

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro darbas

STRUKTŪRINIS SPORTO VARŽYBŲ
MODELIAVIMAS IR JŲ REZULTATŲ
PROGNOZAVIMAS

STRUCTURAL MODELLING AND FORECASTING
OF SPORTS EVENTS

Andrius Padriežas

VILNIUS 2018

Darbo vadovas prof. Marijus Radavičius

Darbo recenzentas Raimondas Malukas

Darbas apgintas 2018 m. sausio 15d.

Darbas įvertintas _____

Registravimo Nr. 111000-9.1-5/_____

2018-01-05 _____

Turinys

1	Santrauka	3
2	Įvadas	4
3	TEORINĖ DALIS	6
3.1	Sporto rungtynių prognozavimas ir lažybos moksliniuose darbuose	6
3.2	NBA statistika ir istorija	8
3.3	Struktūrinių lygčių modeliai	10
3.3.1	Latentiniai kintamieji	11
3.3.2	Parametrų vertinimas	12
3.3.3	Modelio identifikavimas	13
3.3.4	Statistika ir gerumo matai. SEM modelio sudarymo procesas	13
3.3.5	Grafinis modelio atvaizdavimas	15
3.4	Neuroninių tinklų modeliai	17
3.5	Teorinis struktūrinių lygčių modelis	19
3.5.1	Modelio struktūra su varžovų kintamaisiais	22
4	PRAKTINĖ DALIS	25
4.1	Duomenys	25
4.2	Kintamųjų skirstymas	26
4.3	Modelis be varžovų rodiklių, vertinimas	28
4.4	Modelis su varžovų rodikliais, vertinimas	31
4.5	Latentinių kintamųjų vertinimas komandoms	33
4.6	Dirbtinių neuroninių tinklų modeliai	37
4.7	Neuroninių tinklų ir SEM palyginimas - latentinių kintamųjų ir neuronų sąryšis	39
4.8	Prognozavimas	42
4.9	Paklaidų klasifikavimas	44
4.10	Lažybos ir modelių testavimas lažyboms	46
5	Išvados	51
	Literatūros sąrašas	52

1 Santrauka

Metodų ir modelių prognozuoti sporto rungtynių rezultatus per istoriją buvo panaudota daugybė. Tikslios sporto rungtynių prognozės gali būti panaudotos daug kur, o viena iš sričių lažybos. Struktūrinių lygčių modeliai - statistinio modeliavimo technika, kurios pagrindinis tikslas įvertinti priežastinius ryšius tarp stebimų kintamųjų bei teorinių struktūrų. Struktūrinių lygčių modeliai yra naudingi testuojant ir tiriant hipotezes apie įvairias struktūras. Šio darbo tikslas yra sudaryti struktūrinių lygčių modelius, kurie aprašytų faktorius, nulemiančius kiek taškų įmes NBA krepšinio komandos tam tikrose rungtynėse. Paraleliai yra konstruojami neuroninių tinklų modeliai bei ieškomi sąryšiai tarp neuroninių tinklų ir struktūrinių lygčių modelių. Tikrinama, kaip gerai modeliai prognozuoja duomenis, kurie nėra modeliavimo ribose bei atliekama analizė, kuri padeda nusakyti kuriais rodikliais remiantis galima tikėtis, kad prognozė bus tiksli. Galiausiai modeliai yra pritaikomi lažybų strategijoms kurti bei stebima, ar modelių prognozės yra naudingos lažinantis.

There were used a lot of different methods and models to forecast results of sports events during the history. Sports events predictions could be used in various areas and one of them is betting. Structural equation modelling (SEM) - technique of statistical modelling. The main purpose of SEM is to estimate the causal relationships between the observed variables and the theoretical structures. The goal of thesis is to create SEM models by the main factors which affect the amount of points in certain NBA games. Thereby, the neural network models are created and relationship between SEM and neural network models is trying to be find. The accuracy of models's forecast for out of sample data is checked. In the end, models are tested for betting strategies.

2 Įvadas

Nustatyti laiką, kada pirmą kartą buvo pradėta lažintis dėl sporto rungtynių baigčių, yra ganėtinai sunku. Yra žinoma, kad Romos imperijos laikais žmonės statydavo nemažas pinigų sumas už vežimų lenktynes, įvairias kovų dvikovas, atletikos varžybas. Kartais tai darydavo net tada, kai tai būdavo draudžiama įstatymų.

Pirmuosius modernių laikų lažybų etapus apibrėžti kur kas lengviau. 19 amžiaus pirmojoje pusėje, Anglijoje, ypač populiaru buvo sportinis ėjimas. Šalia to atsirado daugybė žmonių, norinčių pasipelnėti iš visuotinio bumo bei sporto šakos populiarumo. Lažintis dėl ėjikų rezultatų tapo labai madinga. Tiesa, ne pačio aukščiausio lygio varžybose, lažybos neretai padarydavo meškos paslaugą ir paveikdavo galutinę lenktynių baigtį, kas ilgainiui sumažino ne tik lažybų, bet ir pačių varžybų populiarumą. Jungtinėse Amerikos Valstijose ėjimas nebuvo toks populiarus kaip kitoje Atlanto vandenyno pusėje, tačiau pati lažybų karštinė pasiekė ir JAV. Čia žmonės daugiausiai lažindavosi dėl beisbolo rungtynių baigčių. Beisbolas ilgą laiką išvengė juodų dėmių dėl galimų rungtynių baigčių sutarimo dar joms neprasidėjus.

Naujas, antrasis rimtas, sporto lažybų etapas prasidėjo su vadinamojo statytojų sklaidos (angliškai point spread) sukūrimu. Lažybų tarpininkai per ilgą laiką patirdavo nemenkus nuostolius, todėl buvo sukurta nauja schema, kuri garantavo jiems didelį pelną. Statytojų sklaidos esmė labai paprasta - esamas koeficientas nusako ne komandų pajėgumo skirtumą, bet tai, kokia dalis statytojų pasirinko vieną ar kitą komandą. Lažybininkų tikslas, jog už abi komandas būtų pastatyta vienoda pinigų suma. Tam tikslui pasiekti atitinkamai yra reguliuojamas koeficientas. Kadangi koeficientai esant lygiems šansams, arba esamoje situacijoje lygiam pasiskirstymui tarp pastatytų pinigų už abi komandas yra mažesni nei 2, dėl to lažybininkai visada tokiu atveju laimi. Tuo tarpu statytojai, net ir atspėdami šiek tiek daugiau nei 50% baigčių, ilgu laikotarpiu patiria nuostolius [10].

Naujausias lažybų etapas - lažybos internete. Kaip ir daugelyje sričių, taip ir lažybose, tai atvėrė naujas galimybes tiek lažybų tarpininkams, tiek lažybininkams. Statymai realiu metu, kuomet mačai dar vyksta, stipriai išaugusi statymų pasiūla. Tai galima apibrėžti kaip trečiąjį sporto lažybų etapą.

Didėjant turimos informacijos kiekiui, dalyvavimas lažybose vis labiau ir la-

biau tampa populiariu ne tik tarp sporto mėgėjų, tačiau ir žmonių, pasiekusių nemažai laimėjimų ir akademinėje veikloje. Praktiškai visi sėkmingiausi pasaulio lažybininkai šiuo metu remiasi tam tikromis, vienokiomis, ar kitokiomis kompiuterinio ir statistinio modeliavimo formomis - pradedant Bayeso ar Monte Carlo metodais, Markovo grandinėmis, k artimiausių kaimynų algoritmais ir kita.

Šio darbo objektas NBA rungtynių rezultatų prognozavimas. Iš esmės tikslas yra ne tai, kuri komanda laimės, bet tai, kiek taškų įmes komandos. Tam tikslui pasiekti yra konstruojamas struktūrinių lygčių modeliai, taip pat dirbtinių neuroninių tinklų modeliai. Lyginami jų rezultatai. Vėliau gauti modelių rezultatai pritaikomi istoriniams NBA rungtynių įmestų taškų prognozavimui ir lažybų simuliacijai. Stebima, ar sukonstruotas modelis tam tikrame periode būtų atnešęs lažybininkui pliusą. Ieškoma strategija, padedanti padidinti pelną.

3 TEORINĖ DALIS

3.1 Sporto rungtynių prognozavimas ir lažybos moksliniuose darbuose

Metodų ir modelių prognozuoti sporto rungtynių rezultatus per istoriją buvo panaudota daugybė. Norui teisingai prognozuoti sporto rungtynių baigtis įtaką daro ne vienas faktorius - bet vieni pagrindinių – noras pasipelnyti ar noras sutaupyti pinigų. Prognozės svarbios sporto aistruoliams, komandų treneriams, rėmėjams, medija atstovams bei paprastiesiems žmonėms, kurie tiesiog stato pinigų sumas lažybų punktuose .

Toli pavyzdžių ieškoti nereikia. Lietuvoje krepšinis yra sportas numeris vienas. Kuomet žaidžia mūsų šalies nacionalinė krepšinio rinktinė Europos, pasaulio čempionato ar Olimpinių žaidynių rungtynes - tuomet televizijų reitingai būna tokiose aukštumose, jog jokios kitos laidos prilygti tam negali. Reklamos rinkoje, didesnių renginių, tokių kaip Europos čempionatas, pasaulio čempionatas, ar Olimpinių žaidynės, reklamos paketai yra parduodami visam čempionatui už iš anksto nustatytą sumą. Todėl reklamuotojams labai svarbu, kaip gerai seksis komandoms, kurios pritraukia dideles žmonių mases prie televizijos ekranų. Ar tai krepšinis Lietuvoje, ar futbolas Vokietijoje, po to, kai nacionalinė komanda iškrenta iš turnyrų, susidomėjimas kitomis čempionato rungtynėmis ženkliai krenta. Įvertinus šansus, kad komanda, kurios pasirodymas yra esminis dideliems reitingams, galima priimti sprendimą, ar televizijos nustatyta pinigų suma už visą čempionato reklamos paketą yra geras sandoris, ar ne.

Sporto lažybos yra iš esmės kur kas paprasčiau suvokiamas procesas, kurio pagalba galima panaudoti tikslus sporto prognozavimo metodus tam, jog užsidirbti pinigų. Be matematinių ir statistinių metodų, galimybės iš to užsidirbti yra gana ribotos. Empiriškai yra ištirta, jog vadinamųjų lažybų ekspertų (angliškai tipsters) sėkmė lažybose yra gana prasto lygio. Šie ekspertai sporto rungtynių baigtis prognozuoja geriau nei atsitiktinis spėjimas, tačiau nusileidžia pačioms paprasčiausioms taktikoms, kaip pavyzdžiui visuomet spėti, jog rungtynes laimės namų komanda. [4] Taigi, ekspertų spėjimai yra labai lengvai nugalimi, todėl stiprių ir matematika pagrįstų modelių svarba, norint pasiekti pelningą rezultatą, čia yra kertinė.

Mokslinių darbų, skirtų prognozuoti sporto rungtynių baigtis, nėra itin daug,

Tyrėjas (Metai)	Duomenų rinkinys, tyrimo objektas	Šaltinis
Kahn (2003)	15 savaitių nacionalinės futbolo lygos rungtynių, 2003 metai	Nacionalinės Futbolo Lygos oficialus puslapis
McCabe ir Travathan (2010)	Australijos futbolo lygos, nacionalinės regbio lygos, Anglijos Premier lygos ir Super Regbio lygos rungtynės.	
Trawinski (2010)	ACB krepšinio lygos statistika iš 2008-2009 sezono	ACB oficialus puslapis
Zdarevski ir Kubakov (2010)	Dviejų NBA sezonų rungtynės	Basketball-Reference puslapis
Miljkovic et. al (2010)	NBA lygos rezultatai 2009-2019	Nba.com, Yahoo.com/NBA
Ivankovic et. al (2010)	Serbijos lygos rezultatai, 2005-2010 metai	Serbijos krepšinio federacijos puslapis
Buursma (2011)	Olandijos futbolo lyga, 15 metų duomenys	Football-data puslapis
Cao (2012)	2005-2012 metų NBA sezonas	Trys skirtingi puslapiai

1 pav.

tačiau pastaruoju metu jų kiekis auga labai sparčiai. Pirmieji straipsniai pasirodė dar praėjusio amžiaus šeštajame dešimtmetyje, kai Moroney (1956), o vėliau Reep et. al (1971) naudojo Puasono ir neigiamą binominį skirstinį tam, jog prognozuotų Anglijos futbolo lygos rungtynėse įmušamų įvarčių kiekį. Abu metodai iš esmės rėmėsi informacija apie dviejų tarpusavyje besivaržančių komandų pajėgumą. Vėliau buvo kuriami sudėtingesni modeliai, tačiau jų patikimumas ir prognozavimo tikslumas nebuvo itin aukštas dėl mažo statistinės informacijos kiekio.

Kai prieinamas duomenų kiekis pradėjo ženkliai augti, rezultatams prognozuoti pradėta naudoti duomenų sijonės metodai, tokie kaip dirbtiniai neuroniniai tinklai, atsitiktiniai medžiai, Bayeso metodai, logistinės regresijos ir kiti. Pastaruoju metu dažniausiai moksliniuose darbose naudotos technikos apžvelgiamos 1 pav. bei 2pav. [5]

Struktūrinių lygčių modeliai taip pat buvo taikyti sporto analitikoje. Tiesa, šiek tiek kitokiame kontekste nei anksčiau minėti darbai. Pavyzdžiui T.Baghal (2012) tyrė daugiau nei 10 NBA sezonų komandų statistinius rodiklius bei juos

susiedamas su komandų atlyginimų dydžiais sudarė struktūrinių lygčių modelį, kurio pagrindinis tikslas prognozuoti komandų pergalių procentą tame pačiame sezone. Sukuriama struktūra, kuri apjungia komandų puolimą, gynybą, atlyginimus ir jų tarpusavio sąveiką iškovotų pergalių kiekiui [1]. Darbe siekiama nustatyti kas nulemia komandų pergalių procentą, bet tai daroma jau po fakto. Pavyzdžiui prognozuojant kiek komanda laimės rungtynių 2000-2001 metų sezone, remiamasi viso to sezono komandų statistika bei to sezono atlyginimų dydžiais. Prognozuoti pergalių procentą galime tik tada, kai sezonas pasibaigęs. Darbe naudojami vos 8 pagrindinės NBA statistinės kategorijos.

Apžvelgus užsienio literatūrą bei pagal turimas autoriaus žinias, struktūrinių lygčių modeliavimas sporto rungtynių baigtims prognozuoti dar iki šiol nebuvo taikytas. Šiame darbe šis metodas bei jo sugretinimas bei palyginimas su viena iš duomenų sijonės technikų yra pagrindinis tikslas.

Tyrėjas (Metai)	Privalumai	Trūkumai
Kahn (2003)	Sukurtų tarpusavyje modelių palyginimas. Taip pat aukštas prognozės tikslumas.	Mažas duomenų rinkinys.
McCabe ir Travathan (2010)	Modelį pritaikė tarptautinėse lažybų varžybose (praktinė nauda).	Prastas prognozės tikslumas.
Trawinski (2010)	Fuzzy algoritmų palyginimas, standartinių paklaidų skaičiavimas kiekvienam modeliui.	Maži, lokalūs duomenys. Tikslumas ne pats geriausias.
Zdarevski ir Kubakov (2010)	Automatinis duomenų surinkimas.	Prastas prognozės tikslumas. Nėra palyginimų su kitais analogiškais tyrimais.
Miljokovic et. al (2010)	Duomenų atnaujinimas kiekvienai dienai.	Prastas prognozės tikslumas.
Ivankovic et. al (2010)	Įvairių tipų modeliavimas. Įvertinami skirtingo tipo puolimo taktikos. Aukštas prognozės tikslumas.	Duomenys nėra viešai prieinami, taip panaikinama galimybė palyginti tyrimą su analogiškais darbais.
Buursma (2011)	Modelio funkcijos parinkimas nurodant funkcijoje esančių parametrų kiekį.	Prognozės tikslumas žemas, nėra palyginimo su kitais darbais.
Cao (2012)	Duomenų valdymas, automatinis surinkimas bei palyginimas su kitais darbais.	Prognozės tikslumas prastas, modelio parametrai priklauso nuo pastarųjų 10 rungtynių, negalima prognozuoti pirmųjų 10 mačų.

2 pav.

3.2 NBA statistika ir istorija

Pirmosios NBA rungtynės įvyko 1946 metais. Jau tada buvo sekamos pagrindinės rungtynių statistinės charakteristikos, kaip pelnyti taškai, išmesti ir

pataikyti metimai, mestos baudos. Laikui bėgant, statistinių kategorijų kiekis augo ir informacijos apie įvykusias rungtynes buvo surenkama vis daugiau ir daugiau.

Nuo 1950-1951 metų NBA sezono pradėta sekti atkovotų kamuolių statistika. Perimtų kamuolių, blokų, gynyboje atkovotų kamuolių rodikliai buvo įtraukti nuo 1973-1974 metų NBA sezono. Dar po metų pradėta sekti kiek kartų komandos ir žaidėjai per rungtynes suklysdavo. Kai 1979-1980 metais buvo įvesta trijų taškų taisyklė, ši statistinė kategorija papildė sekamų rodiklių rinkinį. Nuo 2003 metų sekama daugiau rodiklių, pavyzdžiui kiek esant tam tikram žaidėjui komanda pelnė daugiau ar mažiau taškų nei varžovai. Taip pat įvairūs išvestiniai rodikliai, kaip gynybos ir puolimo reitingai.

Visgi revoliucija buvo padaryta prieš tris metus, kuomet NBA pradėjo sekti ir pateikti taip vadinamąją pažangiąją statistiką, kuri leidžia kur kas detaliau ir giliau analizuoti rungtynių eigą. Pagrindinės naujovės su pažangiosios statistikos įvedimu yra:

- Greitis: Nurodo, kiek atakų per 48 minučių rungtynes turi komanda. Kadangi komandos žaidžia skirtingu greičiu, vienos rengia ilgesnes, kitos trumpesnes atakas, dėl to šis rodiklis gali ženkliai skirtis.
- Efektyvus pataikymo procentas: Šis matavimo dalis leidžia nustatyti, kaip gerai tam tikri žaidėjai pataiko įvertinant ne tik pataikymo procentą, bet ir metimų sudėtingumą bei svarbą.
- Komandos pasirodymai su skirtingais penketais: Matuojama, kaip komanda ginasi, atakuoja ir bendrai, kokius rezultatus demonstruoja su skirtingais žaidėjais aikštėje. Matuojama, kaip pasikeičia komandos puolimas ir gynyba esant skirtingiems žaidėjams aikštėje. Taip galima įvertinti kiekvieno žaidėjo svarbą komandai.
- Žaidėjų sekimas: Kiekvienoje NBA arenoje yra šešios kameros, įrengtos virš aikštės. Jos seka kiekvieną žaidėją, kiekvieną jo judesį. Kamera per sekundę užfiksuoja 25 skirtingus kadrus. Ši technologija leidžia surinkti tokias statistines charakteristikas kaip žaidėjų greitis, nubėgtas atstumas, žaidėjų išsisklaidymas puolime, kiek kartų žaidėjai lietė kamuolį, kiek buvo kamuolys tam tikroje aikštės zonoje ir t.t.

- Žaidimo tipai: Stebima, kiek komandos ir žaidėjai naudoja įvairių žaidimo tipų. Kiek kartų žaidžia du prieš du, kiek kartų buvo pelnyti taškai po pastatytos sienelės, ar komanda daugiau remiasi žaidimu vienas prieš vieną, ar komandiniu krepšiniu ir t.t.

Sunku net suskaičiuoti, kiek NBA oficiali svetainė pateikia skirtingų statistinių rodiklių. Mūsų tyrime bus naudojama 230. Jų aprašymas yra pateiktas priede.

3.3 Struktūrinių lygčių modeliai

Struktūrinių lygčių modeliavimas arba tiesiog SEM yra statistinio modeliavimo technika. Į ją galima žiūrėti kaip į faktorinės, kelių ir regresinės analizės derinį. Pagrindinis SEM tikslas įvertinti priežastinius ryšius tarp stebimų kintamųjų bei teorinių struktūrų, kurios tiesiogiai nėra matuojamos (latentinių kintamųjų). Sąryšiai tarp teorinių struktūrų yra apibrėžiami per regresijos koeficientus, taip pat vertinama kovariacijų struktūra tarp stebimų kintamųjų. Struktūrinių lygčių modeliai yra patogūs ir lengvai suprantami modeliai - dažniausiai jie yra atvaizduojami trajektorijų schemomis. [3]

Biologas Sewell Wright (1918, 1921, 1934) išplėtojo iki tol nematytą modelių tipą - kelių modeliavimą (angliškai path model). Kelių modeliai naudoja korelacijs koeficientus bei regresinę analizę tam, jog modeliuoti sudėtingesnius sąryšius tarp stebimų kintamųjų. Pirmieji trajektorijų modeliai buvo naudojami tirti gyvūnų elgsenai. Vėliau panašaus tipo modeliai buvo naudoti ekonomistų Trygve Haavelmo (1943) bei Hebert Simon (1953) darbuose. Septintojo praejusio amžiaus dešimtmečio pabaigoje ir aštuntojo dešimtmečio pradžioje trys mokslininkai Karl Joreskog, Ward Keesling ir David Wiley padėjo rimtą pamatą SEM modelių atsiradimui. Iš pradžių modelis buvo netgi žinomas kaip JKW (pagal autorių pavardžių pirmąsias raides), bet vėliau išpopuliarėjo ir tapo žinomas kaip tiesinis struktūrinis sąryšiu modelis (angliškas trumpinys LISREL). Vėliau tokiu pačiu pavadinimu, LISREL, buvo sukurta pirmoji programinė įranga šių modelių vertinimui.

SEM yra panašus į tradicinius duomenų analizės metodus kaip koreliacinė, regresinė ir dispersinė analizė ne vienu požymiu. Pirma, tiek SEM, tiek tradiciniai metodai remiasi tiesiniais statistiniais modeliais. Statistiniai testai, nusakantys modelių tinkamumą, yra tikslūs tik tuomet, kai tam tikros prielaidos

galioja, pavyzdžiui normalumo prielaida. SEM daro prielaidą apie daugiamatį normalųjį stebimų kintamųjų pasiskirstymą.

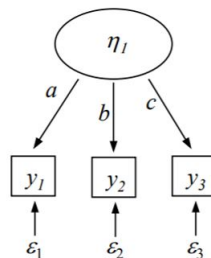
3.3.1 Latentiniai kintamieji

Latentiniai kintamieji yra viena iš pagrindinių struktūrinių modelių dalių. Latentiniai kintamieji, tai kintamieji, kurie nėra tiesiogiai stebimi, bet yra išvedami iš kintamųjų, kurie yra stebimi arba tiesiogiai matuojami. [6] Geriausias latentinio kintamojo pavyzdys yra žmogaus intelektas, kuris tiesiogiai nėra išmatuojamas, tačiau jis gali būti įvertintas, pavyzdžiui, žmogui davus atlikti tam tikrą testą ar atsakius į tam tikrus klausimus.

Latentinių kintamųjų struktūra gali būti paaiškinta klasikine testų teorija, kuri teigia, jog kiekvienas stebimas dydis yra dviejų kintamųjų funkcija. Formaliai:

$$X = T + \epsilon$$

Čia X yra tam tikras gautas eksperimento rezultatas, T – tiriamojo objekto tikrasis lygis, ϵ – paklaida. Šiuo atveju latentiniu kintamuoju būtų T – objekto tikrasis lygis, kurio tiesiogiai išmatuoti negalime. Tam tikro lygio paklaidos gali atsirasti dėl eksperimento netobulo, užduoties formulavimo ir pan. Praktikoje, norint įvertinti latentinį kintamąjį, pageidautina turėti bent tris kintamuosius, kurie aprašo latentinį kintamąjį. Pateiksime tipinį latentinio kintamojo struktūros pavyzdį, su trimis faktoriais. Grafinis atvaizdavimas pateiktas 3 paveiklyje. Graikiška raidė η žymi latentinį kintamąjį, ϵ žymi paklaidas. Raidės a , b , c žymi faktorių svorius. Jų įvertinimo metodai yra aprašyti vėlesniuose skyriuose. Iš esmės interpretacija yra tokia latentinis kintamasis η nusako kokio dydžio rodikliai bus y_1 , y_2 , y_3 . Kadangi visos kintamųjų y_1 , y_2 , y_3 dispersijos latentinis kintamasis nepaaiškina, egzistuoja paklaidos ϵ_1 , ϵ_2 , ϵ_3



3 pav.

3.3.2 Parametrų vertinimas

Skirtingai, nei pavyzdžiui ANOVA ar regresinėje analizėje, SEM naudoja iteracijų vertinimo metodus - dažniausias iš jų yra didžiausio tikėtimumo metodas. Iteracijų būdu bandoma gauti tokias modelio parametrų reikšmes, kurios geriausiai tiktų aprašytam modeliui ir turimiems duomenims.

Žymėjimai:

- Σ - populiacijos variacijos - kovariacijos matrica
- S - imties variacijos koariacijos matrica
- $\Sigma(\theta)$ – populiacijos variacijos - kovariacijos matrica, su apibrėžto modelio apribojimais
- $\Sigma(\hat{\theta})$ - imties variacijos - kovariacijos matrica, su apibrėžto modelio apribojimais

Bendra paklaida: $\Sigma - \Sigma(\hat{\theta})$

Aproksimacijos paklaida: $\Sigma - \Sigma(\theta)$

Įvertinimo paklaida: $\Sigma(\theta) - \Sigma(\hat{\theta})$.

Didžiausio tikėtimumo metodas remiasi idėja, jog žinant tikrąją populiacijos variacijos-kovariacijos matricą Σ galime įvertinti tikimybę gauti bet kurią imties kovariacijų matricą S . Didžiausio tikėtimumo funkcija atrodo taip:

$$F(S, \Sigma(\hat{\theta})) = (\ln | \Sigma(\hat{\theta}) | - \ln | S |) + [\text{trace}(S \cdot \Sigma(\hat{\theta})^{-1}) - k]$$

Čia k – stebimų kintamųjų kiekis modelyje.

SEM modelyje tiek S , tiek $\Sigma(\hat{\theta})$ yra populiacijos variacijos - kovariacijų matricos Σ įverčiai. Skirtumas tas, jog $\Sigma(\hat{\theta})$ matrica yra apribota tyrėjo sudaryto modelio, o S ne. Didžiausio tikėtimumo metodu yra ieškoma tokio parametrų rinkinio, kuris maksimizuotų tikimybę, kad S yra gauta iš $\Sigma(\theta)$, darant prielaidą, jog $\Sigma(\hat{\theta})$ yra geriausias $\Sigma(\theta)$ įvertis. [3]

Pastebėkime, jog jei didžiausio tikėtimumo funkcijoje S ir $\Sigma(\hat{\theta})$ yra identiškos, pirmoji dalis funkcijos yra lygi nuliui. Jeigu abi matricos yra lygios, tuomet S ir $\Sigma(\hat{\theta})^{-1}$ įstrižainės elementų suma bus lygi k , todėl ir antroji reiškinio dalis bus lygi nuliui.

Modelio parametrų vertinimas remiasi iteraciniu metodu. Iš pradžių yra startuojame su tam tikromis pradinėmis reikšmėmis (jos gali būti parinktos tyrėjo, pavyzdžiui 2 žingsnių mažiausių kvadratų metodu ar kitais metodais). Su

pradinėmis reikšmėmis yra išsprendžiamos modelio lygtys bei gauta kovariacijų matrica $\Sigma(\hat{\theta}_1)$ yra lyginama su S . Algoritmo tikslas – surasti tokias modelio parametrų reikšmes, jog jos minimizuotų modelio kvadratinę paklaidą. Gautas įvertinių rinkinys naudojamas vietoje pradinių reikšmių ir yra gaunama modelio apribota variacijos-kovariacijos matrica $\Sigma(\hat{\theta}_2)$. Lyginama ar pirmojo modelio ir antrojo modelio paklaidų dydis sumažėjo. Iteracijos procesas baigiasi, kuomet naujai sudarytas įverčių rinkinys negali sumažinti paklaidų dydžio.

3.3.3 Modelio identifikavimas

Modelio indentifikacijos prasmė gana paprasta - parametrų kiekis modelyje turi būti lygus, arba mažesnis nei stebimų dydžių dispersijų ir kovariacijų kiekiui modelyje. Skirtumas stebimų dispersijų ir kovariacijų bei parametrų kiekio modelyje yra lygus modelio laisvės laipsnių skaičiui.

k raide pažymėkime kiekį stebimų dydžių yra įtrauktų į modelį. Tuomet iš viso variacijos - kovariacijos matricoje bus $k(k+1)/2$ skirtingų kovariacijų arba dispersijų. Jeigu laisvės laipsnių kiekis yra neigiamas, tai yra parametrų kiekis modelyje yra didesnis nei stebimų kintamųjų kiekis - modelis yra neidentifikuotas ir modelio parametrų įvertinti negalime. Jeigu laisvės laipsnių skaičius lygus 0, tuomet modelis yra tiksliai identifikuotas. Tuomet unikalūs vertinamų parametrų rinkinys gali būti rastas, tačiau modelio gerumo charakteristikų įvertinti negalima. Tuo tarpu jeigu laisvės laipsnių kiekis didesnis už 0, modelis yra peridentifikuotas. Tai turi teigiamos prasmės tuo, jog galime įvertinti parametrus taip, jog jie geriausiai tiktų duomenims bei sudarytam modeliui.

3.3.4 Statistika ir gerumo matai. SEM modelio sudarymo procesas

Norint nustatyti struktūrinių lygčių modelių adekvatumą, keletas statistikų yra apskaičiuojama.

χ^2 testas nurodo skirtumą tarp laukiamos ir stebimos kovariacijų matricų. Maža χ^2 statistikos reikšmė nusako, jog skirtumas tarp stebimos ir laukiamos kovariacijos matricos yra nedidelis. Statistikos reikšmė apskaičiuojama pagal formulę, ja remiantis daroma išvada ar tyrėjo sudaryta struktūra atitinka turimus duomenis:

$$\chi^2 = (n - 1)F(S, \Sigma(\hat{\theta}))$$

Kitas modelio gerumo matas - CFI indeksas. CFI kinta intervale $[0; 1]$. Kuo

didesnė CFI reikšmė, tuo modelis yra labiau tinkamas. Modeliai su CFI reikšme, didesne nei 0,9 yra pakankamai geri [9]. CFI indeksas yra apskaičiuojamas remiantis tokiomis formulėmis:

$$d = \chi^2 - df$$

čia df - laisvės laipsnių skaičius modelyje

$$CFI = \frac{d(Null) - d(Model)}{d(Null)}$$

čia $Null$ žymime nulinį modelį, o $Model$ - tyrėjo sudarytą modelio struktūrą. Čia nulinis modelis reiškia, jog kovariacijos tarp latentinių kintamųjų yra nulinės. Kitaip tariant iš anksto aprašyta struktūra iš tiesų neturi jokių tiesioginių ryšių.

Itin populiarus indeksas matuoti SEM modelio adekvatumą yra RMSEA. RMSEA reikšmės taip pat svyruoja intervale [0,1]. Maža RMSEA reikšmė reiškia, jog modelis yra adekvatus. RMSEA reikšmė svyruojanti tarp 0,05 ir 0,1 laikoma požymiu, jog duomenys tinka aprašytai SEM struktūrai [2]. RMSEA apskaičiuojamas remiantis šia formule:

$$RMSEA = \frac{\sqrt{d}}{\sqrt{df(N-1)}}$$

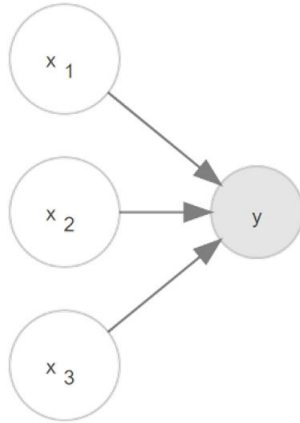
SEM modelio sudarymo procesas:

- Modelio specifikacija, remiantis tam tikra teorija arba ekspertine patirtimi. Sudaromas grafinis modelis, sąryšiai tarp kintamųjų, lygtys.
- Kintamųjų, kurie aprašys modelį parinkimas
- Modelio identifikacija
- Duomenų rinkimas
- Parametrų vertinimas
- Modelio vertinimas
- Interpretacija, panaudojimas

3.3.5 Grafinis modelio atvaizdavimas

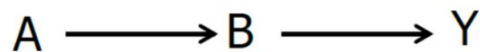
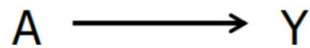
Kaip minėjome, struktūrinių lygčių modeliavime yra svarbus grafinis atvaizdavimas. Šiame skyriuje pateikiami grafinių struktūrinių lygčių pavyzdžiai.

Pradėkime nuo paprasto tiesinio regresinio modelio. Regresinio modelio pagalba norime įvertinti ryšį tarp tam tikrų kintamųjų. Grafiškai tai atrodo taip: siekiama nustatyti priklausomybę tarp y ir jį aprašančių kintamųjų x_1 , x_2 , x_3 (4 pav.).



4 pav.

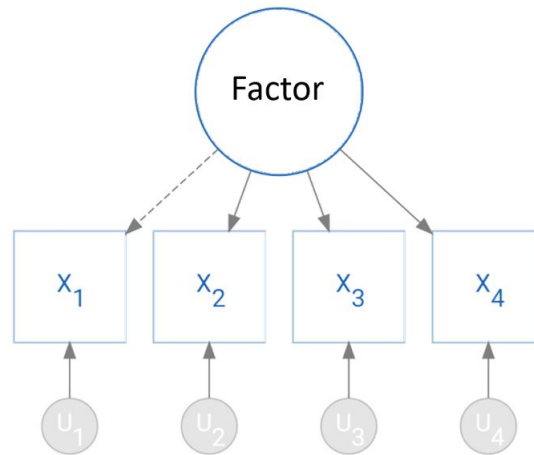
Kelių analizė, parodo kaip kintamieji yra tarpusavyje susiję. Galimi keli variantai: pavyzdžiui tiesioginis sąryšis, netiesioginis sąryšis. Sąryšių grafinis pavyzdys pateiktas 5 pav. Tiesioginio sąryšiu atveju stebime kaip kintamasis A veikia kintamąjį Y . Netiesioginis sąryšis yra toks, kai kintamasis A veikia kintamąjį B , o tas, savo ruožtu veikia kintamąjį Y .



5 pav.

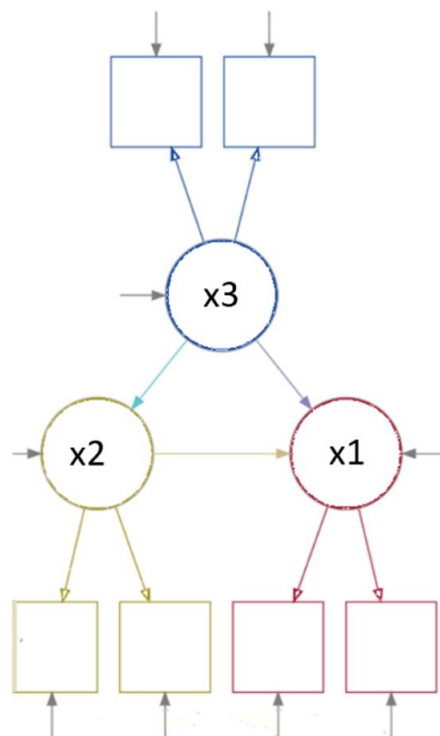
Latentiniai kintamieji schemose vaizduojami apskritime. Pateiktoje schemoje (6 pav.), kiekvienas stebimas dydis x_i $i=1,..4$, yra funkcija nuo latentinio kintamojo. Taip pat modelyje yra paklaidos u_i , $i=1,..4$. Tai dalis stebimų

dydžių variacijos, kurios latentinis kintamasis nepaaiškina. [7]



6 pav.

Struktūrinių lygčių modelio pavyzdys pateiktas 7 paveikslėlyje. Modelyje yra trys latentiniai kintamieji - x_1 , x_2 , x_3 . Visi jie yra aprašomi iš dviejų skirtingų kintamųjų. Tiesa, x_1 yra koreliuotas tiek su x_2 , tiek su x_3 . x_2 yra koreliuotas su x_3 .



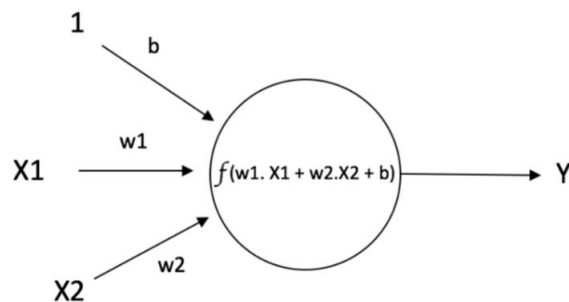
7 pav.

3.4 Neuroninių tinklų modeliai

Šiame skyriuje pateiksime pačią pagrindinę informaciją apie dirbtinius neuroninius tinklus. Kadangi neuroniniai tinklai nėra pagrindinis darbo tikslas, išsamesnės informacijos apie pačius neuroninius tinklus čia nepateiksime. Aprašymas sudarytas remiantis [12].

Dirbtiniai neuroniniai tinklai - skaičiavimo modeliai, kurie yra paremti biologiniu neuronų darbu žmogaus smegenyse. Šie modeliai taikomi klasifikavimo uždaviniuose, prognozavime, optimizavime bei kitose srityse.

Pagrindinis modelio vienetas yra neuronas. Neuronas su jėjimais (stebimais ar matuojamais dydžiais) yra susijęs per tam tikrus svorius. Iš gautų įvesčių su svoriais, rezultatai neuronuose gaunami per funkciją $f()$, kuri įvairiais atvejais yra skirtinga.



8 pav.

8 paveikslėlyje pateikta pati paprasčiausia dirbtinio neuroninio tinklo schema. Du įėjimo taškai su svoriais w_1 bei w_2 (čia w_1 , w_2 yra realieji skaičiai) bei 1 su svoriu b . Jį galime laikyti analogu regresijoje esančios konstantos. Rezultatas yra funkcija $f()$ nuo gautosios tiesinės įėjimo taškų ir svorių kombinacijos. Praktikoje neuroninių tinklų modeliai gali pateikti ne vieną rezultatą, bet ir daugiau, tačiau šiame darbe tokie modeliai nėra aktualūs, todėl jų nepapildinsime. Funkcija $f()$ yra netiesinė ir dažniausiai vadinama aktyvacijos funkcija. Aktyvacijos funkcijos tikslas aprašyti egzistuojantį netiesiškumą tarp įėjimo taškų bei galutinio rezultato. Praktikoje dažniausiai naudojamos kelios aktyvacijos funkcijos:

- *Sigmoidinė funkcija* realias reikšmes paskirsto intervale (0;1):

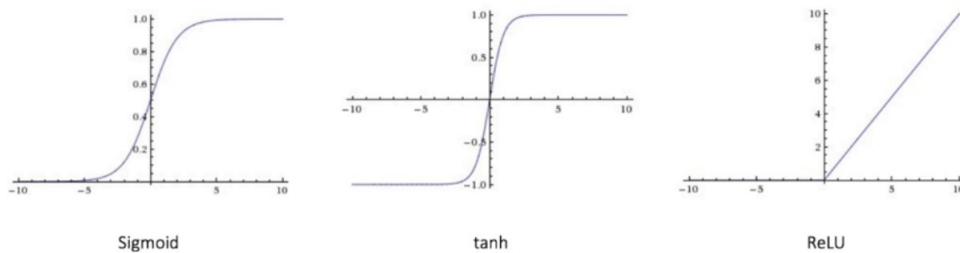
$$f(a) = \frac{1}{1 + e^{-a}}$$

- *Hiperbolinio tangento* funkcija realias reikšmes paskirsto intervale $[-1;1]$:

$$f(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

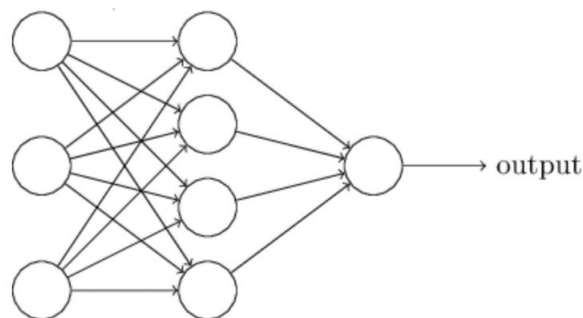
- *ReLU funkcija* realias neigiamas reikšmes pakeičia į nulį, tuo tarp teigiamas reikšmes palieka tokias pačias.

$$f(a) = \max \{0, a\}$$



9 pav.

Modelyje gali egzistuoti ne vienas neuronas. Tuomet neuronai informaciją iš įėjimo taškų gaus su skirtingais svoriais. Neuronai yra taip vadinamajame paslėptajame sluoksnyje. Paslėptų sluoksnių modelyje gali būti daugiau nei vienas, tačiau daugumai atvejų, vieno paslėptojo sluoksnio yra pakankamai, norint gauti gerus rezultatus [11]. Atveju, kuomet slaptajame sluoksnyje yra daugiau nei vienas neuronas, galutinis rezultatas yra aktyvacijos funkcija nuo gautų neuronų reikšmės su tam tikrais svoriais w_i .



10 pav.

3.5 Teorinis struktūrinių lygčių modelis

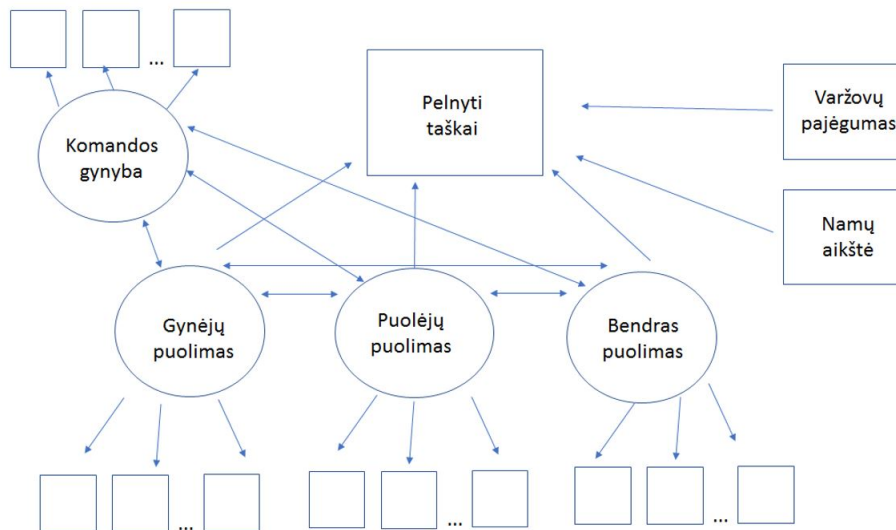
Pagrindinis viso tyrimo tikslas – nustatyti veiksnius, kurie veikia komandų pelnomų taškų kiekį per rungtynes ir šį rodiklį prognozuoti. Visų pirma buvo sukurta pati idėja ir modelis kaip veikia žaidimas ir pačios žaidimo charakteristikos. Kitaip tariant sukurta pati pirminė struktūra, kurią testavome, vėliau tobulinome, o galiausiai jos pagalba prognozavome norimus dydžius.

Labai paprastai tariant rodiklį, kiek komanda įmes taškų, nulemia iš esmės keli pagrindiniai dalykai. Kiek stiprus komandos puolimas yra neabejotinai pats svarbiausias faktorius, nulemiantis tai, kiek komanda įmes taškų tam tikroje rungtynėse. Visgi, taip apibendrinti visko negalime. Komandos yra skirtingos, vienos turi stiprius aukštaūgius ir koncentruojasi į žaidimą po krepšiu, išnaudodami jų meistriškumą, kitos, pasikliauna gynėjų žaidimu, pataikymu iš tritaškio zonos, jų prasiveržimais, žaidimo „skaitymu“. Kitos komandos renkasi šių dviejų, iki šiol įvardintų taktikų mišinį, remiasi ne kažkokios vienos grandies pranašumais, o bendru komandos žaidimu, tarpusavio supratimu, kamuolio judėjimu ir dalinimusi, gerais deriniais, o ne pavienių žaidėjų meistriškumu. Iš to galime sudaryti tris bendrus naujus kintamuosius, kurie geriausiai apibūdintų tris skirtingas puolimo taktikas ir komandų stiprybes – gynėjų puolimas (aprašo dažniausiai žemesnių žaidimą, tritaškius ir t.t.), puolėjų puolimas (aprašo aukštaūgių puolimą – pelnytus taškus arčiau krepšio, atkovotus kamuolius puolime ir t.t.) bei bendrą puolimą (kuris nurodo bendrą komandos lygį, kamuolio dalijimąsi, pataikymą ir pan.). Be to, tikėtina, kad egzistuoja tarpusavio sąveika tarp visų kintamųjų. Tai yra, geri puolėjai atlaisvina rankas mažesniems žaidėjams ir šie gali įmesti daugiau taškų ir atvirkščiai, geri gynėjai sukuria geras situacijas aukštaūgiams. Todėl modelyje bus daroma prielaida, kad šie kintamieji yra tarpusavyje susiję.

Nors tiesiogiai komandos gynyba taškų neprideda, tačiau tai gali būti netiesioginis veiksnys, leidžiantis komandai pelnyti taškus. Pavyzdžiui perimti kamuoliai suteikia progą komandai pabėgti į greitas atakas ir pelnyti lengvus taškus. Kitas pavyzdys – atkovoti kamuoliai. Atkovoti kamuoliai suteikia šansą komandai pradėti ataką, kuomet varžovai dar neišsirikiavo į gynybą ir, tarkime, negalės padėti vienas kitam žaidime 1 prieš 1, kur varžovai galbūt yra labai stiprūs.

Taigi, pradiniam modelyje buvo testuojama struktūra su keturiais laten-

tiniais kintamaisiais – gynėjų puolimas, puolėjų puolimas, bendras komandos puolimas ir komandos gynyba. Tiesa, komandos gynyba savaime nebus įtraukta į regresiją, kaip aiškinantys kintamasis pelnytiems taškams. Bus vertinama jo ir kitų latentinių kintamųjų tarpusavio sąveika. Visgi, tai dar toli gražu ne visi rodikliai, kurie galėtų padėti prognozuoti taškų kiekį, kurį pelnys komanda. Įtakos įmestų taškų kiekiui turi faktas, ar yra žaidžiama namuose, koks yra komandų pajėgumo skirtumas – tai yra tikėtina, jog žaidžiant labai stiprioms ir labai silpnoms komandoms (pergalių prasme) įmestų taškų kiekis bus visai kitoks nei žaidžiant komandos turinčioms panašų pergalių santykį, nors tarkime, puolimo ir gynybos lygiai būtų panašūs abejais atvejais. Pradinis modelis grafiškai pateiktas 11 paveikslėlyje.



11 pav.

Modelio struktūra formaliai pateikta pirmojoje ir antrojoje formulėse.

$$\left\{ \begin{array}{l}
x_{i,1} = f_1(Gpuol) + \epsilon_{i,x1} \\
\quad \dots \\
x_{i,k} = f_k(Gpuol) + \epsilon_{i,xk} \\
z_{i,1} = g_1(Ppuol) + \epsilon_{i,z1} \\
\quad \dots \\
z_{i,l} = g_l(Ppuol) + \epsilon_{i,zl} \\
r_{i,1} = h_1(Bpuol) + \epsilon_{i,r1} \\
\quad \dots \\
r_{i,c} = h_c(Bpuol) + \epsilon_{i,rc} \\
t_{i,1} = j_1(Gyn) + \epsilon_{i,t1} \\
\quad \dots \\
t_{i,s} = j_s(Gyn) + \epsilon_{i,ts} \\
y_i = \beta_1 Gpuol_i + \beta_2 Ppuol_i + \beta_3 Bpuol + \beta_4 w_i + \beta_5 h_i + \epsilon_i
\end{array} \right. \quad (1)$$

Ir:

$$\left\{ \begin{array}{l}
Gpuol \sim Ppuol \\
Gpuol \sim Bpuol \\
Bpuol \sim Ppuol \\
Gpuol \sim Gyn \\
Ppuol \sim Gyn \\
Bpuol \sim Gyn
\end{array} \right. \quad (2)$$

Čia x_1, \dots, x_k žymi k stebimų kintamųjų, kuriais yra aprašomas latentinis kintamasis gynėjų puolimas. Atitinkamai z_1, \dots, z_l , r_1, \dots, r_i bei t_1, \dots, t_s yra stebimi dydžiai, kuriais aprašomi latentiniai kintamieji puolėjų puolimas, bendras puolimas ir gynyba. Funkcijos $f()$, $g()$, $h()$, $j()$ su atitinkamais indeksais, šiuo atveju yra tiesinės funkcijos, nusakančios koku svoriu latentinis kintamasis veikia stebimą dydį. Ko nepaaiškina latentinis kintamasis, tas patenka į paklaidas ϵ su atitinkamais indeksais. Skirtingų indeksų ϵ dispersija gali skirtis. Kintamieji w bei h atitinkamai aprašo stebimos komandos ir varžovų pajėgumo sanykį, o h – indikatorius, įgyjantis reikšmę 1, jeigu komanda žaidžia namuose ir 0 – jei išvykoje. y žymime komandos įmestų taškų kiekį. \sim simboliu žymime faktą, jog kintamieji tarpusavyje yra koreliuoti – tai yra latentiniai kintamieji tarpusavyje turi tam tikrą tiesinį ryšį.

Aprašant struktūrinių lygčių modelio struktūrą, šis faktas yra svarbus, kadangi vertinant parametrų reikšmes, yra lyginamos modelio apribota variaci-

jos - kovariacijų matrica $\Sigma(\hat{\theta})$ bei neapribotoji S . Jeigus šis sąryšis modelio struktūroje būtų neaprašytas, apribotoje variacijos - kovariacijų matricoje prie atitinkamų kintamųjų būtų nuliai.

3.5.1 Modelio struktūra su varžovų kintamaisiais

Komandos puolimas, be abejonių, labai svarbus rodiklis, nusakantis, pelnytų komandų taškų kiekį tam tikrose rungtynėse. Visgi ne ką svarbesnį vaidmenį vaidina tai, prieš kokį varžovą yra žaidžiama - koks yra priešininkų gynybos lygis. Kai kuriais atvejais, tai gali būti net gi svarbesnis veiksnys. Daugelyje atvejų, kuomet bendrai komandos puolimo lygis, bent pagal statistinius rodiklius, atrodo labai gerai – pataikymo procentas, rezultatyvių perdavimų rodikliai, taškų kiekis per 100 atakų, tačiau susidūrus su gerai organizuota ir gerai subalansuota varžovų gynyba, visa puolimo jėga tiesiog dingsta ir matomi rezultatai yra ženkliai prastesni nei iki tol demonstruoti vidutiniai rodikliai. Dėl šių priežasčių į modelį įtrauksime dar vieną latentintį kintamąjį, kuris aprašys varžovų komandos gynybos lygį. Kaip ir anksčiau buvo aprašyta, šalia jo bus dar vienas latentinis kintamasis – nurodantis bendrą varžovų puolimo lygį. Kitaip tariant kintamieji yra koreliuoti. Prieš tai aprašėme tai, jog gera gynyba gali paveikti puolimą.

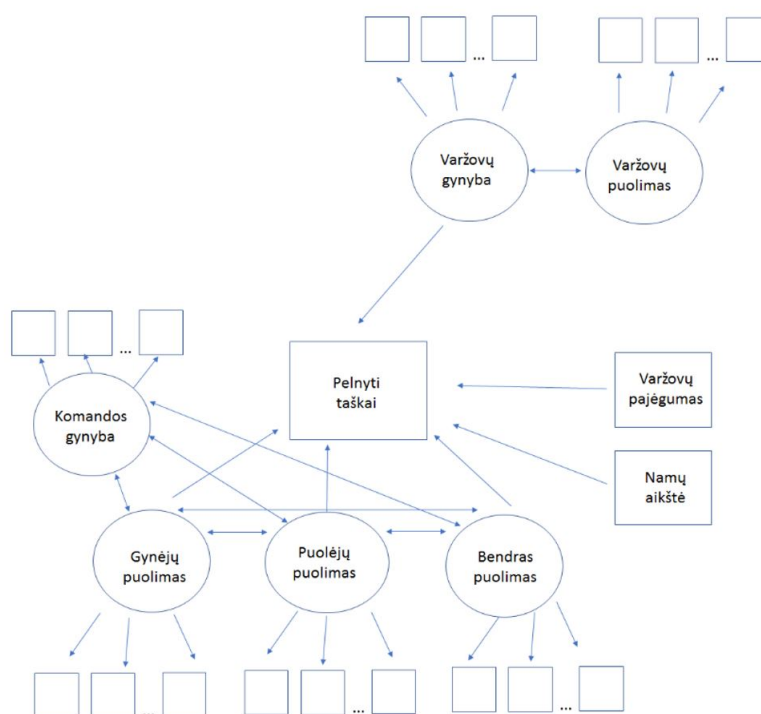
$$\left\{ \begin{array}{l}
x_{i,1} = f_1(Gpuol) + \epsilon_{i,x1} \\
\cdots \\
x_{i,k} = f_k(Gpuol) + \epsilon_{i,xk} \\
z_{i,1} = g_1(Ppuol) + \epsilon_{i,z1} \\
\cdots \\
z_{i,l} = g_l(Ppuol) + \epsilon_{i,zl} \\
r_{i,1} = h_1(Bpuol) + \epsilon_{i,r1} \\
\cdots \\
r_{i,c} = h_c(Bpuol) + \epsilon_{i,rc} \\
t_{i,1} = j_1(Gyn) + \epsilon_{i,t1} \\
\cdots \\
t_{i,s} = j_s(Gyn) + \epsilon_{i,ts} \\
p_{i,1} = j_1(Vgyn) + \epsilon_{i,p1} \\
\cdots \\
p_{i,s} = j_s(Vgyn) + \epsilon_{i,ps} \\
e_{i,1} = u_1(VPuol) + \epsilon_{i,e1} \\
\cdots \\
e_{i,b} = u_b(VPuol) + \epsilon_{i,eb} \\
y_i = \alpha + \beta_1 Gpuol_i + \beta_2 Ppuol_i + \beta_3 Bpuol + \beta_4 w_i + \beta_5 h_i + \beta_6 Vgyn_i + \epsilon_i
\end{array} \right. \quad (3)$$

Ir:

$$\left\{ \begin{array}{l}
Gpuol \sim Ppuol \\
Gpuol \sim Bpuol \\
Bpuol \sim Ppuol \\
Gpuol \sim Gyn \\
Ppuol \sim Gyn \\
Bpuol \sim Gyn \\
Vgyn \sim VPuol
\end{array} \right. \quad (4)$$

Tačiau iš kitos pusės - jeigu puolimas geras, žaidėjai mažai klysta, įmeta metimus, varžovams yra leidžiama kur kas mažiau progų išbėgti į greitą puolimą ir pelnyti lengvus taškus. Tokia logika vadovaujantis latentintis kintamasis varžovų puolimas sąvekauja su varžovų gynyba, bet jis tiesiogiai neveikia stebimos komandos taškų kiekio. Grafinis šio modelio vaizdas pavaizduotas paveikslėly-

je. Formaliai modelio lygtys yra užrašytos formulėje aukščiau. Visas aprašymas ir konstrukcija analogiška modeliui be varžovų kintamųjų. Skirtumas tas, jog prisideda du latentiniai kintamieji. Pastebėkime, jog varžovų gynybos ir stebimos komandos gynybos latentinių kintamųjų struktūra sutampa – funkcijos aprašančios latentinio kintamojo ir stebimo kintamojo sąveiką yra tos pačios. Kintamųjų kiekis, aprašantis patį latentinį kintamąjį yra toks pats. Skirtumas tas, jog varžovų gynyba vėliau yra tiesiogiai naudojama prognozuoti įmestų taškų kiekį.



12 pav.

4 PRAKTINĖ DALIS

4.1 Duomenys

Tyrimo tikslas – nusakyti NBA komandų įmetamų taškų struktūrą, tam tikrose, skirtingose rungtynėse. Tiriamas laikotarpis 2016-12-04 – 2017-03-31.

Per visą tirtą laikotarpį NBA arenose buvo sužaistas 761 mačas. Todėl tiriama objektų yra dvigubai daugiau – 1522, kadangi iš kiekvienerių rungtynių gaunami dviejų komandų rezultatai. Periodas pasirinktas neatsitiktinai. 2016-2017 metų NBA reguliarusis sezonas startavo 2016 metų spalio 25 dieną, o baigėsi kitų metų balandžio 17 dieną. Iki gruodžio 4 dienos visos lygos komandos buvo sužaidusios bent 20 rungtynių. 20 rungtynių yra pakankamas kiekis, jog paaiškėtų komandų stipriosios ir silpnosios pusės, kadangi su įvairaus lygio varžovais per tiek rungtynių tenka susidurti. Paskutinių 12 dienų mačai į tyrimą neįtraukiami dėl keleto priežasčių. Paskutinės reguliariojo sezono rungtynės daugeliui klubui praktiškai nieko nelemia – jų padėtis turnyrinėje lentelėje yra aiški. Kai kurios komandos – norėdamos gauti geresnį šaukimą artėjančių metų naujokų biržoje tyčia taupo savo žaidėjus ir nori pralaimėti rungtynes. Kitos komandos taip pat taupo savo lyderius, kadangi artėja atkrintamosios varžybos, kuriose lyderių gero žaidimo reikės dar labiau.

Duomenų struktūra yra labai paprasta. Ją sudaro iš esmės keturi skirtingi blokai. Pirmasis blokas – modeliuojamas kintamasis – stebimos komandos pelnyti taškai. Antrajame bloke yra pateiktos modeliuojamos komandos statistinės kategorijos. Trečiajame bloke pateiktos modeliuojamos komandos varžovės statistinės kategorijos. Ketvirtajame bloke – kiti kintamieji, kaip namų aikštės indikatorius, santykis tarp abiejų komandų pergalių procento ir kiti.

Duomenų rinkimo idėja taip pat labai nesudėtinga. Jis remiasi tuo, jog rezultatams nusakyti yra naudojama visa įmanoma statistinė informacija, kuri prieinama iki rungtynių pradžios. Kitaip tariant, jeigu rungtynės vyksta 2016-12-04 dieną, visa statistinė informacija iki 2016-12-03 imtinai yra surinktuose duomenyse. Kiekvieną dieną komandų vidutinės statistinės kategorijos šiek tiek koreguojasi.

Duomenyse yra 229 skirtingi stulpeliai. 1 stulpelis skirtas rezultatui, kiti – statistinėms kategorijoms bei kitoms rungtynių charakteristikoms aprašyti. Pusė kintamųjų iš esmės yra tokie patys, tik nusako skirtingų komandų rodiklius.

Pavyzdžiui – vidutinis stebimos komandos pelnytų taškų kiekis bei vidutinis stebimos komandos varžovės pelnytų taškų kiekis. Ir taip visiems kintamiesiems – vienas rodiklis duomenyse sukuria du skirtingus kintamuosius – pirmasis aprašo modeliuojamos komandos rodiklį, antrasis modeliuojamos komandos varžovės rodiklį.

Į modelį buvo įtrauktas ir tam tikras indikatorius, nurodantis rimtą pasikeitimą komandos struktūroje. Pavyzdžiui traumos, arba tenerio pasikeitimas. Visgi, tokią informaciją surinkti ir patį kintamąjį aprašyti objektyviai yra gana sudėtinga, todėl tik pačių didžiausių komandos žvaigždžių traumos buvo įtrauktos į tyrimą.

Visų kintamųjų aprašymas yra pateiktas priede. Aprašyti yra tik unikalūs statistiniai rodikliai. Modeliuojant rodiklio pavadinime pridedant žodį „var“ yra suprantama, kad tai yra rodiklis ne modeliuojamos komandos, o jos varžovės.

Kadangi darbe bus naudojami įvairūs metodai ir vieni iš jų yra jautrūs skirtingoms duomenų dimensijoms, visi duomenų stulpeliai x buvo transformuoti į stulpelį z remiantis šia taisykle:

$$z_i = \frac{x_i - \bar{x}}{sd(x)}$$

čia $sd(x) = \frac{\Sigma(x_i - \bar{x})^2}{N-1}$, N - duomenų kiekis.

4.2 Kintamųjų skirstymas

Pagal sudarytą teorinį modelį, buvo siekiama jį testuoti iš empirinių duomenų. Visi kintamieji buvo suskirstyti į šešias skirtingas grupes, jog patektų bent į vieną grupę, kurią aprašo latentinis kintamasis. Skirstymas buvo atliktas pasitelkiant autoriaus ekspertinį vertinimą. Jų aprašymas, kaip jau buvo minėta anksčiau, yra prieduose.

Bendras Puolimas	Puolėjų puolimas	Gynėjų puolimas	Gynyba
AveragePTS	OREB	3PA	DREB
FGA	OREB%	3P%	REB
FG%	OppOREB	AST	STL
FTA	%PTSITP	AST%	BLK
FT%	PaintPTS	AST/TO	PF
OffRtg	PaintFG%	TOV%	DefRtg
TS%	PostPTS	PACE	DREB%
OppSTL	PostFG%	OppPTSofTOV	REB%
OppBLK	ContestedOreb	%FGA3PT	OppFG%
%FGA2PT	Orebchances	%PTS3PT	OppFTA Rate
%PTSFT	Orebdistance	3FGM%AST	OppTOV%
2FGM%AST	ScreenAst	Drives	OppOREB%
DFGA	PostTouches	DrivesPTS%	OppPTS
Contested2ptShots	RAFGM	DrivesPass%	OppFG%
DistMilesOF	RAFG%	DrivesTO%	Opp3P%
AvgSpeedOF	PaintFGM	DrivesPTS	OppFTA
MidFGM	PaintFG%	DrivesFG%	OppDREB
MidFG%		CatchShootPTS	OppAST
%PTS2PT		CatchShootFG%	OppTOV
%PTS2PT-MR		PullupPTS	OppPFD
TOV		PullupFG%	OppPTS2ndChance
%PTSFBPs		EllbowPTS	OppPTSFB
%PTSOFFTO		EllbowFG%	OppPTSPaint
		Contested3ptShoots	Deflections
			LooseBallRecover
			ChargesDrawn
			DistMilesDF
			AvgSpeedDF
			ContestedDreb
			Drebchances
			Drebdistance

Bendras Puolimas	Puolėjų puolimas	Gynėjų puolimas	Gynyba
			OPRAFGM
			OPRAFG%
			OPPaintFGM
			OPPaintFG%
			OPMidFGM
			OPMidFG%
			OPL3FGM
			OPL3FG%
			OPR3FG%
			OPR3FGM
			OP3OtherFGM
			OP3OtherFG%

Analogiškai buvo suskirstyti ir varžovų komandos kintamieji, tik jie nebuvo skirstomi taip smulkiai, jų suskirstymas yra į dvi dalis - tiesiog varžovų gynyba ir varžovų puolimas. Beje, tiek varžovų, tiek stebimos komandos gynybai aprašyti yra naudojami tie patys kintamieji, todėl jie pačiame struktūrinių lygčių modeliavime bus fiksuoti kaip vienodi. Statistinio R paketo bibliotekoje laavaan tai yra labai paprasta nurodyti, pačiame modelyje, šalia modeliuojamų kintamųjų įrašant tą patį parametrą šalia kintamųjų skirtingose lygtyse.

4.3 Modelis be varžovų rodiklių, vertinimas

Kintamieji, kurie bus įtraukti aprašyti latentinius kintamuosius buvo parinkti remiantis jų tarpusavio koreliacija bei autoriaus ekspertinėmis įžvalgomis.

Visi modelio kintamieji reikšmingi, ženklai atitinka logiką ir neprieštarauja vienas kitam, nors p-reikšmė Chi-square statistikos yra nereikšminga, tai yra, ji sakytu, jog modelis nėra tinkamas, tačiau taip nutinka labai dažnai, imtimis, didesnėms nei 200 [8]. Chi-square statistika didesnio dydžio imtims modelį atmata kaip netinkamą praktiškai visada ir neretai be priežasties. Kadangi kitos modelio charakteristikos RMSEA (0,12), CFI (0,903) atitinka keliamus kriterijus arba yra arti labai arti nustatytų ribų, toliau tęsime tyrimą. Modelio AIC 82601.

Apžvelkime patį modelį ir kintamuosius, kurie nusako latentinius kintamuosius.

Puolimas – jį nusako penki kintamieji – komandos išmetamų baudų kiekis, vidutinis taškų kiekis per rungtynes, puolimo reitingas, taškų kiekis, pelnytas greitose atakose bei pataikytų metimų kiekis iš vidutinio nuotolio. Parametrų interpretacija paprasta – padidėjus vidutiniam taškų kiekiui vienetu, komandos puolimas padidėja taip pat vienetu. Priminsime, kad to nereikia suprasti tiesiogiai, kadangi visi kintamieji buvo transformuoti.

Gynėjų puolimo kintamasis yra sudarytas iš šių stebinių – rezultatyvių perdavimų ir klaidų santykio, tritaškių kiekio procentais nuo visų išmestų metimų iš žaidimo, pataikytų tritaškių iš kairiojo ir dešiniojo atakos krašto, taškų kiekio kai taškai buvo pelnyti pagavus kamuolį ir jo nesumušus į žemę bei klaidų procento, kuomet yra veržiamasi po krepšiu. Interpretacija tokia pati – pavyzdžiui dviejų komandų visi rodikliai yra identiški, tačiau vienetu skiriasi klaidų procentas besiveržiant po krepšiu. Tokių komandų latentinis gynėjų puolimo rodiklis skirsis per 0,407.

Puolėjų puolimą sudaro taškai pelnyti baudos aikštelėje, kamuolio lietimai ir pelnyti taškai apribotoje zonoje, rezultatyvios sienelės, atkovotų kamuolių puolime procentas bei atkovotų kamuolių puolime vidutinis atstumas. Pastarasis rodiklis turi neigiamą ženklą ir yra logiškas. Kuo arčiau yra atkovojami kamuoliai, tuo tai duoda komandai pranašumo pelnyti taškus iškart (apribotoje zonoje esantį žaidėją sustabdyti yra labai sunku), be to, kuo daugiau kamuolių atkovojama toli nuo krepšio, tuo tai daugiau galima paaiškinti kaip atsitiktinumu nei komandos žaidėjų gebėjimu gerai grumtis dėl kamuolių puolime. Kamuolys tiesiog „įkrito“ žaidėjui į rankas, o nebuvo atkovotas. Koeficientas prie rezultatyvių sienelių yra pats didžiausias. Iš tiesų, tai galėtų būti indikatorius, ne tik pačio fakto, jog po geros sienelės buvo įmesti taškai, tačiau neretai po gerų sienelių aukštaūgiai gauna kamuolį iš įžaidėjų ir įmeta lengvus taškus. Bendrai, komandos puolimui stiprios sienelės suteikia labai daug pranašumų.

Paskutinis latentinis faktorius yra gynyba. Ji yra aprašoma šiais rodikliais: kiek vidutiniškai per rungtynes oponentai įmeta į stebimos komandos krepšį, vidutinis pražangų kiekis, varžovų išmestų baudų kiekis, gynybos reitingas bei blokų kiekis. Visų kintamųjų ženklai taip pat neprieštarauja vienas kitam. Visi ženklai su teigiamu ženklu nusako, jog kuo rodiklis yra didesnis, tuo komandos gynyba yra prastesnė. Tiksliau tariant - kuo daugiau praleidžia taškų, kuo daugiau varžovai meta baudų, kuo daugiau kartų yra pažeidžiamos taisyklės ir kuo didesnis gynybos reitingas (didesnis gynybos reitingas rodo prastesnę gynybą,

o mažesnis gera) tuo komandos gynyba yra prastesnė. Blokai turi priešingą ženklą – kuo daugiau blokų – tuo gynyba geresnė ir ženklas yra priešingas iki tol buvusiems.

Regresijos koeficientai, nusakantys pelnytų taškų kiekį yra statistiškai reikšmingi bei jų ženklai neprieštarauja logikai. Didžiausią koeficientą turi gynėjų puolimas, tačiau tai dar nereiškia, kad gynėjų puolimas yra pagrindinis faktorius, nulemiantis pelnytų taškų kiekį. Kiek vėliau pažvelgsime į pačių kintamųjų dydžius ir nustatysime, kokį svorį kiekvienas kintamasis turi šioje struktūroje (pavyzdžiui gynėjų puolimo latentinio kintamojo reikšmės yra tiesiog ženkliai mažesnės už bendro puolimo ir puolėjo puolimo reikšmes, tad didesnis koeficientas savaime nereiškia, kad gynėjų puolimas yra svarbiausias). Žaidimas namie ir komandų pajėgumo skirtumas taip pat buvo įtraukti į regresiją. Jų ženklai ir reikšmingumas taip pat atitinka keliamus kriterijus. Žemiau esančiose lentelėse yra pateikta modelio santrauka.

Latentintis kintamasis: BPuolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
FTA	1		
AveragePTS	2,9	0,12	<0,05
OffRtg	1,6	0,08	<0,05
%PTSFBP	1,4	0,07	<0,05
MidFGM	0,4	0,04	<0,05
Latentintis kintamasis: GPuolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
AST/TO	1		
%FGA3PT	0,7	0,03	<0,05
L3FGM	0,9	0,04	<0,05
R3FGM	1,1	0,04	<0,05
DrivesTO%	-0,4	0,03	<0,05

Latentintis kintamasis: PPUolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
PaintPTS	1		
PostTouches	0,9	0,01	<0,05
PostPTS	1,0	0,02	<0,05
ScreenAst	1,4	0,12	<0,05
OREBPRC%	0,05	0,01	<0,05
Orebdistance	-0,08	0,01	<0,05
Latentintis kintamasis: Gyn			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
OppPTS	1		
OppPFD	0,5	0,03	<0,05
OppFTA	0,5	0,02	<0,05
DefRtg	0,9	0,2	<0,05
BLK	-0,2	0,04	<0,05
Regresija: Pelnyti taškai			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
Intercept	-0,9	0,3	<0,05
BPuolimas	0,04	0,02	<0,05
PPuolimas	0,2	0,03	<0,05
GPuolimas	0,13	0,01	<0,05
Home	0,29	0,05	<0,05
W	0,25	0,04	<0,05

4.4 Modelis su varžovų rodikliais, vertinimas

Pirminis sudarytas modelis buvo koreguojamas pridėdant stebimus varžovų komandų statistinius rodiklius bei iš jų sukonstruojant naujus latentinius kintamuosius – varžovų gynybą bei varžovų puolimą. Modelio RMSEA (0,09), CFI (0,912) atitinka keliamus kriterijus. Modelio AIC 73201 yra mažesnis nei pastarojo modelio. Varžovų gynyba buvo įtraukta į regresiją, tuo tarpu varžovų puolimas į regresiją nėra įtraukiamas, tačiau tariama, kad yra koreliuotas su varžovų gynyba.

Modelyje yra apriboti tam tikri parametrai. Tiek stebimos komandos gynyba, tiek varžovų komandos gynyba yra aprašomi tais pačiais kintamaisiais.

Modelio loginė idėja sako, jog patys latentiniai kintamieji turėtų sutapti, jeigu juos aprašantys kintamieji sutampa, dėl to koeficientai prie stebimos komandos gynybos rodiklių bei varžovų komandos gynybos rodiklių sutampa (modelyje aprašomas apribojimas).

Modelio santrauka pateikta lentelėje. Regresijoje prisidėjęs varžovų gynybos rodiklis turi logišką ženklą pagal jau anksčiau aprašytą jo konstravimo struktūrą. Kuo didesnis gynybos latentinis kintamasis – tuo komandos gynyba yra blogesnė. Kuo varžovų gynyba blogesnė, tuo stebima komanda turėtų įmesti daugiau taškų.

Latentintis kintamasis: BPuolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
FTA	1		
AveragePTS	2,8	0,12	<0,05
OffRtg	1,5	0,08	<0,05
%PTSFBP	1,4	0,07	<0,05
MidFGM	0,3	0,04	<0,05
Latentintis kintamasis: GPuolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
AST/TO	1		
%FGA3PT	0,7	0,03	<0,05
L3FGM	0,8	0,04	<0,05
R3FGM	1,1	0,04	<0,05
DrivesTO%	-0,4	0,03	<0,05
Latentintis kintamasis: PPUolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
PaintPTS	1		
PostTouches	0,9	0,01	<0,05
PostPTS	0,9	0,02	<0,05
ScreenAst	1,4	0,12	<0,05
OREBPRC%	0,05	0,01	<0,05
Orebdistance	-0,08	0,01	<0,05

Latentintiai kintamieji: Gyn ir VarGyn			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
OppPTS	1		
OppPFD	0,5	0,03	<0,05
OppFTA	0,5	0,02	<0,05
DefRtg	1,2	0,2	<0,05
BLK	-0,2	0,04	<0,05
Latentintiai kintamieji: VarPuolimas			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
VarFTA	1		
VarAST/TO	0,8	0,03	<0,05
VarPaintPTS	0,7	0,04	<0,05
VarAveragePTS	1,8	0,04	<0,05
VarOffRTG	1,4	0,03	<0,05
Var%FGA3P	0,6	0,05	<0,05
Regresija: Pelnyti taškai			
Kintamasis	Įvertis	St. paklaida	p-reikšmė
Intercept	-0,8	0,2	<0,05
BPuolimas	0,08	0,02	<0,05
PPuolimas	0,02	0,03	<0,05
GPuolimas	0,12	0,02	<0,05
VGyn	0,26	0,02	<0,05
Home	0,29	0,05	<0,05
W	0,26	0,04	<0,05

4.5 Latentinių kintamųjų vertinimas komandoms

Statistiniame R pakete, lavaan bibliotekoje yra labai patogi funkcija *lavPredict* kuri įvertina latentinių kintamųjų dydį pagal sukonstruotą modelį ir pagal atitinkamas į modelį įtrauktų kintamųjų reikšmes.

Įvertinti latentiniai kintamieji komandoms yra pateikti priede, 25-28 paveikslėliuose. Matyti, kad įvertinti puolimo ir gynybos latentiniai kintamieji yra logiški. Apibendrinus visus kintamuosius pastebėtina, kad ekspertų geriausiomis puolimo komandomis laikomos Golden State Warriors, Boston Celtics išsiskiria gynėjų ir puolėjų kategorijose. Tuo tarpu pačios silpniausios lygos komandos

bei komandos, kurių gynyba iš tiesų nežiba net plika akimi žvelgiant, Lakers, Nuggets, Nets, Suns, jų gynybos latentintis kintamasis buvo įvertintas kaip pats didžiausias (kas nusako silpną gynybą). Spurs, Jazz ir Warriors įvertintos kaip geriausios gynybos komandos. Tai atrodo labai logiškai. Warriors labai agresyvi gynybos komanda ir nors ji kartais praleidžia gana daug taškų, tačiau tai yra ne dėl jų prastos gynybos, o dėl demonstruojamo žaidimo tempo. Labai agresyvi gynyba leidžia perimti daug kamuolių, varžovus išprovokuoti mesti neparuoštus metimus. Tuo tarpu Spurs ir Jazz 2016-2017 metų NBA sezone buvo elitinės gynybos komandos. Jų gynyba stipri ne tik komandiškai, tačiau ir kartu turi žaidėjų, kurie yra vadinami pačiais geriausiais visoje lygoje (Spurs atveju Kawhi Leonard yra vienas geriausių perimetro gynėjų bei kartu yra labai stiprus, jog gali keistis su aukštaūgiais bei ginti sunkiuosius kraštus ar net vidurio puolėjus, Jazz atveju Rudy Gobert yra tikra gynybos siena. Prancūzas yra geriausias blokuotų metimų specialistas, jo ilgos rankos net ir neblokavus metimų, sukelia varžovams labai daug problemų norint įmesti į krepšį iš arčiau).

Grafiškai, dar geriau situaciją apibūdina 13 paveikslėlis. Jame horizonta-lioje ašyje yra įvertintas bendras vidutinis komandų puolimas per sezoną. Jo vertinimas labai paprastas – svertinė tiesinė latentinių kintamųjų (bendro puo-limo, puolėjų puolimo ir gynėjų puolimo) kombinacija. Svoriai prie kintamųjų yra lygūs koeficientams regresijos lygtyje. Vertikaloje ašyje – įvertinta vidutinė latentinio gynybos kintamojo reikšmė per sezoną.

Latentintis gynybos kintamasis buvo konstruotas taip, jog didesnė jo reikšmė reiškia prastesnę komandos gynybą, o latentiniai puolimo kintamieji atvirkščiai – kuo didesnė reikšmė, tuo stipresnis puolimas. Taigi, geriausią puolimą ir ge-riausią gynybą turinčios komandos 13 paveikslėlyje yra dešiniajame apatiniame kampe. Nenuostabu, jog dvi geriausios komandos – San Antonijaus Spurs bei Golden State Warriors. Šios dvi komandos per sezoną demonstravo patį geriau-sią krepšinių lygoje bei susitiko NBA Vakarų konferencijos finale ir ten Warriors serijoje iki keturių pergalių šventė pergalę 4-0. Visgi, tokiam lengvai Warriors pergalei įtaką padarė pagrindinio Spurs žaidėjo Kawhi Leonard patirta trauma pirmosiose serijos rungtynėse. Iki tol, kol K.Leonard nebuvo traumotas, Spurs turėjo dviženklę taškų persvarą.

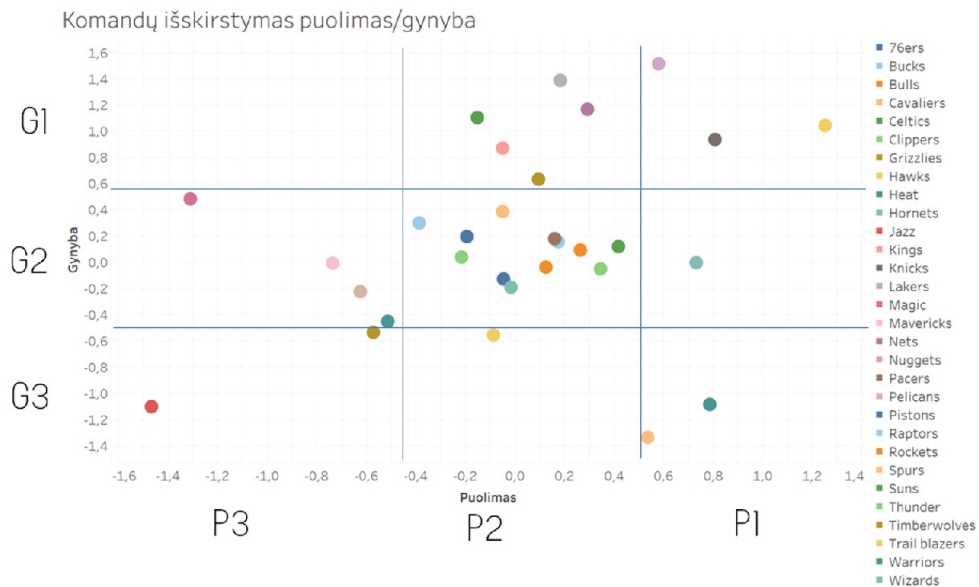
Blogiausios lygos komandos, su prasta gynyba ir prastu puolimu patektų į kairįjį viršutinį 13 pav. stačiakampį, tiesa nė viena iš komandų tokių rodik-lių nedemonstruoja. Orlando Magic yra viena iš komandų buvusi arčiausiai

jos. Tiesa, yra aštuonios komandos, kurių gynyba yra prastesnė nei Magic. Kas dar išsiskiria – tai Jutos Jazz klubas, esantis apatiniame kairiajame kampe. Šioje zonoje esančios komandos turi labai gerą gynybą, tačiau nepasižymi aukštu puolimo įvertinimu. Logiškas atitikimas – Jazz laikoma viena geriausių besiginančių komandų, bet puolimo sferoje nepasižymi įvairiapusiškumu, geru pataikymu ar labai stipriu komandiniu žaidimu.

Įdomu tai, jog NBA finalininkė, Klivlendo Cavaliers komanda atliktame komandų įvertinime nepasižymėjo nei geru puolimu, nei itin stipria gynyba. Tam yra logiškų paaiškinimų. Cavaliers, neabejotinai stipriausia Rytų konferencijos komanda, tačiau reguliariojo sezono metu komanda nerodydavo viso savo potencialo. Ne itin pajėgioje Rytų konferencijoje Cavaliers sugebėjo užimti ant-rąją poziciją. Be to, ji nebuvo iškovota taip paprastai, pačioje sezono finišo tiesiojoje buvo aplenkta Toronto Raptors komanda. Tuo tarpu atkrintamosiose varžybose, iki susitikimo NBA finale su Warriors, Cavaliers patyrė vos vieną pralaimėjimą. Lyderių taupymas, leidimas kai kurioms komandoms žvaigždėms pailsėti kai kuriose, ne itin svarbiose rungtynėse, taip pat keletas žaidėjų traumos – visa tai atsispindi sudarytame grafike. Žodžius patvirtinti faktais – pastebėjime, jog LeBron James reguliariajame NBA sezono (darbe tiriamos būtent reguliariojo sezono rungtynės, atkrintamųjų varžybų mačai į tyrimo laikotarpį nepatenka) vidutiniškai aikštėje praleisdavo po 37 minutes, kai tuo tarpu atkrintamųjų varžybų metu šis rodiklis išaugo daugiau nei 10%, iki beveik 42 minučių vidutiniškai per rungtynes. Taigi, faktas, kad Cavaliers reguliariojo sezono metu nedemonstravo viso savo potencialo, atsispindi gautuose rezultatuose.

Visos komandos padalinamos į devynias skirtingas grupes, tiesa, į vieną iš grupių nepateko nei viena komanda, tai galiausiai liko aštuonios grupės. Ateityje šis padalinimas bus skirtas tirti, kaip tiksliai modelis prognozuoja skirtingo lygio komandų pasirodymus.

- P1&G1 laukas – komandos, kurios pasižymi geru puolimu, tačiau jų gynyba yra tarp prasčiausių lygoje.
- P1&G2 laukas – komandos turinčios gerą puolimą, o gynyba yra vidutinio pajėgumo.
- P1&G3 laukas – komandos turinčios gerą puolimą ir gerą gynybą.
- P2&G1 laukas – komandos turinčios vidutinį puolimą bei prastą gynybą.



13 pav.

P1&G1	P1&G2	P1&G3	P2&G1	P2&G2	P2&G3	P3&G2	P3&G3
Knicks	Wizards	Spurs	Kings	76ers	Hawks	Heat	Grizzlies
Blazers		Warriors	Lakers	Bucks		Magic	Jazz
Nuggets			Nets	Bulls		Mavs	
			Suns	Cavs		Pelicans	
			TWolves	Celtics			
				Clippers			
				Hornets			
				Pacers			
				Pistons			
				Raptors			
				Rockets			
				Thunder			

14 pav.

- P2&G2 laukas – komandos turinčios vidutinį puolimą bei vidutinę gynybą.
- P2&G3 laukas – komandos turinčios vidutinį puolimą bei gerą gynybą.
- P3&G2 laukas – komandos turinčios prastą puolimą bei vidutinę gynybą.
- P3&G3 laukas – komandos turinčios prastą puolimą bei gerą gynybą.

Suskirstymas į laukus atskleidžia įprastą faktą, jog vidutinišką populiacijos savybę turinčių individų kiekis visuomet didžiausias. Analogiškai ir šiame paskirstyme – komandų su vidutiniu puolimu bei vidutine gynyba yra 12, kas sudaro 40% visų lygos komandų. Tuo tarpu išsiskiriančių – su geru puolimu ir gera gynyba, su blogu puolimu ir gera gynyba – vos po dvi. Kaip jau minėjome,

komandų, kurios išsiskirtų tiek savo blogu puolimu, tiek savo prasta gynyba nėra. Vienintelė Atlantos Hawks pateko į lauką P2&G3, kuris nusako komandas su vidutiniu puolimu bei gera gynyba. Norint patekti į elitinį P1&G3 lauką šiai komandai reiktų pagerinti savo puolimo rodiklius.

4.6 Dirbtinių neuroninių tinklų modeliai

Kaip buvo galima pastebėti nagrinėjant kitų autorių darbus, dirbtinių neuroninių tinklų modeliai buvo vienas iš populiariausių metodų, naudojamas sporto rungtynių baigtims prognozuoti. Šiame skyriuje, turimiems duomenims sieksime sudaryti neuroninių tinklų modelį.

Statistiniame R pakete neuroninių tinklų modeliams konstruoti yra sukurta labai patogi neuralnet biblioteka, o pats modelis yra įvertinamas naudojantis neuralnet bibliotekoje esančia tokio pačio pavadinimo funkcija. Čia numatytoji aktyvacijos funkcija yra sigmoidinė. Ją nesunkiai galima koreguoti neuralnet funkcijoje įrašant papildomą parametą, pavyzdžiui `act.fct= tanh`. Mūsų tyrime bus naudojama sigmoidinė aktyvacijos funkcija.

Surasti geriausiai duomenims aprašyti tinkamą neuroninių tinklų modelį buvo naudojami įvairūs statistiniai ir ekonometriniai metodai. Visų pirma, buvo atrinkti kintamieji, kurie bus įtraukiami į neuroninių tinklų modelį. Tam tikslui buvo sudaromas laipsniškosios regresijos pažingsnis parinkimas (angliškai *stepwise regression forward selection*). Kintamieji į regresiją atrenkami pažingsniui – pradžioje stebima ar kuris nors iš galimų kintamųjų pagerina modelį lyginant su modeliu be kintamųjų. Jeigu taip, ieškoma ar iš likusių kintamųjų egzistuoja toks, kuris padeda pagerinti iki tol geriausią modelį. Procedūra kartojama tol, kol neegzistuoja kintamasis, kad jis pagerintų iki tol buvusį geriausią modelį. Gerumo vertinimo matu buvo naudojamas AIC kriterijus.

Kintamieji, kurie buvo atrinkti į modelį pateikti žemiau. Iš viso, iš 230 kintamųjų, buvo atrinkti 40 geriausių. Su jais toliau buvo konstruojamas dirbtinių neuroninių tinklų modelis. Jie yra išrikiuoti eilės tvarka, pagal svarbumą. Į modelį buvo įtraukta nemažiau kaip keturi kintamieji, kadangi rezultatai su mažiau kintamųjų buvo pastebimai prastesni. Pirmieji keturi kintamieji buvo įtraukti į visus kitus modelius. Pirmieji penki kintamieji buvo įtraukti į visus modelius, išskyrus į modelius, su keturiais kintamaisiais ir t.t. Geriausio modelio parinkimo strategijos aprašymas yra pateiktas žemiau.

Kintamieji: 1.VarOppPTS, 2.AveragePTS, 3.VarPaintFGPRC, 4.Namai, 5.VarF-

GA, 6.AvgSpeedOf, 7VarL3FGPRC, 8.VarContested3ptShoots, 9.FGPRC, 10.LooseBallRecover, 11.OPL3FGM, 12.OppAST, 13.PaintFGM, 14.EllbowFGPRC, 15.VarBLK, 16.VarOppOREB, 17.OppPTS2ndChance, 18.Deflections, 19.VarPullupFGPRC, 20.VarDrivesTOPRC, 21.OPRFGM, 22.PRCPTSFBPs, 23.OppPTSPaint, 24.X3FGMPRCast, 25.ContestedDreb, 26.VarOP3OTHERFGM, 27.DrivesTOPRC, 28.VarPostFGPRC, 29.VarDefRtg, 30.VarPACE, 31.VarDrebchances, 32.VarOppPTS2ndChance, 33.VarPostTouches, 34.VarOPL3FGPRC, 35.OREB, 36. OppPTSFB, 37. VarRFGPRC, 38.TSPRC, 39. OppDREB, 40. OP3OTHERFGM.

Kaip ir buvo galima tikėtis, regresija pagrindinius du kintamuosius išrinko tai, kiek stebimos komandos priešininkų varžovai vidutiniškai pelno taškų. Šis rodiklis gerai atspindi tai, kokio lygio yra priešininkų gynyba. Antras svarbiausias kintamasis – stebimos komandos vidutinis pelnomų taškų kiekis. Kiti svarbūs faktoriai kaip varžovų pataikymo procentas bei pataikymo procentas baudos aikštelėje, namų aikštės pranašumas, vidutinis puolimo greitis, prarastų kamuolių susigrąžinimo kiekis ir t.t.

Nustatyti patį geriausią dirbtinių neuroninių tinklų modelį buvo naudojama 990 rezultatų imtis. 600 iš jų atsitiktinai buvo įtraukiami į mokymosi aibę, likę 390 buvo paliekami testuoti modelio tikslumui. Buvo ieškomas geriausias modelis, prognozuojantis rungtynių rezultatą, priklausomai nuo įtrauktų kintamųjų kiekio, bei skirtingo neuronų kiekio slaptame modelio sluoksnyje. Buvo naudojamas vienas slaptas sluoksnius modelyje. Kiekvienam kintamųjų rinkiniui buvo konstruojamas modelis su skirtingu kiekiu neuronų. Tai daroma 50 kartų. Kiekvienu iš atvejų skaičiuojama vidutinė kvadratinė testinės aibės rezultatų paklaida ir ji suvidurkinama. Rezultatai pateikti 15 pav.

Matant tendencijas, jog didesnis neuronų kiekis viename slaptajame sluoksnyje duoda prastesnius prognozės rezultatus bei didesnis kintamųjų kiekis nei 10 taip pat rezultatų nepagerina, lentelėje pateikti tik esminiai rezultatai. Geriausius rezultatus demonstravo modeliai su 8 bei 9 kintamaisiais, kurie turėjo du neuronus. Visiems kintamųjų kiekiams, neuronų kiekis didesnis nei du, davė prastesnius rezultatus. Kitų modelių, kurie nepateikti lentelėje, tikslumo rezultatai yra ženkliai prastesni nei pateiktų 15 paveikslėlyje, todėl detalčiau čia jų neaptarinėjame.

Kintamieji/ Neuronai	1	2	3	4	5	6
4	0,9177	0,8579	0,8963	0,8830	0,9148	0,9238
5	0,8997	0,8426	0,8649	0,8844	0,9175	0,9548
6	0,9121	0,8380	0,8681	0,9010	0,9348	0,9815
7	0,9137	0,8381	0,8858	0,9277	0,9617	1,0125
8	0,8909	0,8395	0,8945	0,9556	1,1109	1,5809
9	0,8887	0,8370	0,8979	0,9554	1,0014	1,0818
10	0,8949	0,8782	0,9394	0,9736	1,0328	1,0788
15	0,9243	0,9266	0,9739	1,0466	1,1144	1,1871
20	0,9174	0,9271	1,0362	1,1314	1,2021	1,3083

15 pav.

4.7 Neuroninių tinklų ir SEM palyginimas - latentinių kintamųjų ir neuronų sąryšis

Iš esmės sudaryti neuroninių tinklų modeliai bei struktūrinių lygčių modeliai turi labai panašią struktūrą. Bent grafiniame atvaizdavime struktūra atrodo labai analogiška. Stebime pasirinktus kintamuosius, iš kurių, vėliau, gaunami tam tikri paprastais žodžiais ne itin lengvai paaiškinami dydžiai. Struktūrinių lygčių modelių atveju egzistuoja latentiniai kintamieji, kurių prasmė yra aiški, nes ji buvo nurodyta dar prieš sudarant modelį. Tuo tarpu neuroninių tinklų atveju, egzistuoja neuronai slaptajame sluoksnyje, tačiau ką jie reiškia, kokius sąryšius apibūdina, aišku nėra.

Prieš tyrimą, buvo iškelta hipotezė, kad egzistuoja kažkoks stiprus ryšys tarp neuroninių tinklų modelių bei struktūrinių lygčių modelių, tai yra dirbtinių neuroninių tinklų modeliuose esantys neuronai turi sąsajų su SEM modelyje esančiais latentiniais kintamaisiais ir tai netgi galima surasti egzistuojančio sąryšio tam tikrą dėsninę.

Tyrimui naudosime 532 duomenų eilutes ir du pačius geriausius modelius – SEM su stebimos komandos bendru puolimu, puolėjų puolimu, gynėjų puolimu, bei varžovų gynyba. Tuo tarpu neuroninių tinklų modelį parinksime modelį, kuris rodė geriausius rezultatus prognozuojant testavimo aibę – su devyniais kintamaisiais bei dviem neuronais slaptajame sluoksnyje. Iš pažiūros, modeliai yra ganėtinai skirtingi – SEM modelyje yra penki struktūriniai vienetai, kuriuos tirsime, tuo tarpu dirbtinių neuroninių tinklų – du. Skirtumų išvelgti galime ir

daugiau – SEM modelį sudaro 28 kintamieji, tuo tarpu neuroninių tinklų – tik devyni.

Visgi, prielaida daroma tokia, jog tie du neuronai savyje surenka visą reikiamą informaciją, todėl egzistuoja ryšys tarp įvertintų struktūrų. Be to, nors kintamųjų kiekis modeliuose yra ženkliai skirtingas, tikėtina, kad devyni kintamieji taip pat savyje neša didesnę dalį visų duomenų informacijos, o kita dalis kintamųjų, suteikia ne tiek daug naujų įžvalgų.

Pačioje pradžioje, sudarysime paprastą tiesinę regresiją, tam, jog pažvelgtume ar tiesinė latentinių kintamųjų kombinacija aprašo dirbtinių neuroninių tinklų modelio neuronus. Gautų modelių rezultatai pateikti 16 bei 17 paveikėliuose.

Regresija: Pirmas neuronas			
Kintamasis	Įvertis	St. Paklaida	P reikšmė
Intercept	0,46	0,002	<0,05
Puolimas	-0,04	0.02	<0,05
Gpuolimas	-0,004	0.01	<0,05
Gyn	-0,01	0.03	<0,05
VGyn	0,02	0.05	<0,05

16 pav.

Regresija: Antras neuronas			
Kintamasis	Įvertis	St. Paklaida	P reikšmė
Intercept	170	3,7	<0,05
Puolimas	-63	7,5	<0,05
Gpuolimas	19	2,3	<0,05
PPuolimas	5	1,8	<0,05
Gyn	17	4,6	<0,05
VGyn	-10	4,1	<0,05

17 pav.

Matyti, jog labai stipraus ryšio tarp latentinių kintamųjų bei neuronų slaptuose neuroninių tinklų sluoksnuose nėra. Koreguotasis R-kvadratas pirmajam neuronui siekia vos 0,31 (regresijoje kintamieji yra reikšmingi). Iš esmės ženkliai visų kintamųjų yra logiški. Visi puolimą aprašantys kintamieji turi neigiamą ženklą, o varžovų gynyba teigiamą (tai yra visų kintamųjų kryptis yra ta pati).

Galima daryti interpretaciją, jog kuo stipresnis komandos puolimas – tuo mažesnė pirmojo neurono reikšmė slaptajame neuroninių tinklų modelio sluoksnyje. Logišką ženklą prideda ir varžovų gynyba. Priminsime, jog latentintis kintamasis buvo sukonstruotas taip – jog kuo didesnė jo reikšmė, tuo prastesnė varžovų gynyba. Taigi, kuo geresnė varžovų gynyba – tuo mažesnę reikšmę įgaus neuronas.

Tolesnei analizei nagrinėsime tik pirmo neurono ir latentinių kintamųjų sąryšį. Iš pirmo žvilgsnio, galima būtų daryti prielaidą, kad pakankamai mažą R-kvadratą nulemia tai, jog neuronuose esanti informacija yra apdorojama netiesiniais metodais - mūsų atveju sigmoidine funkcija. Galima būtų daryti prielaidą, kad tam tikroje aplinkoje, kur sigmoidinė funkcija elgiasi gana tiesiškai, sudarytos regresijos R-kvadratas galėtų būti didesnis.

Visgi, atlikus tokį eksperimentą, rezultatai yra priešingi nei buvo galima tikėtis. Didžiausias tiesinis sąryšis egzistuoja tada, kai regresija yra sudaroma pačioms didžiausiomis ir pačioms mažiausiomis neuronų reikšmėms. Lentelėje pateikta informacija apie koreguotąjį R-kvadratą skirtingiems duomenų poaibiams. Pavyzdžiui poaibis 50 didžiausių ir mažiausių reikšmių reiškia, jog buvo modeliuoti 50 didžiausių pirmojo neurono reikšmių ir 50 mažiausių pirmojo neuronų reikšmių su atitinkamais latentiniais kintamaisiais.

Iš lentelės galima susidaryti aiškų vaizdą, jog latentiniai kintamieji gana tiksliai gali tiesiškai aprašyti pirmajame neurone esančią informaciją, jeigu tiriamo tik pačias didžiausias ir pačias mažiausias reikšmes. Tuo tarpu rezultatai, kuomet tiriamo tik vidurines neuronų reikšmes, pavyzdžiui atmetus 150 didžiausių ir 150 mažiausių reikšmių, koreguotasis R-kvadratas yra vos 0,06.

Poaibis	Koreguotasis R^2
Visi duomenys	0,31
50 didžiausių ir mažiausių reikšmių	0,66
100 didžiausių ir mažiausių reikšmių	0,51
150 didžiausių ir mažiausių reikšmių	0,46
Be 50 didžiausių ir mažiausių reikšmių	0,16
Be 100 didžiausių ir mažiausių reikšmių	0,10
Be 150 didžiausių ir mažiausių reikšmių	0,06

Galima daryti išvadą, jog pirmasis neuronas ir latentiniai kintamieji turi šią bendrą. Kuomet pirmojo neurono reikšmės yra didelės arba mažos, tuomet

sąryšis egzistuoja kur kas stipresnis nei visiems duomenims. Viso to interpretacija galėtų būti tokia: tiek pirmasis neuronas, tiek tiesinė latentinių kintamųjų kombinacija gana vienodai aprašo informaciją, kai susiduria dvi labai skirtingo lygio komandos - stiprų puolimą turinti komanda prieš silpną gynybą bei silpną puolimą turinti komanda prieš stiprią gynybą.

4.8 Prognozavimas

Kaip jau minėjome, visus turimus duomenis prieš tyrimą buvome padalinę į dvi dalis. Iš viso tyrime yra 1522 rezultatų eilutės. 990 eilutės skirtos modeliavimui, likusios 532 testavimui. Šių testavimo duomenų modeliai iki pat pabaigos nebuvo matę.

Iš 990 eilučių, taikydami anksčiau aprašytą metodą, išrinkome geriausiai duomenims tinkantį dirbtinių neuroninių tinklų modelį. Struktūrinių lygčių modelis buvo konstruotas su visomis 990 duomenų eilučių ir atrinktas pagal AIC kriterijų. Kitas žingsnis - patikrinti, kaip gerai modeliai veikia duomenims, esantiems už modeliavimo ribų.

Geriausias neuroninių tinklų modelis turi du neuronus ir devynis kintamuosius. Jo vidutinė kvadratinė paklaida galutinio testavimo duomenimis, kuriuos sudaro 532 eilučių yra - 0,83. Tuo tarpu struktūrinių lygčių modelio, su puolimo bei gynybos latentiniais kintamaisiais – vidutinė standartinė paklaida siekia 0,95. Modelio, be gynybos latentinio kintamojo, 0,99.

Priedo 29 paveikslėlyje pateiktas komandų sąrašas, kurių rezultatai buvo prognozuojami prasčiausiai. Jose pateiktos vidutinių kvadratinių paklaidų reikšmės testinei imčiai. Stebima ne tik kurios komandos įmestų taškų rezultatas yra labiausiai nukrypęs nuo tikrosios reikšmės, bet taip pat kurios besiginančios komandos varžovų rezultatas labiausiai nukrypęs nuo tikrosios reikšmės. Gali būti, kad kai kurių komandų gynybos sugebėjimai yra pervertinami arba nuvertinami. Analogiškai ir su puolimu. Galbūt kai kurių komandų puolimo talentas yra neįvertinamas tiek, kiek jis vertas.

Matyti, jog dirbtinių neuroninių tinklų modelis itin prastai prognozavo Warriors komandos tikslumą, prastai įvertino komandų, kurios žaidė prieš 76ers ir Timberwolves rezultatus. Tuo tarpu SEM modelis prasčiau prognozavo Rockets pasirodymus bei komandų taškų kiekį, kuomet jų varžovė buvo Lakers komanda. Ypač gerai dirbtinių neuroninių tinklų modelis prognozavo Celtics bei Wizards rezultatus.

18 ir 19 paveikslėliuose pateikta vidutinių kvadratinių paklaidų neuroninių tinklų rezultatai testinei imčiai, pagal suranguotus komandų puolimus ir gynybas. Rangai buvo pateikti kiek anksčiau lentelėje. Pastebėtina, kad neuroninių tinklų modelis itin tiksliai spėja rungtynių rezultatus, kuriose dalyvauja P1&G1 lauko komandos. Nesvarbu, ar jų pačių, ar komandų, prieš kurias ginasi šio lauko komandos, rezultatai yra tiksliausi tarp visų grupių (neskaitnat P1&G2 lauko puolančiai komandai, tačiau tame lauke yra tik viena komanda).

	Puolanti komanda	Besiginanti komanda
P1&G1	0,6154	0,6954
P1&G2	0,3876	0,7726
P1&G3	1,2614	0,8060
P2&G1	0,6965	1,0126
P2&G2	0,8483	0,8758
P2&G3	0,7767	0,7746
P3&G2	1,0953	0,7635
P3&G3	0,7582	0,6153

18 pav.

Puolanti komanda	Besiginanti komanda	
P1	G1	0,8528
	G2	0,7424
	G3	0,8614
P2	G1	0,8826
	G2	0,7917
	G3	0,7315
P3	G1	0,9586
	G2	1,1031
	G3	0,5456

19 pav.

Svarbus pastebėjimas iš 19 paveikslėlio, jog neuroninių tinklų prognozavimo tikslumas, komandoms turinčioms patį prasčiausią puolimą (P3), kai jos rungtyniauja prieš komandas, turinčias pačią geriausią gynybą G3, yra labai geras. Tiesa, vidutinė kvadratinė prognozės paklaida yra suskaičiuota tik iš 15 rungtynių, nes tik tiek rungtynių su tokiais laukų deriniais pateko į testinę aibę.

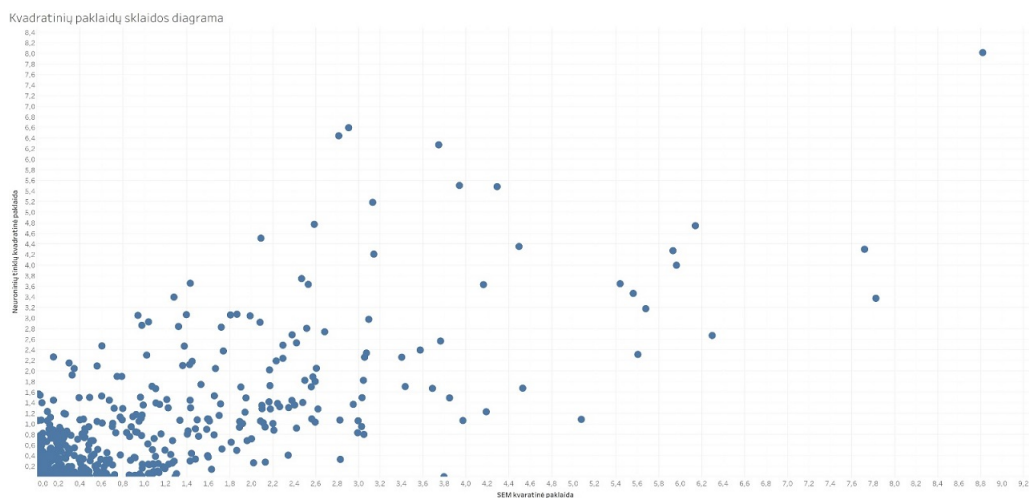
Kuomet stebimos komandos varžovė yra iš pačio stipriausio gynybos lauko,

tuomet prognozės tikslumas yra labai geras. Tokių rungtynių imtyje buvo mažiausiai iš visų trijų laukų (trečiajame stipriausiame lauke yra 17% rungtynių, antrajame gynybos lauke - 57%, o pirmajame, silpniausiame, 25%), tačiau pakankamai daug, jo daryti patikimas išvadas. Šių rungtynių vidutinė kvadratinė paklaida buvo 0,72, kai tuo tarpu iš vidutiniškai besiginančio lauko komandų 0,84, o iš pačio silpniausio gynybos lauko – 0,89.

4.9 Paklaidų klasifikavimas

Šiame skyriuje nagrinėsime klasifikavimo uždavinį, kurio tikslas nustatyti veiksnius, kuomet anksčiau sudaryti modeliai rungtynių baigtis prognozuoja blogiau nei įprasta. 20 paveikslėlyje yra pateikta kvadratinė paklaidų grafikas. Vertikalioje ašyje geriausiai duomenims tikusio neuroninių tinklų modelio kvadratinės paklaidos, o horizantaloje - SEM modelio su varžovų rodikliais kvadratinės paklaidos. Visos rungtynės yra dalinamos į dvi dalis: 1 dalis - paklaidos dydis viršija modelio paklaidų vidurkį, 2 dalis - neviršija modelio paklaidų vidurkio. Rungtynės, kurias abu modeliai prognozuoja blogiau nei vidutiniškai žymime 0, o rungtynes, kurių kvadratinė paklaida yra mažesnė abiem modeliams nei vidutiniškai - 1. Priminsime, jog neuroninių tinklų atveju vidutinė kvadratinė paklaida siekė 0,83, o SEM - 0,95.

Sukoduotoms rungtynėms sieksime sudaryti dvinarės logistinės regresijos modelį, kuris leistų išsiaiškinti rodiklius, kurie padėtų nusakyti tikėtinumą, kada abu modeliai rungtynių taškų kiekį prognozuos tiksliai, o kada netiksliai.



20 pav.

Į logistinės regresijos modelį buvo įtraukti visi surinkti duomenys. Geriausiai duomenims tinkantys faktoriai buvo įtraukiami pasinaudojant pažingsninę

regresiją - iš pradžių modelį sudaro visi kintamieji, vėliau jų kiekis mažinamas. Algoritmas veikia iki tol, kol AIC kriterijus modeliui negali būti geresnis iš modelio pašalinus kurį nors kintamąjį.

Rungtynių kiekis, kurios įvertintos, kaip gana tiksliai atspėtos (jų kvadratinė paklaida yra mažesnė nei 0,83 neuroninių tinklų atveju bei mažesnė nei 0,95 SEM atveju) dalis yra 54%.

Modelio rezultatai pateikti lentelėje. Modelio Chi-kvadrato statistika yra ženkliai mažesnė už 0,05, todėl galime teigti, jog regresoriai gerai tinka duomenims.

Kintamasis	Įvertis	St.nuokrypis	z-reikšmė	P-reikšmė
Intercept	-653	156	-4,2	< 0,05
AveragePTS	-3,6	1,2	-2,9	< 0,05
FTA	1,8	0,6	2,9	<0,05
AST%	0,08	0,02	2,9	<0,05
TS%	0,08	0,02	2,9	<0,05
VarAveragePTS	-0,14	0,05	-2,7	<0,05
VarFGA	0,36	0,1	3,4	<0,05
VarREB	-2,1	0,5	-3,9	<0,05
VarOREB %	0,9	0,2	3,8	<0,05
VarDREB %	3,4	1,0	3,4	<0,05
VarOpeFG %	-0,65	0,3	-2,4	<0,05
VarOpp3p %	0,44	0,15	2,9	<0,05
VarOppOREB	5,7	1,7	3,3	<0,05
VarOppAST	0,34	0,1	2,6	<0,05
Var%PTS2PT	-0,11	0,03	-2,9	<0,05

Teigiamas ženklas modelyje reiškia, jog kuo didesnis tas rodiklis, tuo didesni šansai, kad abu modeliai rungtynes atspės gana tiksliai (tiksliau nei vidurkis). Ir priešingai - jeigu ženklas neigiamas, kuo didesnis rodiklis, tuo didesnė tikimybė, kad modeliai rungtynes prognozuos ne taip tiksliai.

Logistinės regresijoje itin svarbi klasifikavimo rezultatų lentelė. Kadangi žinome, jog mūsų duomenyse 54% rungtynių yra pažymėtos kaip 1, todėl jei modelio tikslų identifikuotų rungtynių kiekis nesiektų 54% galėtumėme daryti išvadą, jog jo vertė yra ne itin didelė, arba iš viso jos nėra. Klasifikavimo lentelė pateikta žemiau. Bendrai modelis tiksliai klasifikavo 62% rungtynių. Tiesa, itin

tiksliai identifikuoti rungtynių, kurios SEM ir neuroninių tinklų modelių nebus tiksliai atspėtos, nepavyko. Tokių prognozių tikslumas siekia 52%, kai tuo tarpu tiksliai atspėti, kada modeliai rungtynes prognozuos geriau nei vidurkis pavyko 71% kartų.

	0	1
FALSE	124 (0,52)	86 (0,29)
TRUE	114 (0,48)	208 (0,71)

Svarbus pastebėjimas, jog į modelį buvo įtraukti net 10 kintamųjų, kurie apibūdina varžovų komandos rodiklius, tuo tarpu stebimos komandos rodiklių kiekis yra 4. Atsižvelgiant į tai, galime daryti išvadą, jog galbūt kai kuriais atvejais modeliuose per mažai buvo įtraukta ir įvertinta varžovų įtaka. Nors logistinės regresijos modelis nesuteikia labai aiškaus vaizdo, kada SEM ir neuroninių tinklų modelių paklaida gali būti didelė, tačiau kartus, kuomet tikėtina, kad paklaida bus nedidelė, jis sugeba identifikuoti gana gerai.

4.10 Lažybos ir modelių testavimas lažyboms

Šiame skyriuje testuosime ir aprašysime, kaip sudaryti modeliai veiktų, jeigu juos bandytumėme pritaikyti lažybų punktuose statant už tam tikrus įvykius. Taip pat, vėliau, kuriamos strategijos, padedančios išlošti didesnę kiekį pinigų.

Nors šioje tokia įžanga apie lažybas ir koeficientus buvo pateikta pačioje darbo pradžioje, šioje vietoje šiek tiek aiškiau ir konkrečiau pateiksime, kaip veikia lažybų kontoros, kaip yra nustatomi koeficientai ir kitus, šiam skyriui svarbius dalykus.

Pradėkime nuo to, jog lažybų kontorose klientams yra pateikiama pačių įvairiausių pasiūlymų. Žmonės gali statyti už įvairiausio tipo įvykius. Kadangi darbo objektas yra NBA krepšinis, tai pateiksime keletą krepšinio pasiūlymų pavyzdžių.

Komanda X įveiks komandą Y.

Komanda X pirmoji pelnys taškus rungtynėse.

Rezultatyviausias rungtynių kėlinys.

Bendra komandų pelnytų taškų suma bus lyginė.

Žaidėjas A iš komandos Y pelnys daugiau taškų nei 10,5.

Komanda X pelnytų taškų suma 99,5.

Būtent paskutinis pavyzdys mums yra aktualiausias. Kadangi sukonstruoti modeliai prognozuoja kiek tam tikra komanda pelnys taškų specifinėse rungtynėse, tokio tipo pasiūlymus, kaip „Komanda X pelnytų taškų suma 99,5” galime prognozuoti ir stebėti, kaip gerai modelis veikia.

Lažybų kontorų pateikiamų koeficientų reikšmės teoriškai svyruoja intervale $(1; +\infty)$, tačiau realybėje didesnių nei 1000 koeficientų pasitaiko itin retai. Koeficiento prasmė labai paprasta. Tarkime įvykiui „Komanda X pelnytų taškų suma didesnė nei 99,5” lažybų kontora paskyrė koeficientą 1,9. Jeigu pasirenkate šį įvykį ir pastatote 100 eurų už šį įvykį, laimėjimo atveju jūsų sąskaitoje bus 190 eurų. Pralaimėjimo atveju jūs pralaimite visus 100 eurų.

Lažybų kontoros, netgi žinodamos, jog šansai kad įvykis įvyks yra $\frac{1}{2}$, tokiam įvykiui priskiria koeficientą mažesnį nei 2. Ilgu laikotarpiu atsitiktinai statydami už tokiu įvykius bankrotas būtų neišvengiamas. Geriausias to pavyzdys pasiūlymas – Bendra komandų pelnytų taškų suma bus lyginė. Empiriškai patvirtina, jog tokio įvykio krepšinyje tikimybė yra $\frac{1}{2}$, tačiau lažybų kontoros tokiam įvykiui duoda 1,9 arba net mažesnį koeficientą – likusi dalis atitenka pačiai lažybų kontorai. Be tarpininkų, tokio įvykio koeficientas būtų lygus 2. Priklausomai nuo lažybų kontorų, statant už tokius įvykius, sėkmingai pastatytų statymų kiekis turi būti didesnis nei 50%, norint ilgame laikotarpyje nebankrotuoti. Kai koeficientas yra 1,9, tokiu atveju reikia tiksliai atspėti $1/1,9 \cdot 100 = 52,6\%$ baigčių, norint nebankrotuoti ilgame periode.

Istorinius lažybų koeficientus pateikia interneto puslapis oddsportal.com. Remdamiesi šio puslapio koeficientais stebėsime, kaip sudaryti modeliai veikia tikroje lažybų aplinkoje. Tyrimo imtis - 532 baigtys.

Iš praėjusio skyriaus matyti, jog kai kurioms komandoms, arba skirtingą puolimo ir gynybos lygį turinčioms komandoms modeliai rezultatą prognozuoja geriau nei kitoms. Užsimerkti ir akiai statyti už tokių įvykių baigtis galime, tačiau tai nebūtų pats racionaliausias sprendimas. Gali būti, jog lažybų kontoros pateikia tokius pasiūlymus, kurių tikslumas yra netgi geresnis nei mūsų sukonstruoto modelio. Pavyzdžiui mūsų neuroninių tinklų modelio, rungtynėms, kuriose yra žaidžiama prieš geriausią gynybą turinčias komandas, vidutinė kvadratinė paklaida siekia 0,72. Jeigu lažybų kontorų pateikiamų pasiūlymų paklaida yra mažesnė, tokiu atveju mūsų statymai patirtų nuostolį.

Jau iš testavimo paklaidų buvo matyti, jog neuroninių tinklų modelis pelnytų taškų kiekį prognozuoja tiksliau nei struktūrinių lygčių modelis. Dėl to,

tikėtis, jog lažybose SEM veiks geriau už neuroninių tinklų modelį, neverta.

21 paveikslėlyje pateikta kaip gerai neuroninių tinklų modelis ir struktūrinių lygčių modelis veikia lažybų sąlygomis bei kaip gerai jie veikia, kai skirtumas tarp modelio prognozuotos reikšmės ir lažybininkų siūlomoms reikšmės skiriasi per tam tikrą dydį.

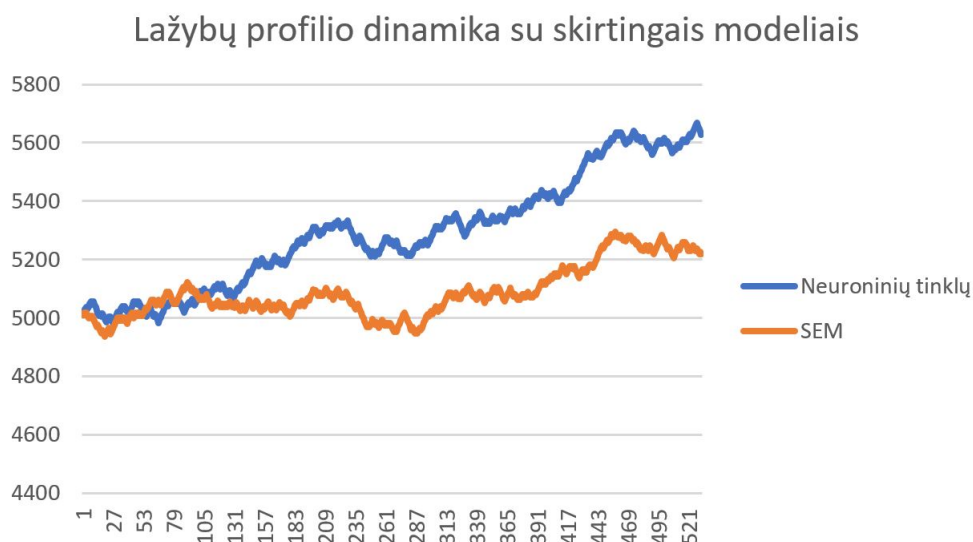
Pastebėtina, jog dirbtinių neuroninių tinklų modelis didėjant skirtumui tarp siūlomo lažybininkų taškų kiekio ir pačio modelio prognozuojamos reikšmės, laimi vis didesnę kartą kiekių. Jeigu skirtumas tarp lažybininkų pasiūlyto taškų kiekio ir modelio išprognozuoto yra didesnis nei 10, tokiais atvejais net 71% kartų modelis buvo teisingas sakdamas, kad iš tiesų komandos įmestų taškų kiekis bus didesnis arba mažesnis nei siūlė lažybų kontoros. Tiesa, tokių rungtynių dalis imtyje nebuvo itin didelė, 11,6%. Tuo tarpu rungtynių skaičius, kuriose lažybininkų ir modelio prognozuojamų taškų skirtumas buvo didesnis nei 5 taškai sudaro 49%. Jeigu statymai būtų daromi tik už tokio tipo rungtynes, tuomet laimėjimų procentas siektų 62%.

		SEM	Rungtynių kiekis	Neuroninių tinklų	Rungtynių kiekis
	Iš viso% laimėjimų OVER/UNDER	54%	532	57%	532
Minimalus taškų skirtumas tarp modelio ir lažybininkų pasiūlymo	1	53%	467	59%	471
	2	54%	425	59%	424
	3	54%	373	58%	378
	4	55%	320	60%	319
	5	54%	270	62%	260
	6	57%	222	63%	209
	7	58%	180	63%	149
	8	59%	141	65%	124
	9	58%	95	67%	86
	10	54%	72	71%	62

21 pav.

22 paveikslėlyje pateikta lažybų profilio dinamika didėjant rungtynių kiekiui. Pradinis biudžetas 5000 eurų. Iš esmės jis nėra svarbus, svarbu tik tai, jog jis nebūtų pakankamai mažas palyginti su statoma suma, nes tokiu atveju patekus į nesėkmingą ruožą galima bankroto tikimybė išauga. Už kiekvieną mačą yra statoma 10 eurų. SEM modelis, po 532 rungtynių, su tokio dydžio statymais būtų sugeneravęs 228 eurus pliuso. Neuroninių tinklų modelis - 628 eurus pelno. Paminėtina tai, jog SEM modelis per visas 532 rungtynių daugiausiai pralaimėjo 5 kartus iš eilės. Norint grįžti į pradinį 5000 eurų tašką, reiktų jog nesėkmių

serija truktų 23 rungtynes.



22 pav.

Strategijos

Iš pirminės analizės matyti, jog neuroninių tinklų modelis prognozuoja geriau nei struktūrinių lygčių modelis, todėl visos lažybų strategijos bus bandomos neuroninių tinklų modeliui.

1 strategija: naivusis statymas, statoma aklai – jeigu modelis prognozuoja daugiau taškų nei lažybininkai statoma „over“, jeigu mažiau – „under“

2 strategija: statoma ne aklai, o tik už tuos įvykius, kur skirtumas tarp lažybininkų ir modelių prognozuojamo taškų kiekis yra didesnis nei tam tikras nustatytas kiekis.

3 strategija: naivusis metodas, tačiau statoma už tam tikro gynybos lygio ir tam tikro puolimo lygio ekipas, kuomet prognozuojame jų rezultatą. Ekipų puolimo ir gynybos lygiai paimti iš sudaryto SEM modelio.

Daroma viena prielaida, jog koeficientas už tokius įvykius yra lygus 1,95. Tokio dydžio koeficientą šiuo metu siūlo viena populiariausių lažybų tarpininkų Lietuvoje. Statoma suma už kiekvieną baigtį 10 eurų. Grafiniam atvaizdavimui, kuris pateiktas priede naudojamas pradinis biudžetas 5000 eurų.

Iš 22 paveikslėlio matyti, jog statant už įvykius, kur skirtumas tarp lažybininkų ir siūlomo modelio rezultato viršija 6, pasiekama 22% pelnas, jeigu skaičiuotumėme nuo visos sumos, kuria buvo rizikuota. Yra strategijų, kurios duoda pakankamai didelį pelningumą, tačiau tokių rungtynių kiekis yra labai nedidelis. Pavyzdžiui komandos puolimas P3 ir varžovų gynyba G3 (susiduria blogiausią puolimą turinčios komandos, prieš geriausią gynybą turinčias koman-

das), tokių rungtynių laimėjimų atvejis yra labai auštas, tačiau tokių rungtynių dalis vos vos viršija 5%.

Strategija	Tipas	Rungtynių kiekis	Pastatyta pinigų suma	Laimėta pinigų	Santykis	
	Visi	532	5320	5947,5	1,12	
	Namų komandos	266	2660	2964	1,11	
	Išvykos komandos	266	2660	2983,5	1,12	
	Puolimas P1	107	1070	1072	1,00	
	Puolimas P2	322	3220	3627	1,13	
	Puolimas P3	103	1030	1248	1,21	
	Varžovų gynyba G1	136	1360	1443	1,06	
	Varžovų gynyba G2	307	3070	3451,5	1,12	
	Varžovų gynyba G3	89	890	1053	1,18	
	Puolimas P2 & varžovų gynyba G2	186	1860	2184	1,17	
	Puolimas P2 & varžovų gynyba G1	80	800	819	1,02	
	Puolimas P2 & varžovų gynyba G3	56	560	624	1,11	
	Puolimas P1 & varžovų gynyba G2	60	600	585	0,98	
	Puolimas P1 & varžovų gynyba G1	29	290	273	0,94	
	Puolimas P1 & varžovų gynyba G3	18	180	214	1,19	
	Puolimas P3 & varžovų gynyba G2	61	610	682,5	1,12	
	Puolimas P3 & varžovų gynyba G1	27	270	351	1,30	
	Puolimas P3 & varžovų gynyba G3	15	150	214,5	1,43	
	Naivusis					
	Skirtumas tarp lažybininkų ir prognozės	2	424	4240	4855,5	1,15
4		319	3190	3724,5	1,17	
6		209	2090	2554,5	1,22	
8		124	1240	1560	1,26	
10		62	620	858	1,38	

23 pav.

Paskutinė testuota strategija - neuronų tinklų ir SEM modelių panaudojimas kartu. Statymas atliekamas tik tada, jei abu modeliai prognozuoja vienodą baigtį, pavyzdžiui, kad komanda įmes daugiau nei siūlo lažybų kontora. Tokių rungtynių kiekis sudaro 67% visų testuotų mačų. Rezultatai yra geresni nei statant remiantis atskirais SEM ir neuroninių tinklų modeliais - laimėjimų kiekis yra 58% kartų (SEM atveju 54%, o neuroninių tinklų atveju 57%). Visgi, tai nebūtų pati geriausia strategija. Paprastu neuronų tinklų atveju, kuomet skirtumas tarp lažybininkų pasiūlymo ir neuroninių tinklų prognozės viršija 1 tašką, rungtynių dalis, už kurias pagal tokią strategiją statytumėme yra 87%, o laimėjimų kiekis taip pat didesnis - 59%.

5 Išvados

Darbe tiriami tam tikrose NBA rungtynėse įmestų taškų struktūra. Naudojami metodai - struktūrinių lygčių modeliavimas bei neuroninių tinklų modeliai. Siekiama ne tik kuo geriau prognozuoti taškų kiekį, kurį įmeta komandos tam tikrose rungtynėse, tačiau kartu suprasti veiksnius ir struktūrą, kuri tai nulemia. Testuojama autoriaus sudaryta SEM modelio struktūra gana gerai tinka duomenims. Atrandami kintamieji, kurie tinka sudarytai struktūrai ir iš esmės nusako, komandų įmestų taškų kiekį. Prienama prie išvados, jog varžovų komandos gynybos ir puolimo latentiniai kintamieji yra svarbūs bendroje struktūroje, jie pagerina modelio statistines charakteristikas, o taip pat jo prognozę. Nors SEM modeliai įmestų taškų kiekį prognozuoja blogiau nei neuroninių tinklų modeliai, jų nauda galima išvelgti kitur. Pasinaudojant struktūra ir įvertintais latentiniais komandų puolimo ir gynybos kintamaisiais, komandos yra suskirstomos į skirtingus gynybos ir puolimo lygius, o tai padeda geriau suprasti ir pačius neuroninių tinklų modelius, kaip jie veikia, kuriose vietose jie prognozuoja geriau, o kur prognozė nėra tokia tiksli. Atlikus prognozių paklaidų analizę, matyti, kad varžovų komandos rodikliai yra labai svarbūs nusakant tikėtinumo lygį, kada modeliai tiksliai prognozuos rungtynių taškų kiekį. Pritaikius sudarytų modelių prognozes lažyboms, tiek SEM, tiek neuroninių tinklų modelis per tirtą laikotarpį net ir statant naiviai demonstravo teigiamus rezultatus. Įvedus įvairias strategijas, laimėjimų santykis tampa dar aukštesnis. Galime daryti išvadą, jog tinkamai parinkus duomenų rinkimo strategiją (naudojami visi prieinami duomenys iki rungtynių pradžios) bei sudarius tinkamus modelius, laimėti statant lažybų kontorose yra pasiekiamas tikslas.

Literatūra

- [1] T.Baghal Are the "Four Factors" Indicators of One Factor? An Application of Structural Equation Modeling Methodology to NBA Data in Prediction of Winning Percentage, *Journal of Quantitative Analysis in Sports*, 2012.
- [2] J.Coughlan, M.R. Mullen, D.Hooper Structural Equation Modelling: Guidelines for Determining Model Fit, *Electronic Journal of Business Research Methods*, Volume 6 Issue 1, 53-60, 2008.
- [3] B.Chen, J.Pearl Graphical Tools for Linear Structural Equation Modeling, Technical Report, 2015.
- [4] D.Forrest, R.Simmons, Forecasting Sport: The behaviour of performance of football tipsters, *International Journal of Forecasting*, 16, 331-337, 2000.
- [5] M.Haghighat, H.Rastegari, N.Nouzafza A review of data mining techniques for result of prediction in sports, *Advanced in Computer Science: an International Journal* 2, 7-12, 2013.
- [6] D.Harville, M.H. Smith The home court advantage: How large is it and does it vary from team to team? *The American Statistician*, 48, 22-28, 1994.
- [7] J.J. Hox, T.M. Bechger Structural Equation Modeling, *Family Science Review* 11, 354-373. 2004.
- [8] R.E. Schumacker, R.G. Lomax A Beginner's guide Structural Equation Modeling Third, 2010.
- [9] L. Hu, P.Bentler Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55, 1999.
- [10] M.Spann, B. Skiera Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets Betting Odds and Tipsters, *Journal of Forecasting* 28, 55-72 2009.
- [11] K.O. Stanley, R.Mikkulainen Evolving Neural Networks Through Arguing Topologies, *The MIT Press Journals* 2002

- [12] S.Warren, Neural Networks and Statistical Models, Nineteenth Annual SAS Users Group International Conference, 1994.

A Priedas

1 priedas: Rodiklių paaiškinimas

AveragePTS - vidutiniškai pelnomų taškų kiekis

FGA - išmetamų metimų kiekis per rungtynes

FG% - pataikymo procentas

3PA - išmetamų tritaškių kiekis

3P% - tritaškių pataikymo procentas

FTA - išmetamų baudų kiekis

FT% - baudų pataikymo procentas

OREB - atkovotų kamuolių kiekis puolime

DREB - atkovotų kamuolių kiekis gynyboje

REB - bendras atkovotų kamuolių kiekis

AST - rezultatyvių perdavimų kiekis

TOV - klaidų kiekis

STL - perimtų kamuolių kiekis

BLK - blokuotų metimų kiekis

PF - pražangų kiekis

OffRtg - komandos puolimo reitingas

DefRtg - komandos gynybos reitingas

AST% - procentas metimų, įmestų po rezultatyvaus perdavimo

AST/TO - rezultatyvių perdavimų ir klaidų santykis

OREB% - atkovotų kamuolių procentas puolime

DREB% - atkovotų kamuolių procentas gynyboje

REB% - atkovotų kamuolių procentas

TOV% - klaidų procentas

TS% - išvestinis rodiklis - tikrasis pataikymo procentas

PACE - greitis, skaičiuojamas kaip atakų kiekis per 48 žaidimo minutes

OpeFG% - varžovų pataikymo procentas

OpFTA Rate - varžovų komandos baudų metimų kiekio ir iš viso mestų metimo santykis

OpTOV% - varžovų komandos atakų kiekis procentais, kuris pasibaigia klaidomis

OpOREB% - varžovų komandos atkovotų kamuolių kiekis procentais puolime

OppPTS - varžovų komandos vidutiniškai pelnomų taškų kiekis

OppFG% - varžovų komandos pataikymo procentas

Opp3P% - varžovų komandos pataikymo procentas nuo trijų taškų zonos

OppFTA - varžovų išmestų baudų kiekis

OppOREB - varžovų puolime atkovotų kamuolių kiekis

OppDREB - varžovų gynyboje atkovotų kamuolių kiekis

OppAST - varžovų komandos vidutiniškai atliekamų rezultatyvių perdavimų kiekis

OppTOV - varžovų komandos padarytų klaidų kiekis

OppSTL - varžovų komandos perimtų kamuolių kiekis

OppBLK - varžovų komandos blokuotų metimų kiekis

OppPTSofTOV - varžovų komandos taškai pelnyti po priešininkų klaidos

OppPTS2ndchance - varžovų taškai, pelnyti atkovojus kamuolį puolime

OppPTSFB - varžovų komandos taškai pelnyti greitose atakose

OppPTSPaint - varžovų komandos taškai, pelnyti baudos aikštelėje

%FGA2PT - išmestų dvitaškių kiekis procentais

%FGA3PT - išmestų tritaškių kiekis

%PTS2PT - taškų kiekis procentais, įmestas iš dviejų taškų zonos

%PTS2PT-MR - taškų kiekis procentais, pelnytas iš vidutinio nuotolio

%PTS3PT - taškų kiekis procentais, pelnytas nuo trijų taškų zonos

%PTSFBPs - taškų kiekis procentais, pelnytų greitose atakose

%PTSoffTO - taškų kiekis, pelnytas po varžovų klaidų

%PTSITP - taškų kiekis procentais, pelnytas baudoje aikštelėje

2FGM%AST - įmesti dvitaškiai, po rezultatyvių perdavimų

3FGM%AST - įmestų tirtaškių kiekis procentais, po rezultatyvių perdavimų

Drives - prasiveržimų kiekis

DrivesPTS% - taškų kiekis procentais, pelnytas pro prasiveržimų

DrivesPass% - procentas atakų, kuomet besiveržiant yra atliekamas perdavimas

DrivesTO% - procentas atakų, kuomet besiveržiant yra suklystama

DFGA - išmestų varžovų metimų kiekis, kuomet metimas buvo atliktas esant ne laisvam

DrivesPTS - taškų kiekis, vidutiniškai pelnomas besiveržiant

DrivesFG% - pataikymo procentas, kuomet yra veržiamasi

CatcShootPTS - taškų kiekis, kuomet yra taškai pelnomi iškart po perdavimo, nesumušant kamuolio į žemę

CatchandShootPTS% - taškų kiekis procentais, kuomet yra taškai pelnomi iškart po perdavimo, nesumušant kamuolio į žemę

PullupPTS - taškai pelnyti sugavus kamuolį bei iškart išmetus arba padarius ne daugiau kaip vieną papildomą judesį

PullupPTS% - taškų procentas, kurie pelnyti sugavus kamuolį bei iškart išmetus arba padarius ne daugiau kaip vieną papildomą judesį

PullupPTSFG% -pataikymo procentas metimų, kurie buvo išmesti sugavus kamuolį bei iškart išmetus arba padarius ne daugiau kaip vieną papildomą judesį

PaintPTS - taškai pelnyti baudos aikštelėje

PaintFG% - taškų kiekis procentais, pelnytas baudos aikštelėje

PostFG% - pataikymo procentas baudos aikštelėje

PostPTS - taškai pelnyti apribotoje zonoje

PostFG% - pataikymo procentas apribotoje zonoje

ContestedOreb - atkovoti kamuoliai puolime, nurungiant besivaržantį priešiniką

Orebchances - galimybių kiekis atkovoti kamuolį puolime

ScreenAst - pataikytų metimų kiekis, kai tam progą sukūrė pastatyta sienelė

Deflections - kiekis, kuomet yra paliečiamas kamuolys jį kontroliuojant varžovams

LooseBallRecover - kiekis kartų, kuomet kamuolys yra iškart perimams jį praradus

ChargesDrawn - išprovokuotų pražangų kiekis puolime

Contested2ptsshoots - metimai iš dviejų taškų zonos, kada varžovai metė būdami nelaisvi

Contested3ptsShoots - metimai iš trijų taškų zonos, kada varžovai metė būdami nelaisvi

DistMilesOf - nubėgtas komandos atstumas puolant

DistMilesDF - nubėgtas atstumas komandai besiginant

AvgSpeedOf - vidutinis greitis komandai puolant

AvgSpeedDF - vidutinis greitis komandai besiginant

ContestedDreb - atkovotų kamuolių kiekis gynyboje, kuomet buvo varžovamasi su kitos komandos žaidėju

Drebchances -galimybės atkovoti kamuolį gynyboje

DrebDistance - atstumas nuo krepšio, kuomet kamuolys yra atkovojamasis gynyboje

PostTouches - kiekis kartų, kuomet kamuolys atakos metu yra liečiamas apribotoje zonoje

RAFGM - kiekis metimų, kurie buvo įmesti apribotoje zonoje

RAFG% - pataikymo procentas iš apribotosios zonos

PaintFGM - metimų kiekis, kurie buvo pelnyti baudos aikštelėje

PaintFG% - pataikymo procentas baudos aikštelėje

EllbowPTS - taškai pelnyti iš zonos, kur susikerta baudos linija ir baudos aikštelės linijos

EllbowFG% - pataikymo procentas iš zonos, kur susikerta baudos linija ir baudos aikštelės linijos

MidFG% - pataikymo procentas iš vidutinio nuotolio

MidFGM - pataikytų metimų kiekis iš vidutinio nuotolio

L3FGM - pataikytų tritaškių kiekis iš kairiojo atakos krašto

L3FG% - pataikytų tritaškių procentas iš kairiojo atakos krašto

R3FGM - pataikytų tritaškių kiekis iš dešiniojo atakos krašto

R3FG% - pataikytų tritaškių procentas iš dešiniojo atakos krašto

3OTHERFGM - pataikytų tritaškių kiekis ne iš kraštų

3OTHERFG% - pataikytų tritaškių procentas ne iš kraštų

OPRAFGM - oponentų taškų kiekis pelnytas apribotoje zonoje

OPRAFG% - oponentų pataikymo procentas apribotoje zonoje

OPPaintFGM - oponentų pataikytų metimų kiekis baudos aikštelėje

OPPaintFG% - oponentų pataikytų metimų procentas baudos aikštelėje

OPMidFGM - oponentų pataikytų metimų kiekis iš vidutinio nuotolio

OPMidFG% - oponentų pataikytų metimų procentas iš vidutinio nuotolio

OPL3FGM - oponentų pataikytų tritaškių kiekis iš kairiojo krašto

OPL3FG% - oponentų pataikytų tritaškių procentas iš kairiojo krašto

OPR3FGM - oponentų pataikytų tritaškių kiekis iš dešiniojo krašto

OPR3FG% - oponentų pataikytų tritaškių procentas iš dešiniojo krašto

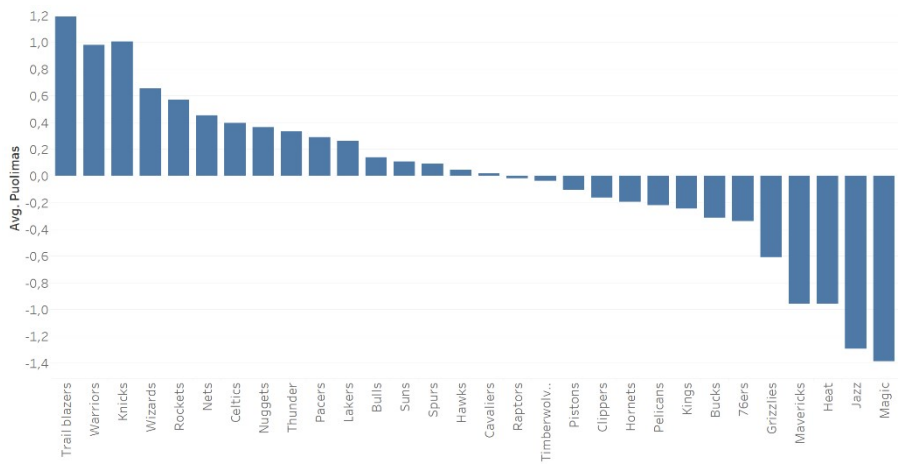
OP3OTHERFGM - oponentų pataikytų tritaškių kiekis ne iš kraštų

OP3OTHERFG% - oponentų pataikytų tritaškių procentas ne iš kraštų

W% - pergalių procentas

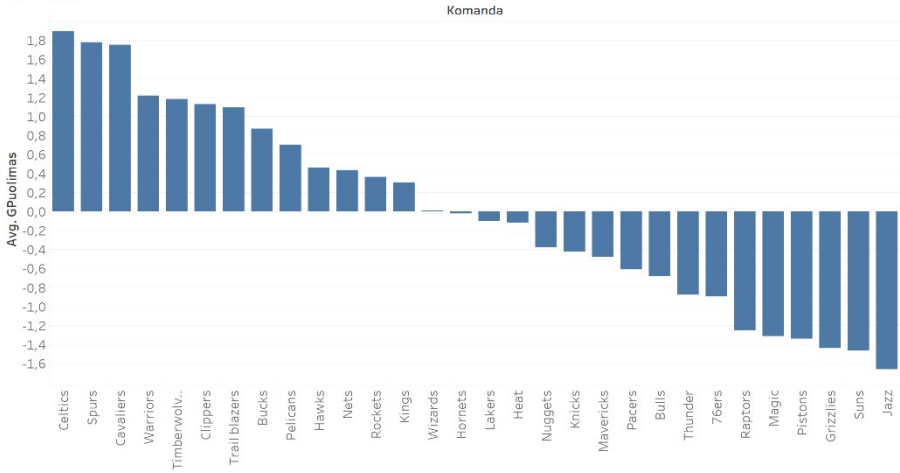
Visi kiti kintamieji buvo surinkti ir stebimos komandos varžovams. Duomenų lentelėje šalia kiekvieno rodiklio pridėtas varžovo (Var) trumpinys.

Puolimo latentinis kintamasis



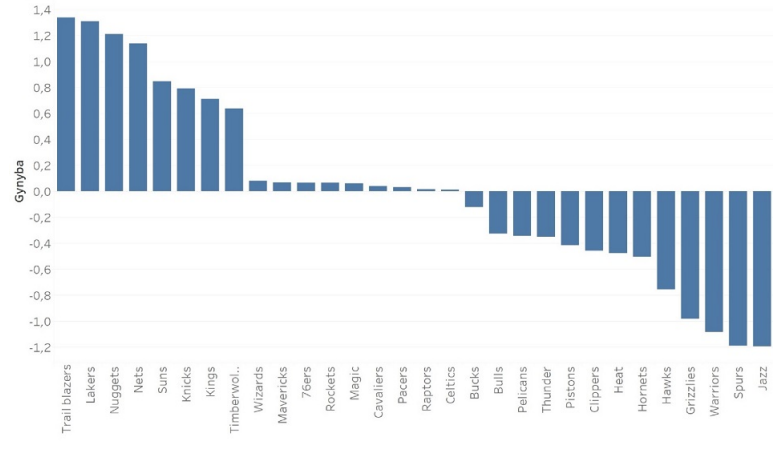
24 pav.

Gynėjų puolimo latentinis kintamasis



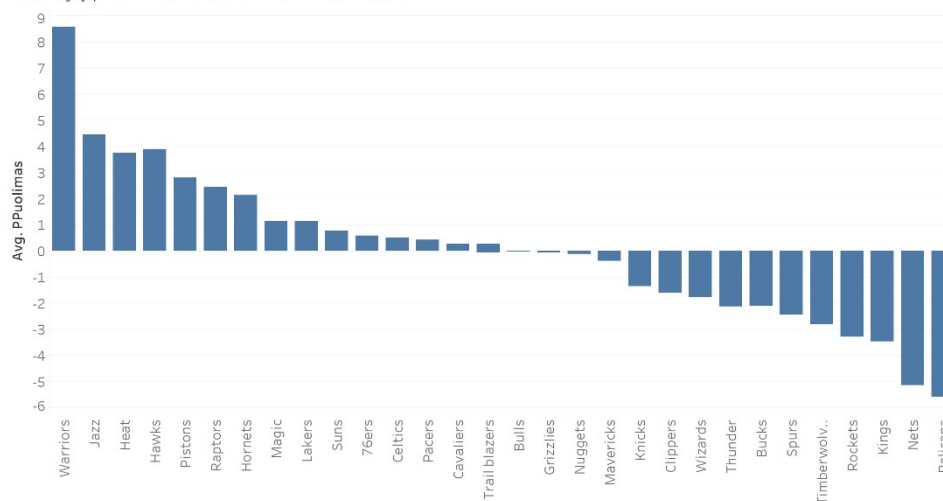
25 pav.

Gynybos latentinis kintamasis komandoms



26 pav.

Puolėjų puolimo latentinis kintamasis



27 pav.

	Puolimas		Gynyba	
	SEM	MSE	SEM	MSE
76ers	0,402712438	0,743026468	1,282458	1,346799619
Bucks	0,667981582	0,862501444	0,758665	0,749856093
Bulls	1,29505081	0,988827903	0,928943	0,810020205
Cavaliers	1,499627026	1,058678634	0,819427	1,01195633
Celtics	0,472417794	0,16229273	0,739855	0,808292502
Clippers	0,8578746	0,464313701	1,457948	0,924615077
Grizzlies	0,686557439	0,803053737	0,865353	0,580859341
Hawks	1,044855236	0,776673347	0,875298	0,774597622
Heat	0,808006582	1,387191064	0,769976	0,887953343
Hornets	0,849726984	0,762423638	1,018336	1,209991266
Jazz	1,15956157	0,718341146	0,932074	0,645855509
Kings	1,068718638	1,03895257	0,737468	0,423936445
Knicks	0,509962881	0,327303662	0,77189	1,024838905
Lakers	0,613270153	0,578502222	2,037622	1,331473226
Magic	1,030443168	0,941868834	0,697163	0,482582839
Mavericks	1,148861561	0,843995407	0,537367	0,524817375
Nets	0,832571745	0,441686647	0,854473	0,761639211
Nuggets	1,397708365	0,868959143	0,682744	0,47325615
Pacers	0,516161422	1,143610046	0,680603	0,660533045
Pelicans	1,09837782	1,19689024	1,24889	1,151361883
Pistons	1,372179428	1,191857573	0,563805	0,68712532
Raptors	0,612950261	1,365318763	1,37614	0,942355164
Rockets	1,95084548	0,75502317	1,187716	0,854686466
Spurs	0,676622683	0,799835949	0,805295	0,919488292
Suns	1,12320732	0,713581757	1,18618	1,044027149
Thunder	0,906588889	0,714539373	0,828235	0,500097217
Timberwolves	0,77196137	0,650877583	1,732756	1,568637698
Trail blazers	1,26072526	0,635158958	0,730316	0,601215544
Warriors	1,060688848	1,722940896	0,929606	0,692606505
Wizards	1,059518294	0,387595424	0,828024	0,772589829

28 pav.