$

$$\text{ir } \varepsilon_2 = qc \quad (1.20)$$

Įsistatom gautas išraiškas į (1.12) ir (1.13) formules gauname:

$$k_1 = \frac{1}{p(1+c)} \quad (1.21)$$

$$\text{ir } k_2 = \frac{1}{q(1+c)} \quad (1.22)$$

Pasikaičiuokime vidutinį išlošį (1.14) įsistačius (1.18) ir (1.19):

$$100(k_1 - 1)p - 100q = 100 \left[ \frac{1}{1+c} - (p+q) \right] = 100 \left( \frac{1}{1+c} - 1 \right) = 100 \left( \frac{-c}{1+c} \right)$$

Matome, kad vidutinis lošėjo išlošis yra neigiamas!

Toliau pateiksiu pavyzdį kaip tai atrodo praktikoje.

Tarkime turime dvi komandas A ir B. Jų 'tikrosios' tikimybės laimėti yra:

Komandos A laimėti tikimybė: 60%

Komandos B laimėti tikimybė: 40%

Sudėję tikimybes gauname 100%. Tokias tikimybes siūlydama lažybų bendrovė nieko neuždirbtų. Siekdama pelno ji truputį pakoreguotų tikimybes:

Komandos A laimėti tikimybė: 60% +  $\varepsilon_1$  %

Komandos B laimėti tikimybė: 40% +  $\varepsilon_2$  %

Lažybų kontoros pelno užsideda įvairiai - nuo 6% iki 10%. Šiuo atveju įsivaizduokime, kad lažybų kontora užsideda 8%. Vadinasi  $\varepsilon_1 + \varepsilon_2 = 8\%$ . Maržą pasiskaičiuoti labai nesunku nuėjus į bet kurios lažybų kontoros internetinę svetainę, nes šias tikimybes galime pasiversti į lažybų kontorų koeficientus ir atvirkščiai.

Lažybų kontora tikisi, kad žmonės statys savo pinigus panašiais santykiais, kaip jų paskaičiuotos tikrosios tikimybės, t.y. iš  $x$  pastatytų pinigų  $0,6x$  bus pastatyta už A komandos pergalę,  $0,4x$  už B komandą. Tada bukmekeriai pelną paskirsto taip, kad nepaisant koks bus

rezultatas, vidutinis išlošis bus lygus visais atvejais ir, aišku, mažesnis už pastatytą pinigų sumą x. Pagal (1.21) ir (1.22) galime pasiskaičiuoti  $\varepsilon_1$  ir  $\varepsilon_2$ :

$$\varepsilon_1 = pc = 0.6 * 0.08 = 0.048$$

$$\varepsilon_2 = qc = 0.4 * 0.08 = 0.032$$

Taip lažybų bendrovė padidina tikimybes laimėti:

Komandos A laimėti tikimybė: 64.8%

Komandos B laimėti tikimybė: 43.2%

Iš to gauname ir koeficientus:

$$64.8\% = 1/0.648 \sim 1.54$$

$$43.2\% = 1/0.432 \sim 2.31$$

Taigi, jeigu pvz iš 100lt 60% pastatoma už A komanda, o likusieji už B, tada:

Jei A komanda laimi, bendrovė sumoka  $60 * 1.54 = 92.4$ lt

Jei B komanda laimi, bendrovė sumoka  $40 * 2.31 = 92.4$ lt

Taip lažybų bendrovė bet kuriuo atveju pasiima 7,6lt.

Jei koeficientai būtų buvę sudėlioti pagal 'tikrasias' tikimybes, tada bendrovei kiekvienu atveju būtų tekę sumokėti po 100lt. Vadinasi koreguodama koeficientus lažybų bendrovė uždirba!

## 5.2. Tikimybės

Kaip jau minėjau pirmo skyriaus pradžioje labai daug informacijos įvairiuose šaltiniuose galima rasti apie koeficientų sudarymą, bet praktiškai nieko apie tikimybių apskaičiavimą. Į bet kokius užklausimus lažybų bendrovėms ir jų bukmekeriams nebuvo visiškai reaguojama. Visgi per trečius asmenis pavyko susirašyti su vienos didžiausių Lietuvos bendrovės buvusiu bukmekeriu.

Jo teigimu, tiek Lietuvos, tiek užsienio lažybų bendrovės savo koeficientų nekuria, o tik kopijuoja juos vienos nuo kitų tik truputi juos pagal save pasikoreguodami, kad nebūtu koeficientai vienodi. Taip darome todėl, kad koeficientai ženkliai nesiskirtų, nes pvz jei vienoje bendrovėje už A komandos pergalę koeficientas būtų 2,5, o už B komandos 1,5, kai tuo tarpu kitoje bendrovėje jau už A komandą koeficientas būtų 1,5, o už B komandą - 2,5. Tokiu atveju

klientas abejoje bendrovėse pastatęs už A komandos pergalę, išloštų be galimybės pralaimėti. Yra kelios bendrovės didžiausios Rusijoje ar azijos valstybėse iš kurios likusios ir kopjuoja.

Taigi bukmekeris jokių formulių pateikti negalėjo, nes jam niekada jokių ir neprireikė.

### 5.3. Lažybų punkto keoficientai

Žemiau esančioje lentelėje pateikiau tam tikros lažybų bendrovės (tikrasis bedrovės pavadinimas autoriui žinomas) lažybinius koeficientus. Lentelę bus galima rasti ir cd esančioje direktorijoje 'keoficientai' faile 3 lentele.xls. Paaiškinimai:

kom1 - namie žaidžiusios komandos pavadinimas

kom2 - svečiuose žaidžiusios komandos pavadinimas

kof1 - lažybų bendrovės koeficientas statant pinigus už kom1 komandą (pastačius x pinigų, pergalės atveju laimėsime kof1\*x pinigų)

kof2 - lažybų bendrovės koeficientas statant pinigus už kom2 komandą (pastačius x pinigų, pergalės atveju laimėsime kof2\*x pinigų)

l - rungtynių laimėtojas (0 - namų komanda pralaimėjo, 1 - namų komanda laimėjo)

$p1=1/kof1$

$p2=1/kof2$

suma - sudėjus p1 ir p2 matyti, kad lažybų bendrovės pelnas yra apie 6%

3 lentelė

Kom1	Kom2	kof1	kof2	l	p1	p2	suma
ASVEL	Ulker	2,36	1,58	0	0,423729	0,632911	1,05664
Barcelona	Cibona	1,02	12,29	1	0,980392	0,081367	1,061759
MdP	Zalgiris	1,04	9,84	1	0,961538	0,101626	1,063164
ASVEL	MdP	7,62	1,08	0	0,131234	0,925926	1,05716
Ulker	Cibona	1,23	4,06	1	0,813008	0,246305	1,059314
Zalgiris	Barcelona	4,15	1,22	0	0,240964	0,819672	1,060636
Cibona	Zalgiris	1,57	2,36	1	0,636943	0,423729	1,060671
Ulker	MdP	4,67	1,18	0	0,214133	0,847458	1,06159
Barcelona	ASVEL	1,01	13,57	1	0,990099	0,073692	1,063791
Zalgiris	Ulker	1,46	2,62	0	0,684932	0,381679	1,066611
ASVEL	Cibona	1,49	2,56	1	0,671141	0,390625	1,061766
MdP	Barcelona	1,66	2,17	0	0,60241	0,460829	1,063239
MdP	Cibona	1,02	13,33	1	0,980392	0,075019	1,055411
Barcelona	Ulker	1,03	11,43	1	0,970874	0,087489	1,058363

ASVEL	Zalgiris	1,85	1,92	1	0,540541	0,520833	1,061374
Ulker	ASVEL	1,15	5,24	0	0,869565	0,19084	1,060405
Cibona	Barcelona	6,53	1,11	0	0,153139	0,900901	1,05404
Zalgiris	MdP	4,27	1,21	0	0,234192	0,826446	1,060638
Cibona	Ulker	1,44	2,73	1	0,694444	0,3663	1,060745
MdP	ASVEL	1,03	12,37	1	0,970874	0,080841	1,051715
Barcelona	Zalgiris	1,02	13,06	1	0,980392	0,07657	1,056962
MdP	Ulker	1,04	10,6	1	0,961538	0,09434	1,055878
Zalgiris	Cibona	1,36	3,11	1	0,735294	0,321543	1,056838
ASVEL	Barcelona	6,29	1,13	0	0,158983	0,884956	1,043938
Cibona	ASVEL	1,47	2,63	1	0,680272	0,380228	1,0605
Ulker	Zalgiris	1,23	4,06	0	0,813008	0,246305	1,059314
Barcelona	MdP	1,15	5,42	1	0,869565	0,184502	1,054067
Unicaja	Olympiacos	2,69	1,45	1	0,371747	0,689655	1,061402
Efes Pilsen	Partizan	1,11	6,35	1	0,900901	0,15748	1,058381
Entente	L.Rytas	1,71	2,1	0	0,584795	0,47619	1,060986
Olympiacos	L.Rytas	1,07	8,06	1	0,934579	0,124069	1,058649
Partizan	Entente	1,25	3,92	1	0,8	0,255102	1,055102
Unicaja	Efes Pilsen	1,39	2,95	1	0,719424	0,338983	1,058408
Entente	Unicaja	3,2	1,34	0	0,3125	0,746269	1,058769
L.Rytas	Partizan	1,33	3,29	1	0,75188	0,303951	1,055831
Olympiacos	Efes Pilsen	1,12	6,09	1	0,892857	0,164204	1,057061
Efes Pilsen	Entente	1,03	11,65	1	0,970874	0,085837	1,056711
Unicaja	L.Rytas	1,18	4,74	1	0,847458	0,21097	1,058428
Partizan	Olympiacos	4,27	1,21	1	0,234192	0,826446	1,060638
Entente	Olympiacos	6,97	1,1	0	0,143472	0,909091	1,052563
Unicaja	Partizan	1,15	5,49	0	0,869565	0,182149	1,051715
Efes Pilsen	L.Rytas	1,2	4,52	1	0,833333	0,221239	1,054572
L.Rytas	Entente	1,1	6,72	1	0,909091	0,14881	1,0579
Olympiacos	Unicaja	1,12	6,28	1	0,892857	0,159236	1,052093
Partizan	Efes Pilsen	1,65	2,19	1	0,606061	0,456621	1,062682
L.Rytas	Olympiacos	3,11	1,36	0	0,321543	0,735294	1,056838
Efes Pilsen	Unicaja	1,4	2,89	0	0,714286	0,346021	1,060306
Entente	Partizan	2,22	1,64	1	0,45045	0,609756	1,060207
Partizan	L.Rytas	1,31	3,4	1	0,763359	0,294118	1,057476
Efes Pilsen	Olympiacos	1,92	1,85	0	0,520833	0,540541	1,061374
Unicaja	Entente	1,04	10,08	0	0,961538	0,099206	1,060745
Entente	Efes Pilsen	2,62	1,48	0	0,381679	0,675676	1,057355
L.Rytas	Unicaja	1,32	3,33	0	0,757576	0,3003	1,057876
Olympiacos	Partizan	1,04	10,39	1	0,961538	0,096246	1,057785
CSKA	Roma	1,12	6	0	0,892857	0,166667	1,059524
Vitorija	Maccabi	1,35	3,11	1	0,740741	0,321543	1,062284
Olimpija	Maroussi	1,26	3,83	0	0,793651	0,261097	1,054747
Maroussi	Roma	1,71	2,1	0	0,584795	0,47619	1,060986
Maccabi	CSKA	1,49	2,56	1	0,671141	0,390625	1,061766
Olimpija	Vitorija	3,26	1,33	0	0,306748	0,75188	1,058628
Maroussi	Vitorija	2,49	1,52	0	0,401606	0,657895	1,059501
CSKA	Olimpija	1,07	7,7	1	0,934579	0,12987	1,06445
Roma	Maccabi	2,06	1,72	0	0,485437	0,581395	1,066832
Olimpija	Roma	2,11	1,7	1	0,473934	0,588235	1,062169
Maccabi	Maroussi	1,04	10,42	1	0,961538	0,095969	1,057508
Vitorija	CSKA	1,39	2,91	0	0,719424	0,343643	1,063067
CSKA	Maroussi	1,08	8,21	1	0,925926	0,121803	1,047729
Vitorija	Roma	1,22	4,2	1	0,819672	0,238095	1,057767

Olimpija	Maccabi	2,81	1,42	0	0,355872	0,704225	1,060097
Maroussi	Olimpija	1,42	2,81	1	0,704225	0,355872	1,060097
Maccabi	Vitorija	1,28	3,6	0	0,78125	0,277778	1,059028
Roma	CSKA	2,21	1,64	0	0,452489	0,609756	1,062245
Vitorija	Olimpija	1,05	10,16	1	0,952381	0,098425	1,050806
Roma	Maroussi	1,42	2,8	0	0,704225	0,357143	1,061368
CSKA	Maccabi	1,39	2,94	1	0,719424	0,340136	1,059561
Vitorija	Maroussi	1,07	8,49	1	0,934579	0,117786	1,052365
Maccabi	Roma	1,12	6,27	1	0,892857	0,15949	1,052347
Olimpija	CSKA	3,47	1,3	0	0,288184	0,769231	1,057415
CSKA	Vitorija	1,48	2,59	1	0,675676	0,3861	1,061776
Roma	Olimpija	1,12	6,14	1	0,892857	0,162866	1,055724
Maroussi	Maccabi	1,78	1,99	1	0,561798	0,502513	1,06431
EWE	Milano	1,8	1,97	0	0,555556	0,507614	1,06317
Real	Prokom	1,03	11,6	1	0,970874	0,086207	1,057081
Panathinaikos	Chimki	1,08	7,45	1	0,925926	0,134228	1,060154
Chimki	EWE	1,11	6,29	1	0,900901	0,158983	1,059883
Prokom	Milano	1,73	2,06	1	0,578035	0,485437	1,063472
Real	Panathinaikos	1,65	2,22	1	0,606061	0,45045	1,056511
Prokom	Panathinaikos	5,71	1,13	0	0,175131	0,884956	1,060087
EWE	Real	5,52	1,14	0	0,181159	0,877193	1,058352
Milano	Chimki			0	#DIV/0!	#DIV/0!	#DIV/0!
Chimki	Prokom	1,11	6,18	1	0,900901	0,161812	1,062713
Panathinaikos	EWE	1,01	12,63	1	0,990099	0,079177	1,069276
Real	Milano	1,05	10,08	1	0,952381	0,099206	1,051587
Real	Chimki	1,06	8,87	1	0,943396	0,11274	1,056136
Panathinaikos	Milano	1,02	12,94	1	0,980392	0,07728	1,057672
EWE	Prokom	1,59	2,33	0	0,628931	0,429185	1,058115
Chimki	Panathinaikos	2,86	1,41	0	0,34965	0,70922	1,05887
Prokom	Real	6,15	1,12	1	0,162602	0,892857	1,055459
Milano	EWE	1,21	4,37	1	0,826446	0,228833	1,055279
EWE	Chimki	3,23	1,34	0	0,309598	0,746269	1,055866
Panathinaikos	Real	1,3	3,45	0	0,769231	0,289855	1,059086
Milano	Prokom	1,36	3,07	1	0,735294	0,325733	1,061027
Panathinaikos	Prokom	1,02	12,83	1	0,980392	0,077942	1,058334
Real	EWE	1,02	14,73	1	0,980392	0,067889	1,048281
Chimki	Milano			1	#DIV/0!	#DIV/0!	#DIV/0!
Prokom	Chimki	1,41	2,86	1	0,70922	0,34965	1,05887
Milano	Real	2,97	1,38	0	0,3367	0,724638	1,061338
EWE	Panathinaikos	5,7	1,13	0	0,175439	0,884956	1,060394
Panathinaikos	Partizan	1,05	10	0	0,952381	0,1	1,052381
Barcelona	Maroussi	1,02	14,05	1	0,980392	0,071174	1,051567
Partizan	Barcelona	4,64	1,18	1	0,215517	0,847458	1,062975
Maroussi	Panathinaikos	5,72	1,14	1	0,174825	0,877193	1,052018
Maroussi	Partizan	1,56	2,39	1	0,641026	0,41841	1,059436
Barcelona	Panathinaikos	1,26	3,72	1	0,793651	0,268817	1,062468
Partizan	Maroussi	1,25	3,89	1	0,8	0,257069	1,057069
Panathinaikos	Barcelona	1,84	1,92	0	0,543478	0,520833	1,064312
Maroussi	Barcelona	4,57	1,19	0	0,218818	0,840336	1,059155
Partizan	Panathinaikos	1,6	2,3	0	0,625	0,434783	1,059783
Panathinaikos	Maroussi	1,62	2,26	1	0,617284	0,442478	1,059762
Barcelona	Partizan	1,04	10,8	1	0,961538	0,092593	1,054131
Real	Efes Pilsen	1,1	7,02	1	0,909091	0,14245	1,051541
Mdp	Maccabi	1,17	4,93	1	0,854701	0,20284	1,057541



Efes Pilsen	MdP	2,27	1,6	1	0,440529	0,625	1,065529
Maccabi	Real	1,76	2,02	1	0,568182	0,49505	1,063231
Maccabi	Efes Pilsen	1,32	3,3	1	0,757576	0,30303	1,060606
Mdp	Real	1,64	2,2	1	0,609756	0,454545	1,064302
Efes Pilsen	Maccabi	1,59	2,3	1	0,628931	0,434783	1,063713
Real	MdP	1,35	3,12	1	0,740741	0,320513	1,061254
Maccabi	MdP	1,56	2,38	1	0,641026	0,420168	1,061194
Efes Pilsen	Real	2,15	1,67	0	0,465116	0,598802	1,063919
Mdp	Efes Pilsen	1,18	4,63	1	0,847458	0,215983	1,06344
Real	Maccabi	1,23	4,02	0	0,813008	0,248756	1,061764
CSKA	Unicaja	1,24	3,88	1	0,806452	0,257732	1,064184
Prokom	Zalgiris	1,39	2,96	1	0,719424	0,337838	1,057262
Zalgiris	CSKA	3,24	1,33	0	0,308642	0,75188	1,060522
Unicaja	Prokom	1,18	4,71	0	0,847458	0,212314	1,059772
CSKA	Prokom	1,12	5,96	1	0,892857	0,167785	1,060642
Unicaja	Zalgiris	1,09	7,31	1	0,917431	0,136799	1,05423
Prokom	CSKA	2,6	1,48	1	0,384615	0,675676	1,060291
Zalgiris	Unicaja	2,72	1,44	1	0,367647	0,694444	1,062092
Zalgiris	Prokom	1,73	2,07	1	0,578035	0,483092	1,061126
Unicaja	CSKA	2,13	1,68	0	0,469484	0,595238	1,064722
CSKA	Zalgiris	1,06	8,43	1	0,943396	0,118624	1,06202
Prokom	Unicaja	1,4	2,89	0	0,714286	0,346021	1,060306
Chimki	Cibona	1,13	5,94	1	0,884956	0,16835	1,053306
Vitorija	Olympiacos	1,64	2,21	0	0,609756	0,452489	1,062245
Olympiacos	Chimki	1,09	7,03	1	0,917431	0,142248	1,059679
Cibona	Vitorija	2,87	1,4	0	0,348432	0,714286	1,062718
Olympiacos	Cibona	1,03	12,22	1	0,970874	0,081833	1,052707
Vitorija	Chimki	1,27	3,65	0	0,787402	0,273973	1,061374
Cibona	Olympiacos	3,55	1,29	0	0,28169	0,775194	1,056884
Chimki	Vitorija	1,55	2,4	0	0,645161	0,416667	1,061828
Cibona	Chimki	2,13	1,68	1	0,469484	0,595238	1,064722
Olympiacos	Vitorija	1,21	4,29	1	0,826446	0,2331	1,059547
Vitorija	Cibona	1,12	6,17	1	0,892857	0,162075	1,054932
Chimki	Olympiacos	1,7	2,1	1	0,588235	0,47619	1,064426
Barcelona	Real	1,16	5,09	1	0,862069	0,196464	1,058533
Barcelona	Real	1,16	5,08	0	0,862069	0,19685	1,058919
Real	Barcelona	2,03	1,74	0	0,492611	0,574713	1,067323
Real	Barcelona	2,15	1,67	0	0,465116	0,598802	1,063919
Maccabi	Partizan	1,19	4,59	0	0,840336	0,217865	1,058201
Maccabi	Partizan	1,19	4,68	1	0,840336	0,213675	1,054011
Partizan	Maccabi	1,64	2,21	1	0,609756	0,452489	1,062245
Partizan	Maccabi	1,56	2,38	1	0,641026	0,420168	1,061194
CSKA	Vitorija	1,3	3,42	1	0,769231	0,292398	1,061628
CSKA	Vitorija	1,22	4,19	1	0,819672	0,238663	1,058336
Vitorija	CSKA	1,74	2,03	1	0,574713	0,492611	1,067323
Vitorija	CSKA	1,79	1,98	0	0,558659	0,505051	1,06371
Olympiacos	Prokom	1,05	9,88	1	0,952381	0,101215	1,053596
Olympiacos	Prokom	1,05	9,9	1	0,952381	0,10101	1,053391
Prokom	Olympiacos	2,72	1,44	1	0,367647	0,694444	1,062092
Prokom	Olympiacos	2,85	1,41	0	0,350877	0,70922	1,060097

Poroje rungtynių nėra visiškai koeficientų. Abejose rungtynėse žaidė Rusų klubas 'Chimki'.  
Nėra informacijos kodėl šių duomenų trūksta.

Tiksliausiai prognozuojančio modelio ieškojau tam, kad tikimybės laimėti vienai ar kitai komandai taip pat nurodytų į favoritą tiksliau. Tikimybes nesunku išsireikšti iš (1.2) formulės:

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m)}$$

Visus dydžius turime, taigi pasirašęs nesudėtingą SAS kodą, paskaičiuoju komandų tikimybes laimėti kiekvienas rungtynes. Kadangi šios tikimybės nėra tikros, o tik mano paskaičiuotos, tai jos bus įverčiai. Tikimybės įvertį kad namų komanda pralaimės žymesiu p1, kad laimės - p2. Pateikiu gautų rezultatų fragmentą (išskiriant Žalgirio namų rungtynes):

p1	p2	Kom1	Kom2
0,744057	0,255943	Zalgiris	Barcelona
0,603348	0,396652	Zalgiris	Ulker
0,558492	0,441508	Zalgiris	MdP
0,533974	0,466026	Zalgiris	Cibona
0,638371	0,361629	Zalgiris	CSKA
0,639856	0,360144	Zalgiris	Unicaja
0,822265	0,177735	Zalgiris	Prokom

Iš p1 tikimybių matome, kad Žalgiris, pasak mano sukurto modelio, visose namų rungtynėse buvo favoritas, bet laimėjo iš jų tik trejas varžybas.

#### 5.4. Lažybų kontoros tikimybių palyginimas su tik. įverčiais bei žaidėjo pelno paskaičiavimas

Žinome, kad galiojant (1.16) lygybei vidutinis išlošis yra vienodas statant už bet kurią komandą. Vadinas, turint tikrasias tikimybes statyti reikia už tą komandą, kurios vidutinis išlošis didesnis. Iš (1.16) turime lygybę  $\frac{p}{p + \varepsilon_1} = \frac{q}{q + \varepsilon_2}$  (1.23)

Vietoje skaitiklyje esančių tikimybių imkime paskaičiuotus tik. įverčius. Iš trečios lentelės pagal lygybes (1.19) ir (1.20) pasiskaičiuoju tikrasias lažybų punkto paskaičiuotas tikimybes. Į (1.23)

lygybės vardiklius įsistatę tikrasias tikimybes gaunu  $\frac{\hat{p}}{p} = \frac{\hat{q}}{q} \Rightarrow \frac{\hat{p}}{\hat{q}} = \frac{p}{q}$

Remdamasis šia lygybę bandysiu statyti pinigų. Pingiai bus statomi už namų komandą, jei

$$\frac{\hat{p}}{\hat{q}} > \frac{p}{q} \text{ ir už svečių komandą, jei } \frac{\hat{p}}{\hat{q}} < \frac{p}{q}$$

Bus statoma po 2lt kiekvienose varžybose.

Žemiau pateiktoje lentelėje bus parodyti statymų rezultatai bei palygintos lažybų kontoros tikimybės su paskaičiuotais tikimybių įverčiais. Šią lentelę taip pat bus galima rasti cd direktorijoje 'koeficientai' pavadinimu 4 lentele.xls Paaiškinimai:

kom1 - namie žaidžiusios komandos pavadinimas;

kom2 - svečiuose žaidžiusios komandos pavadinimas;

p - tikroji tikimybė laimėti kom1. Tikimybė atstatyta pagal lažybų kontoros koeficientus

q - tikroji tikimybė laimėti kom2. Tikimybė atstatyta pagal lažybų kontoros koeficientus

p\_iv. - tik. įvertis laimėti kom1. Tikimybės paskaičiuotos prognozuojant su logistine regresija.

q\_iv. - tik. įvertis laimėti kom2. Tikimybės paskaičiuotos prognozuojant su logistine regresija.

p-p\_iv - lažybų bendrovės ir tik. įverčių skirtumas;

$$p/q = \frac{p}{q};$$

$$p_{iv}/q_{iv} = \frac{\hat{p}}{\hat{q}};$$

kof1 - lažybų bendrovės koeficientas statant pinigų už kom1 komandą (pastačius x pinigų, pergalės atveju laimėsime kof1\*x pinigų);

kof2 - lažybų bendrovės koeficientas statant pinigų už kom2 komandą (pastačius x pinigų, pergalės atveju laimėsime kof2\*x pinigų);

l - rungtynių laimėtojas (0 - namų komanda pralaimėjo, 1 - namų komanda laimėjo);

pelnas - uždirbti pinigai po rungtynių

Kom1	Kom2	p	q	p_iv.	q_iv	p-p_iv	p/q	p_iv/q_iv	kof1	kof2	l	pelnas
ASVEL	Ulker	0,40	0,60	0,38	0,62	0,02	0,67	0,62	2,36	1,58	0	1,16
Barca	Cibona	0,92	0,08	0,99	0,01	-0,07	12,05	91,28	1,02	12,29	1	0,04
MdP	Zalgiris	0,90	0,10	0,81	0,19	0,09	9,46	4,39	1,04	9,84	1	-2
ASVEL	MdP	0,12	0,88	0,53	0,47	-0,40	0,14	1,11	7,62	1,08	0	-2
Ulker	Cibona	0,77	0,23	0,80	0,20	-0,04	3,30	4,10	1,23	4,06	1	0,46
Zalgiris	Barca	0,23	0,77	0,74	0,26	-0,52	0,29	2,91	4,15	1,22	0	-2
Cibona	Zalgiris	0,60	0,40	0,15	0,85	0,45	1,50	0,18	1,57	2,36	1	-2
Ulker	MdP	0,20	0,80	0,67	0,33	-0,47	0,25	2,07	4,67	1,18	0	-2
Barca	ASVEL	0,93	0,07	0,93	0,07	0,00	13,44	13,83	1,01	13,57	1	0,02

Zalgiris	Ulker	0,64	0,36	0,60	0,40	0,04	1,79	1,52	1,46	2,62	0	3,24
ASVEL	Cibona	0,63	0,37	0,74	0,26	-0,11	1,72	2,81	1,49	2,56	1	0,98
MdP	Barca	0,57	0,43	0,61	0,39	-0,04	1,31	1,53	1,66	2,17	0	-2
MdP	Cibona	0,93	0,07	0,52	0,48	0,41	13,07	1,09	1,02	13,33	1	-2
Barca	Ulker	0,92	0,08	0,57	0,43	0,35	11,10	1,30	1,03	11,43	1	-2
ASVEL	Zalgiris	0,51	0,49	0,56	0,44	-0,05	1,04	1,29	1,85	1,92	1	1,7
Ulker	ASVEL	0,82	0,18	0,61	0,39	0,21	4,56	1,53	1,15	5,24	0	8,48
Cibona	Barca	0,15	0,85	0,27	0,73	-0,13	0,17	0,37	6,53	1,11	0	-2
Zalgiris	MdP	0,22	0,78	0,56	0,44	-0,34	0,28	1,26	4,27	1,21	0	-2
Cibona	Ulker	0,65	0,35	0,63	0,37	0,03	1,90	1,69	1,44	2,73	1	-2
MdP	ASVEL	0,92	0,08	0,74	0,26	0,19	12,01	2,81	1,03	12,37	1	-2
Barca	Zalgiris	0,93	0,07	0,85	0,15	0,08	12,80	5,54	1,02	13,06	1	-2
MdP	Ulker	0,91	0,09	0,84	0,16	0,07	10,19	5,12	1,04	10,6	1	-2
Zalgiris	Cibona	0,70	0,30	0,53	0,47	0,16	2,29	1,15	1,36	3,11	1	-2
ASVEL	Barca	0,15	0,85	0,24	0,76	-0,09	0,18	0,32	6,29	1,13	0	-2
Cibona	ASVEL	0,64	0,36	0,57	0,43	0,07	1,79	1,34	1,47	2,63	1	-2
Ulker	Zalgiris	0,77	0,23	0,21	0,79	0,56	3,30	0,26	1,23	4,06	0	6,12
Barca	MdP	0,82	0,18	0,97	0,03	-0,14	4,71	31,55	1,15	5,42	1	0,3
Unicaja	OLY	0,35	0,65	0,10	0,90	0,25	0,54	0,11	2,69	1,45	1	-2
Efes	Partizan	0,85	0,15	0,85	0,15	0,00	5,72	5,87	1,11	6,35	1	0,22
Entente	L.Rytas	0,55	0,45	0,28	0,72	0,27	1,23	0,38	1,71	2,1	0	2,2
OLY	L.Rytas	0,88	0,12	0,91	0,09	-0,03	7,53	10,76	1,07	8,06	1	0,14
Partizan	Entente	0,76	0,24	0,65	0,35	0,11	3,14	1,86	1,25	3,92	1	-2
Unicaja	Efes	0,68	0,32	0,55	0,45	0,13	2,12	1,20	1,39	2,95	1	-2
Entente	Unicaja	0,30	0,70	0,52	0,48	-0,23	0,42	1,09	3,2	1,34	0	-2
L.Rytas	Partizan	0,71	0,29	0,75	0,25	-0,03	2,47	2,95	1,33	3,29	1	0,66
OLY	Efes	0,84	0,16	0,89	0,11	-0,05	5,44	8,47	1,12	6,09	1	0,24
Efes	Entente	0,92	0,08	0,90	0,10	0,02	11,31	8,63	1,03	11,65	1	-2
Unicaja	L.Rytas	0,80	0,20	0,90	0,10	-0,09	4,02	8,56	1,18	4,74	1	0,36
Partizan	OLY	0,22	0,78	0,54	0,46	-0,32	0,28	1,16	4,27	1,21	1	6,54
Entente	OLY	0,14	0,86	0,30	0,70	-0,16	0,16	0,43	6,97	1,1	0	-2
Unicaja	Partizan	0,83	0,17	0,85	0,15	-0,03	4,77	5,75	1,15	5,49	0	-2
Efes	L.Rytas	0,79	0,21	0,91	0,09	-0,12	3,77	10,05	1,2	4,52	1	0,4
L.Rytas	Entente	0,86	0,14	0,37	0,63	0,49	6,11	0,58	1,1	6,72	1	-2
OLY	Unicaja	0,85	0,15	0,89	0,11	-0,04	5,61	7,86	1,12	6,28	1	0,24
Partizan	Efes	0,57	0,43	0,77	0,23	-0,20	1,33	3,26	1,65	2,19	1	1,3
L.Rytas	OLY	0,30	0,70	0,16	0,84	0,14	0,44	0,20	3,11	1,36	0	0,72
Efes	Unicaja	0,67	0,33	0,52	0,48	0,16	2,06	1,08	1,4	2,89	0	3,78
Entente	Partizan	0,42	0,58	0,58	0,42	-0,15	0,74	1,36	2,22	1,64	1	2,44
Partizan	L.Rytas	0,72	0,28	0,89	0,11	-0,17	2,60	8,43	1,31	3,4	1	0,62
Efes	OLY	0,49	0,51	0,36	0,64	0,13	0,96	0,57	1,92	1,85	0	1,7
Unicaja	Entente	0,91	0,09	0,13	0,87	0,78	9,69	0,15	1,04	10,08	0	18,16
Entente	Efes	0,36	0,64	0,12	0,88	0,24	0,56	0,13	2,62	1,48	0	0,96
L.Rytas	Unicaja	0,72	0,28	0,81	0,19	-0,09	2,52	4,20	1,32	3,33	0	-2
OLY	Partizan	0,91	0,09	0,83	0,17	0,08	9,99	4,83	1,04	10,39	1	-2
CSKA	Roma	0,84	0,16	0,45	0,55	0,39	5,36	0,82	1,12	6	0	10
Vitorija	Maccabi	0,70	0,30	0,82	0,18	-0,12	2,30	4,56	1,35	3,11	1	0,7
Olimpija	Maroussi	0,75	0,25	0,82	0,18	-0,07	3,04	4,52	1,26	3,83	0	-2
Maroussi	Roma	0,55	0,45	0,69	0,31	-0,13	1,23	2,18	1,71	2,1	0	-2

Maccabi	CSKA	0,63	0,37	0,95	0,05	-0,32	1,72	18,71	1,49	2,56	1	0,98
Olimpija	Vitorija	0,29	0,71	0,64	0,36	-0,35	0,41	1,77	3,26	1,33	0	-2
Maroussi	Vitorija	0,38	0,62	0,68	0,32	-0,30	0,61	2,08	2,49	1,52	0	-2
CSKA	Olimpija	0,88	0,12	0,83	0,17	0,05	7,20	4,83	1,07	7,7	1	-2
Roma	Maccabi	0,46	0,54	0,37	0,63	0,08	0,83	0,60	2,06	1,72	0	1,44
Olimpija	Roma	0,45	0,55	0,27	0,73	0,18	0,81	0,36	2,11	1,7	1	-2
Maccabi	Maroussi	0,91	0,09	0,95	0,05	-0,04	10,02	19,83	1,04	10,42	1	0,08
Vitorija	CSKA	0,68	0,32	0,63	0,37	0,05	2,09	1,69	1,39	2,91	0	3,82
CSKA	Maroussi	0,88	0,12	0,83	0,17	0,05	7,60	4,93	1,08	8,21	1	-2
Vitorija	Roma	0,77	0,23	0,61	0,39	0,17	3,44	1,54	1,22	4,2	1	-2
Olimpija	Maccabi	0,34	0,66	0,46	0,54	-0,13	0,51	0,86	2,81	1,42	0	-2
Maroussi	Olimpija	0,66	0,34	0,80	0,20	-0,14	1,98	3,99	1,42	2,81	1	0,84
Maccabi	Vitorija	0,74	0,26	0,86	0,14	-0,12	2,81	6,05	1,28	3,6	0	-2
Roma	CSKA	0,43	0,57	0,23	0,77	0,20	0,74	0,30	2,21	1,64	0	1,28
Vitorija	Olimpija	0,91	0,09	0,88	0,12	0,03	9,68	7,21	1,05	10,16	1	-2
Roma	Maroussi	0,66	0,34	0,29	0,71	0,37	1,97	0,41	1,42	2,8	0	3,6
CSKA	Maccabi	0,68	0,32	0,81	0,19	-0,13	2,12	4,34	1,39	2,94	1	0,78
Vitorija	Maroussi	0,89	0,11	0,70	0,30	0,19	7,93	2,35	1,07	8,49	1	-2
Maccabi	Roma	0,85	0,15	0,88	0,12	-0,03	5,60	7,02	1,12	6,27	1	0,24
Olimpija	CSKA	0,27	0,73	0,39	0,61	-0,12	0,37	0,64	3,47	1,3	0	-2
CSKA	Vitorija	0,64	0,36	0,52	0,48	0,11	1,75	1,10	1,48	2,59	1	-2
Roma	Olimpija	0,85	0,15	0,32	0,68	0,53	5,48	0,47	1,12	6,14	1	-2
Maroussi	Maccabi	0,53	0,47	0,79	0,21	-0,27	1,12	3,85	1,78	1,99	1	1,56
EWE	Milano	0,52	0,48	0,85	0,15	-0,32	1,09	5,46	1,8	1,97	0	-2
Real	Prokom	0,92	0,08	0,79	0,21	0,13	11,26	3,78	1,03	11,6	1	-2
PAO	Chimki	0,87	0,13	0,91	0,09	-0,04	6,90	10,31	1,08	7,45	1	0,16
Chimki	EWE	0,85	0,15	0,45	0,55	0,40	5,67	0,82	1,11	6,29	1	-2
Prokom	Milano	0,54	0,46	0,66	0,34	-0,12	1,19	1,97	1,73	2,06	1	1,46
Real	PAO	0,57	0,43	0,94	0,06	-0,36	1,35	14,72	1,65	2,22	1	1,3
Prokom	PAO	0,17	0,83	0,74	0,26	-0,58	0,20	2,90	5,71	1,13	0	-2
EWE	Real	0,17	0,83	0,33	0,67	-0,16	0,21	0,49	5,52	1,14	0	-2
Milano	Chimki	#####	#####	0,23	0,77	#####	#####	0,30			0	0
Chimki	Prokom	0,85	0,15	0,51	0,49	0,34	5,57	1,02	1,11	6,18	1	-2
PAO	EWE	0,93	0,07	0,95	0,05	-0,02	12,50	18,72	1,01	12,63	1	0,02
Real	Milano	0,91	0,09	0,96	0,04	-0,06	9,60	25,31	1,05	10,08	1	0,1
Real	Chimki	0,89	0,11	0,95	0,05	-0,05	8,37	17,65	1,06	8,87	1	0,12
PAO	Milano	0,93	0,07	0,81	0,19	0,12	12,69	4,31	1,02	12,94	1	-2
EWE	Prokom	0,59	0,41	0,58	0,42	0,01	1,47	1,40	1,59	2,33	0	2,66
Chimki	PAO	0,33	0,67	0,40	0,60	-0,06	0,49	0,65	2,86	1,41	0	-2
Prokom	Real	0,15	0,85	0,56	0,44	-0,40	0,18	1,26	6,15	1,12	1	10,3
Milano	EWE	0,78	0,22	0,50	0,50	0,28	3,61	1,00	1,21	4,37	1	-2
EWE	Chimki	0,29	0,71	0,51	0,49	-0,21	0,41	1,03	3,23	1,34	0	-2
PAO	Real	0,73	0,27	0,44	0,56	0,29	2,65	0,78	1,3	3,45	0	4,9
Milano	Prokom	0,69	0,31	0,70	0,30	-0,01	2,26	2,35	1,36	3,07	1	0,72
PAO	Prokom	0,93	0,07	0,87	0,13	0,06	12,58	6,55	1,02	12,83	1	-2
Real	EWE	0,94	0,06	0,71	0,29	0,23	14,44	2,41	1,02	14,73	1	-2
Chimki	Milano	#####	#####	0,82	0,18	#####	#####	4,64			1	0
Prokom	Chimki	0,67	0,33	0,56	0,44	0,11	2,03	1,28	1,41	2,86	1	-2
Milano	Real	0,32	0,68	0,33	0,67	-0,01	0,46	0,48	2,97	1,38	0	-2

EWE	PAO	0,17	0,83	0,10	0,90	0,07	0,20	0,11	5,7	1,13	0	0,26
PAO	Partizan	0,90	0,10	0,67	0,33	0,23	9,52	2,05	1,05	10	0	18
Barca	Maroussi	0,93	0,07	0,94	0,06	-0,01	13,77	16,88	1,02	14,05	1	0,04
Partizan	Barca	0,20	0,80	0,94	0,06	-0,74	0,25	16,36	4,64	1,18	1	7,28
Maroussi	PAO	0,17	0,83	0,65	0,35	-0,48	0,20	1,84	5,72	1,14	1	9,44
Maroussi	Partizan	0,61	0,39	0,73	0,27	-0,12	1,53	2,65	1,56	2,39	1	1,12
Barca	PAO	0,75	0,25	0,88	0,12	-0,14	2,95	7,55	1,26	3,72	1	0,52
Partizan	Maroussi	0,76	0,24	0,83	0,17	-0,08	3,11	5,06	1,25	3,89	1	0,5
PAO	Barca	0,51	0,49	0,34	0,66	0,17	1,04	0,51	1,84	1,92	0	1,84
Maroussi	Barca	0,21	0,79	0,58	0,42	-0,38	0,26	1,39	4,57	1,19	0	-2
Partizan	PAO	0,59	0,41	0,52	0,48	0,07	1,44	1,09	1,6	2,3	0	2,6
PAO	Maroussi	0,58	0,42	0,70	0,30	-0,12	1,40	2,35	1,62	2,26	1	1,24
Barca	Partizan	0,91	0,09	0,73	0,27	0,18	10,38	2,77	1,04	10,8	1	-2
Real	Efes	0,86	0,14	0,25	0,75	0,61	6,38	0,34	1,1	7,02	1	-2
Mdp	Maccabi	0,81	0,19	0,67	0,33	0,14	4,21	2,05	1,17	4,93	1	-2
Efes	MdP	0,41	0,59	0,62	0,38	-0,20	0,70	1,61	2,27	1,6	1	2,54
Maccabi	Real	0,53	0,47	0,79	0,21	-0,25	1,15	3,71	1,76	2,02	1	1,52
Maccabi	Efes	0,71	0,29	0,86	0,14	-0,14	2,50	5,91	1,32	3,3	1	0,64
Mdp	Real	0,57	0,43	0,45	0,55	0,12	1,34	0,83	1,64	2,2	1	-2
Efes	Maccabi	0,59	0,41	0,71	0,29	-0,12	1,45	2,41	1,59	2,3	1	1,18
Real	MdP	0,70	0,30	0,38	0,62	0,32	2,31	0,60	1,35	3,12	1	-2
Maccabi	MdP	0,60	0,40	0,95	0,05	-0,35	1,53	20,40	1,56	2,38	1	1,12
Efes	Real	0,44	0,56	0,80	0,20	-0,37	0,78	4,07	2,15	1,67	0	-2
Mdp	Efes	0,80	0,20	0,79	0,21	0,01	3,92	3,74	1,18	4,63	1	-2
Real	Maccabi	0,77	0,23	0,63	0,37	0,14	3,27	1,71	1,23	4,02	0	6,04
CSKA	Unicaja	0,76	0,24	0,71	0,29	0,05	3,13	2,44	1,24	3,88	1	-2
Prokom	Zalgiris	0,68	0,32	0,57	0,43	0,11	2,13	1,31	1,39	2,96	1	-2
Zalgiris	CSKA	0,29	0,71	0,64	0,36	-0,35	0,41	1,77	3,24	1,33	0	-2
Unicaja	Prokom	0,80	0,20	0,67	0,33	0,13	3,99	2,00	1,18	4,71	0	7,42
CSKA	Prokom	0,84	0,16	0,80	0,20	0,04	5,32	3,97	1,12	5,96	1	-2
Unicaja	Zalgiris	0,87	0,13	0,43	0,57	0,44	6,71	0,74	1,09	7,31	1	-2
Prokom	CSKA	0,36	0,64	0,72	0,28	-0,35	0,57	2,51	2,6	1,48	1	3,2
Zalgiris	Unicaja	0,35	0,65	0,64	0,36	-0,29	0,53	1,78	2,72	1,44	1	3,44
Zalgiris	Prokom	0,54	0,46	0,82	0,18	-0,28	1,20	4,63	1,73	2,07	1	1,46
Unicaja	CSKA	0,44	0,56	0,56	0,44	-0,12	0,79	1,29	2,13	1,68	0	-2
CSKA	Zalgiris	0,89	0,11	0,92	0,08	-0,03	7,95	10,96	1,06	8,43	1	0,12
Prokom	Unicaja	0,67	0,33	0,54	0,46	0,14	2,06	1,16	1,4	2,89	0	3,78
Chimki	Cibona	0,84	0,16	0,89	0,11	-0,05	5,26	8,03	1,13	5,94	1	0,26
Vitorija	OLY	0,57	0,43	0,42	0,58	0,15	1,35	0,74	1,64	2,21	0	2,42
OLY	Chimki	0,87	0,13	0,84	0,16	0,03	6,45	5,09	1,09	7,03	1	-2
Cibona	Vitorija	0,33	0,67	0,21	0,79	0,12	0,49	0,26	2,87	1,4	0	0,8
OLY	Cibona	0,92	0,08	0,88	0,12	0,04	11,86	7,28	1,03	12,22	1	-2
Vitorija	Chimki	0,74	0,26	0,62	0,38	0,12	2,87	1,65	1,27	3,65	0	5,3
Cibona	OLY	0,27	0,73	0,53	0,47	-0,26	0,36	1,12	3,55	1,29	0	-2
Chimki	Vitorija	0,61	0,39	0,29	0,71	0,32	1,55	0,41	1,55	2,4	0	2,8
Cibona	Chimki	0,44	0,56	0,49	0,51	-0,05	0,79	0,95	2,13	1,68	1	2,26
OLY	Vitorija	0,78	0,22	0,64	0,36	0,14	3,55	1,76	1,21	4,29	1	-2
Vitorija	Cibona	0,85	0,15	0,84	0,16	0,01	5,51	5,19	1,12	6,17	1	-2
Chimki	OLY	0,55	0,45	0,44	0,56	0,12	1,24	0,78	1,7	2,1	1	-2

Barca	Real	0,81	0,19	0,53	0,47	0,28	4,39	1,13	1,16	5,09	1	-2
Barca	Real	0,81	0,19	0,69	0,31	0,12	4,38	2,25	1,16	5,08	0	8,16
Real	Barca	0,46	0,54	0,45	0,55	0,01	0,86	0,82	2,03	1,74	0	1,48
Real	Barca	0,44	0,56	0,44	0,56	-0,01	0,78	0,80	2,15	1,67	0	-2
Maccabi	Partizan	0,79	0,21	0,81	0,19	-0,02	3,86	4,26	1,19	4,59	0	-2
Maccabi	Partizan	0,80	0,20	0,94	0,06	-0,14	3,93	14,88	1,19	4,68	1	0,38
Partizan	Maccabi	0,57	0,43	0,74	0,26	-0,17	1,35	2,89	1,64	2,21	1	1,28
Partizan	Maccabi	0,60	0,40	0,71	0,29	-0,10	1,53	2,40	1,56	2,38	1	1,12
CSKA	Vitorija	0,72	0,28	0,66	0,34	0,06	2,63	1,98	1,3	3,42	1	-2
CSKA	Vitorija	0,77	0,23	0,69	0,31	0,08	3,43	2,26	1,22	4,19	1	-2
Vitorija	CSKA	0,54	0,46	0,71	0,29	-0,17	1,17	2,46	1,74	2,03	1	1,48
Vitorija	CSKA	0,53	0,47	0,44	0,56	0,08	1,11	0,80	1,79	1,98	0	1,96
OLY	Prokom	0,90	0,10	0,95	0,05	-0,05	9,41	18,66	1,05	9,88	1	0,1
OLY	Prokom	0,90	0,10	0,79	0,21	0,12	9,43	3,72	1,05	9,9	1	-2
Prokom	OLY	0,35	0,65	0,12	0,88	0,22	0,53	0,14	2,72	1,44	1	-2
Prokom	OLY	0,33	0,67	0,29	0,71	0,04	0,49	0,40	2,85	1,41	0	0,82
											su ma	50,16

4 lentelė

Gale sezono viso pelno gavosi 50.16lt !

Matome, kad didesni laimėjimai susigeneravo iš kelių ypač skirtingų mano tikimybių ir lažybų bendrovės koeficientų. Didžiausias laimėjimas buvo rungtynėse tarp Unicaja ir Entente. Lažybų bendrovė ryškiu favoritu įvardijo Unicaja, kai mano tikimybių įverčiai, tuo tarpu, kaip tik rodė, kad favoritai Entente. Pastatęs pinigų ne už lažybų kontoros favoritą 18.16lt.

Taip pat pabandžius su neuronų tinklų rastomis tikimybėmis laimėjimo suma buvo 27.58lt. Matome, kad beveik dvigubai mažesnis laimėjimas šis. Taip yra dėl to, kad nors neuroniniai tinklai ir truputį tiksliau prognozavo laimėtoją, tačiau praščiau prognozavo 'netikėtas pergales', kurios ir sunėšė didžiausią pelną. Lentelę su neuroninių tinklų rastomis tikimybėmis ir išlošius rasite cd direktorijoje 'koeficientai' faile neuro.xls.

## Išvados

Buvo prognozuojamos dviejų krepšinio lygų rungtynių nugalėtojai. Eurolygos rungtynių nugalėtojas buvo tiksliau prognozuojamas nei VTB krepšinio lygos. Atitinkamai 74.2% ir 69%. Abu rezultatai pasiekti naudojant neuroninius tinklus. Logistinės regresijos rezultatai buvo kiek kuklesni - 71.5% ir 69%. Prastesnes VTB lygos prognozės sieju su 'netvarka' pačioje lygoje. Tvarakaraštis sudėliotas netolygiai. Kartais viena komanda sužaisdavo iki 3 rungtynių per savaitę, kai kitos, tuo tarpu, tik vienas rungtynes arba nė vienu. Komandos, kurios be VTB lygos, dar žaisdavo ir kitose Europos lygose, mažiau kreipdavo dėmesio į VTB, dėl ko taupydavo žaidėjus ir to pasekoje ne vienas rezultatas kiek netikėtas būdavo. Kaip pvz galėčiau duoti kad ir netikėtus Žalgirio kluptelėjimus prieš ne pačias stipriausias Rusijos komandas.

Vėlgi, prognozuojant yra kur tobulėti. Yra daug faktorių, kuriuos galima įtraukti į modelius. Nėra atsižvelgta į konkrečius žaidėjus, o tik į jų pozicijas. Neatsižvelgta ir į komandų rezultatus nacionalinėse pirmenybėse, žaidėjų traumas, komandos vidaus santykius tarp žaidėjų, trenerių pajėgumą ar jų kaitą vidurį sezono ir t.t. Įtraukius šią informaciją, manau, ne tik kad pagerintume prognozes, bet galbūt ir tiksliau prognozuotume, negu prognozuoja lažybų kontoros.

Idomių rezultatų pavyko gauti sudarant savo tikimybes laimėti vienai ar kitai komandai. Statant pinigus pagal savo paskaičiuotas tikimybes galima užsidirbti pinigų. Tiesa, pirmiausia reikėtų atlikti panašius bandymus su kitomis lygomis, kad įsitikinti, jog šis laimėjimas ne atsitiktinis.

Šis darbas yra įvadas bet kuriam norinčiam prognozuoti rungtynių baigtis ar lažintis lažybų kontoroje. Pasinaudojus atliktais tyrimais ir išvadomis, darbą galima plėtoti ieškant dar tikslesnių prognozių ar didesnių išlošimų. Darbą galima pritaikyti, bet kuriems sportiniams žaidimams.



## Santrauka

Darbe bandoma prognozuoti stipriausios Europos krepšinio lygos Eurolygos ir VTB krepšinio lygos rungtynių baigtis. Prognozuota buvo naudojant pažingsninės logistinės regresijos metodus ir neuroninius tinklus. Taip pat, remiantis prognozavimo rezultatais, buvo rastos tikimybės laimėti vienai ar kitai komandai, kurios vėliau buvo panaudotos bandant lažintis lažybų kontoroje.

Darbo tikslas - išsiaiškinti ar skiriasi prognozės prognozuojant su logistine regresija ir naudojant neuroninius tinklus. Taip pat, bandymas išlošti pinigų lyginant paskaičiuotus tikimybių įverčius su lažybų kontoros siūlomomis tikimybėmis. Gavome, kad neuroniniai tinklai prognozuoja kiek geriau už logistinę regresiją. Neuroniniai tinklai prognozavo 74.2% tikslumu, logistinė regresija - 71.5%. Bandant išlošti, didesnis išlošis gavosi pasinaudojus logistinės regresijos prognozuotų rungtynių baigčių tikimybių įverčiais.

Ateityje plečiant tyrimus ir suradus galimybę greičiau duomenis susirinkti, bus galima bandyti pasiekti geresnius rezultatus.

## Summary

In this work there is predicting basketball game winner by using data from the strongest European basketball league Euroleague and VTB basketball league protocols. There were used stepwise logistic regression methods and neural networks for predicting results. Also, based on forecast results there were found the probabilities to win a game for the team, which were used for betting in the betting firms.

The aim of the work is to find out which mathematical methods are better for forecasting: logistic regression or neural networks. Also to find out is it possible to make money from betting companies, if you know your own probabilities of winning.

The result is that better forecasting is from neural networks, but more money can be made using logistic regressions.

In the future, working with more data it is possible to get better results.

## Literatūra ir šaltiniai

1. Eurolygos internetinė svetainė, <http://www.euroleague.net>
2. V. Kazakevičius. Kokybinių duomenų analizė. 52p.  
<http://uosis.mif.vu.lt/~vytas/kokyb/kokyb-2005>
3. J. Kruopis. Matematinė statistika. Vilnius, Mokslas, 1993. 331 p.
4. R. Levulienė. Statistika taikymai naudojant SAS. Vilnius, VU, 2009. 322 p.
5. Ugniaus Paukščio baigiamasis bakalaурinis darbas. Krepšinio varžybų laimėtojo prognozavimas. 2010m
6. A. Verikas, A. Gelžinis. Neuroniniai tinklai ir neuroniniai skaičiavimai. Kaunas, KTU, 2008. 8p.
7. VTB lygos internetinė svetainė <http://www.vtb-league.com/en/home/home.htm>.