



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
MULTIMEDIJOS INŽINERIJOS KATEDRA

ADAS URBONAS

**SPORTO ĮVYKIŲ PROGNOZAVIMO GALIMYBIŲ TYRIMAS
NAUDOJANT NEURONINIUS TINKLUS**

Magistro darbas

Darbo vadovas
prof. dr. A.Misevičius

KAUNAS, 2016



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
MULTIMEDIJOS INŽINERIJOS KATEDRA

ADAS URBONAS

**SPORTO ĮVYKIŲ PROGNOZAVIMO GALIMYBIŲ TYRIMAS
NAUDOJANT NEURONINIUS TINKLUS**

Magistro darbas

Darbo vadovas:
prof. dr. A.Misevičius

(data, parašas)

Recenzentas:

(data, parašas)

Studentas:
A.Urbonas

(data, parašas)

KAUNAS, 2016

AUTORIŲ GARANTINIS RAŠTAS DĖL PATEIKIAMO KŪRINIO
20.. - - d.
Kaunas

Autorius, _____
(vardas, pavardė)

patvirtina, kad Kauno technologijos universitetui pateiktas baigiamasis bakalauro (magistro) darbas (toliau vadinama – kūrinys)

(kūrinio pavadinimas)

pagal Lietuvos Respublikos autorių ir gretutinių teisių įstatymą yra originalus ir užtikrina,
kad

- 1) jį sukūrė ir parašė Kūrinyje įvardyti autoriai;
- 2) Kūrinys nėra ir nebus įteiktas kitoms institucijoms (universitetams) (tiek lietuvių, tiek užsienio kalba);
- 3) Kūrinyje nėra teiginių, neatitinkančių tikrovės, ar medžiagos, kuri galėtų pažeisti kito fizinio ar juridinio asmens intelektinės nuosavybės teises, leidėjų bei finansuotojų reikalavimus ir sąlygas;
- 4) visi Kūrinyje naudojami šaltiniai yra cituojami (su nuoroda į pirminį šaltinį ir autorių);
- 5) neprieštarauja dėl Kūrinio platinimo visomis oficialiomis sklaidos priemonėmis;
- 6) atlygins Kauno technologijos universitetui ir tretiesiems asmenims žalą ir nuostolius, atsiradusius dėl pažeidimų, susijusių su aukščiau išvardytų Autorių garantijų nesilaikymu;
- 7) Autoriai už šiame rašte pateiktos informacijos teisingumą atsako Lietuvos Respublikos įstatymų nustatyta tvarka.

Autorius

(vardas, pavardė)

(parašas)

SANTRAUKA

Šiuo metu, kai informacinės technologijos itin ištobulėjusios ir išplitusios, informacijos sklaida yra be galo didelė. Kiekvieną dieną gauname labai daug informacijos, kurios dalį mūsų smegenys natūraliai ignoruoja, nes tiesiog nėra pajėgios analizuoti ir apdoroti tokio didelio informacijos srauto. Žmogus iš viso informacijos srauto, kuris mus pasiekia, natūraliai atsirenka tik tai, kas jam naudinga ir aktualu, bet tai mūsų smegenys daro paviršutiniškai. Jeigu informacijos srautas didelis, tai žmogaus smegenys nesugeba jos akimirksniu analizuoti, pastebėti tendencijas ir suformuluoti tinkamas išvadas. Būtent dėl pastarosios priežasties mes nesugebame tiksliai numatyti ateities įvykių, nes bet kokio įvykio baigčiai daro įtaką daugybė parametru. Visgi ilginiui analizuodamas žmogus įgauna patirties iš tyrinėjamų duomenų ir gali apytiksliai prognozuoti bei numanyti galimą baigtį.

Pasitelkus kompiuterius, įmanoma modeliuoti įvairius įvykius ir pritaikius prognozavimo algoritmus numatyti juos gana tiksliai. Tikslumas priklauso nuo įvykio srities, turimos informacijos kiekio ir parametru, kurie tiesiogiai daro įtaką įvykio baigčiai. Šio darbo esmė yra nustatyti galimybes prognozuoti sporto įvykius, tam pasitelkiant dirbtinius neuroninius tinklus, kurie imituoja biologinių neuroninių tinklų darbą. Prognozavimo galimybę apima įvykio baigties parametru įtakos nustatymas galutiniai baigčiai. Jeigu turimi parametrai yra svarbūs galutinei baigčiai, galima būtų teigti, kad prognozės bus efektyvios.

SUMMARY

Nowadays, when IT is widely developed, dissemination of information is very high. Each day we are getting a lot of information, part of which our brain normally ignores, because it just is not able to analyze and handle such a large flow of information. Human is naturally choosing useful and true information from the flow that surrounds us; however, our brain is doing it only skin-deep. If information flow is too large, human's brain cannot quickly analyze a lot of information, get sight of trends and make appropriate conclusions. It is for this latter reason we cannot precisely foresee future events, because any kind of event's end is influenced by multiple parameters. While in the long run human being gets experience from exploratory data and approximately can forecast and foreknow the end of certain event.

With the help of computers it is possible to simulate different events and apply forecasting algorithms to predict events pretty accurately. Accuracy depends on event's area, amount of information available and number of parameters, which directly influences occurrence outcome. This work gist is to determine opportunities to predict sports events while using artificial neural networks that mimics biological neural networks. Predicting ability includes event's outcome parameters impact setting for final outcome. If available parameters are solid for final outcome we can state that forecast will be effective.

TURINYS

1. ĮVADAS.....	1
1.1. Probleminė sritis.....	1
1.2. Tyrimo sritis ir objektai.....	1
1.3. Tikslas ir uždaviniai	1
1.4. Dokumento struktūra.....	2
2. ANALIZĖ.....	3
2.1. Biologinis neuronas.....	3
2.2. Dirbtinis neuroninis tinklas	4
2.3. Dirbtinių neuroninių tinklų klasifikacija	5
2.3.1. Tiesioginio sklidimo tinklai	5
2.3.2. Grįžtamojo ryšio tinklai.....	5
2.4. Dirbtinio neuroninio tinklo mokymas	6
2.4.1. Mokymas su mokytoju	6
2.4.2. Mokymas be mokytoju	6
2.4.3. Hibridinis mokymas.....	6
2.4.4. Klaidos atgalinio sklidimo algoritmas	6
2.5. Prognozavimo metodai, paremti DNT	7
2.5.1. DNT taikymas prognozavime	10
2.6. Dirbtinių neuroninių tinklų architektūra.....	10
2.7. Prognozavimo PĮ naudojanti DNT	11
2.7.1. „Aluyda Forecaster“	11
2.7.2. „NeuroSolutions“	12
2.7.3. „EasyNN“	13
2.7.4. „AgielNN“	14
2.7.5. „GMDH Shell“	14
2.7.6. „Orange“	15
3. SPECIFIKACIJA	16
3.1. Projekto planas	16
3.2. Funkciniai reikalavimai.....	18
3.3. Nefunkciniai reikalavimai	18
3.4. Kokybės kriterijai	19
3.5. Projekto duomenų bazė	19
3.6. Duomenų gavyba.....	21
3.7. Duomenų generavimas	22
3.8. Duomenų eksportavimas	24
3.9. PĮ DNT prognozavimui	25
3.10. Testavimas.....	28

3.11. Naudingumo balai	29
3.12. Neuroninio tinklo modelis.....	32
4. EKSPERIMENTINĖ DALIS	34
4.1. Neuroninio tinklo modelio patikrinimas	34
4.2. Duomenų paruošimas	36
4.3. Prognozavimas naudojant DNT	38
4.4. Prognozavimas naudojant regresiją.....	39
4.5. Rezultatų analizė	42
5. IŠVADOS.....	51
6. LITERTŪRA.....	52
7. PRIEDAI	53

LENTELIŲ SĄRAŠAS

Lent. 2.5.1. Prognozavimo modelių pranašumai ir trūkumai.....	9
Lent. 3.5.1. „Varžybos“ lentelės laukelių paaiškinimai	20
Lent. 3.5.2. „Žaidėjai_Varžybose“ lentelės laukelių paaiškinimai	20
Lent. 3.5.3. „Žaidėjai“ lentelės laukelių paaiškinimai	20
Lent. 3.11.1. Naudingumo balo formulės parametrai ir jų vertės	32
Lent. 4.1.1. DNT modelio testavimo rezultatai su pseudoduomenimis	35
Lent. 4.3.1. Kiekvieno sezono prognozavimo rezultatai su DNT naudojant 70 % duomenų apmokymui	39
Lent. 4.3.2. Kiekvieno sezono prognozavimo rezultatai su DNT naudojant 85 % duomenų apmokymui	39
Lent. 4.3.3. Prognozavimo rezultatai su DNT naudojant visų sezonų duomenis	39
Lent. 4.4.1. Kiekvieno sezono kintamųjų reikšmės, gautos iš regresijos naudojant 70 % apmokymo duomenų	40
Lent. 4.4.2. Kiekvieno sezono kintamųjų reikšmės, gautos iš regresijos naudojant 85% apmokymo duomenų	40
Lent. 4.4.3. Visų sezonų kintamųjų reikšmės, gautos iš regresijos naudojant 70 % ir 85 % apmokymo duomenų	41
Lent. 4.4.4. Kiekvieno sezono prognozavimo rezultatai su regresija naudojant 70 % duomenų apmokymui	41
Lent. 4.4.5. Kiekvieno sezono prognozavimo rezultatai su regresija naudojant 85 % duomenų apmokymui	42
Lent. 4.4.6. Prognozavimo rezultatai su regresija naudojant visų sezonų duomenis	42
Lent. 4.5.1. Prognozavimo rezultatų palyginimas.....	42

PAVEIKSLŲ SĄRAŠAS

2.1.1 pav. Biologinio neurono schema	3
2.3.1 pav. Neuroninių tinklų modelių klasifikacija	5
2.4.1 pav. Klaidos atgalinio sklidimo algoritmas	7
2.5.1 pav. Prognozavimo metodai	8
2.7.1 pav. „Aluyda Forecaster“ PĮ programos langas	12
2.7.2 pav. „NeuroSolutions“ programa, integruota Microsoft Excel aplinkoje	13
2.7.3 pav. „EasyNN“ programos langas	13
2.7.4 pav. „AgielNN“ programos langas	14
2.7.5 pav. „GMDH Shell“ programos langas	15
2.7.6 pav. „Orange“ programos langas	15
3.1.1 pav. Projekto plano schema	17
3.5.1 pav. Duomenų bazės schema	19
3.6.1 pav. Duomenų parsisiuntimo programos panaudos atvejų schema	21
3.6.2 pav. Duomenų parsisiuntimo PĮ grafinė sąsaja	22
3.6.3 pav. Duomenų parsisiuntimo procesas	22
3.7.1 pav. Duomenų generavimo programos panaudos atvejų schema	23
3.7.2 pav. Duomenų generavimo programos grafinė sąsaja	23
3.8.1 pav. Duomenų eksportavimo programa	24
3.8.2 pav. Apmokymo duomenų failo išsaugojimas	24
3.8.3 pav. Duomenų eksportavimo panaudos atvejų schema	25
3.9.1 pav. Duomenų importavimas „AgielNN“ programoje	25
3.9.2 pav. DNT suformavimas „AgielNN“ programoje	26
3.9.3 pav. DNT apmokymas „AgielNN“ programoje	27
3.9.4 pav. Prognozavimas su DNT „AgielNN“ programoje	28
3.10.1 pav. Prognozės rezultatų testavimas	29
3.10.2 pav. Rezultatų testavimo procesas	29
3.12.1 pav. Eksperimente naudojamo DNT modelio schema	32
4.1.1 pav. Prognozavimo rezultatų dalis „AgielNN“ programoje naudojant 3 paslėptus sluoksnius	36
4.2.1 pav. Varžybose registruotų žaidėjų naudingumo balai duomenų bazėje	37
4.2.2 pav. Eksportuoti apmokymo duomenys faile	38
4.2.3 pav. Individualaus žaidėjo naudingumo balo apskaičiavimas testavimui	38
4.5.1 pav. DNT ir regresijos prognozavimo tikslumo palyginimas (70 % apmokymo duomenų)	43
4.5.2 pav. DNT ir regresijos prognozavimo tikslumo palyginimas (85 % apmokymo duomenų)	43
4.5.3 pav. Efektyvumo įvertinimas prognozuojant su DNT (70 % apmokymo duomenų)	44

4.5.4 pav. Efektyvumo įvertinimas prognozuojant su DNT (85 % apmokymo duomenų).	45
4.5.5 pav. Efektyvumo įvertinimas prognozuojant su regresija (70 % apmokymo duomenų)	45
4.5.6 pav. Efektyvumo įvertinimas prognozuojant su regresija (85 % apmokymo duomenų)	46
4.5.7 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant regresiją (70 % apmokymo duomenų)	46
4.5.8 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant regresiją (85 % apmokymo duomenų)	47
4.5.9 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant DNT (70 % apmokymo duomenų)	47
4.5.10 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant DNT (85 % apmokymo duomenų)	48
4.5.11 pav. 2009–2015 sezonų efektyvumas prognozuojant su regresija (naudojant visų duomenų 70 %)	49
4.5.12 pav. 2009–2015 sezonų efektyvumas prognozuojant su regresija (naudojant visų duomenų 85 %)	49
4.5.13 pav. 2009–2015 sezonų efektyvumas prognozuojant su DNT (naudojant visų duomenų 70 %)	50
4.5.14 pav. 2009–2015 sezonų efektyvumas prognozuojant su DNT (naudojant visų duomenų 85 %)	50

TERMINŲ IR SANTRUMPŲ ŽODYNAS

DNT – dirbtinis neuroninis tinklas.

PI – programinė įranga.

BNT – biologinis neuroninis tinklas.

Perceptronas – smegenų modelis, kontekste apie dirbtinį intelektą reiškia neuroninį tinklą.

Pseudo – informatikos mokslo šakoje vartojamas priešdėlis, kuris reiškia įvardijamo dalyko netikrumą.

BPM – angl. trump. „Box Plus Minus“ – tai naudingumo balas, naudojamas įvertinti žaidėjo efektyvumą varžybose.

IQ – intelekto trumpinys.

1. ĮVADAS

Šiame tiriamajame darbe nagrinėjamos galimybės prognozuoti įvairių šakų sporto susitikimų baigtis iš sportininkų duomenų istorijos pasitelkiant dirbtinius neuroninius tinklus. Darbo įvade apžvelgiama nagrinėjamos temos probleminė sritis, pagrindžiamas jos aktualumas, aprašomi tyrimo objektai, nurodomi darbo tikslai bei išskelti darbo uždaviniai. Taip pat aptariama viso baigiamojo projekto struktūra.

1.1. Probleminė sritis

Dirbtiniai neuroniniai tinklai (toliau DNT) naudojami objektams klasifikuoti. Skirstant objektus pagal jų požymius yra sprendžiami plataus spektro uždaviniai įvairiose mokslo šakose. Populiariausios pritaikymo sritys, kuriose DNT jau yra sėkmingai taikomi, – tai grafinis vaizdų atpažinimas, medicininė diagnostika, automobilių diagnostika, karinės programos, finansų analizė, bioinformatika, dirbtinis intelektas. Sporte galutinę baigtį lemia labai daug veiksnių, visų jų įvertinti šiuo metu neįmanoma, tačiau naudinga iš duomenų gausos išskirti esminius parametrus, kurie labiausiai daro įtaką baigčiai, ir išsiaiškinti, koks galėtų būti apytikris šių parametrų svoris numatant galutinę baigtį. Kai duomenų yra labai daug, matematiškai apskaičiuoti ir pastebėti vienokias ar kitokias tendencijas gali būti labai sunku.

Šiame darbe nagrinėjama problema – nustatyti, ar įmanoma prognozuoti sporto varžybų baigtis taip, kad prognozių rezultatai teiktų realios praktinės naudos. Problemai spręsti pasitelkiami DNT, regresija ir duomenų gavyba. Efektyvūs darbo rezultatai gali būti naudojami sporto žvalgyboje, efektyviam trenerių darbui, sportininkų savikontrolei ir asmeniniam tobulėjimui, lažybų koeficientams nustatyti ir lažyboms atspėti.

1.2. Tyrimo sritis ir objektai

Šio darbo tiriamoji sritis yra prognozavimų uždavinių sprendimai. Prognozavimo uždaviniai gali skirtis nuo srities pobūdžio. Šiame darbe nagrinėjami uždaviniai, kurie susiję su sporto įvykiais.

Tyrimo objektai – dirbtinių neuroninių tinklų modeliai, prognozavimo metodai, kurie yra paremti DNT, bei matematinė regresija, kuri įprastai naudojama tiriamo objekto parametrų vertėms nustatyti ir taip spręsti prognozavimo uždavinius.

1.3. Tikslas ir uždaviniai

Suformuluotas darbo tikslas – pagrįsti arba paneigti galimybę efektyviai prognozuoti sporto įvykių baigtis pasinaudojant dirbtiniais neuroniniais tinklais. Tikslui pasiekti išskelti šie darbo uždaviniai:

1. Apmokyti dirbtinį neuroninį tinklą pasiremiant pasirinktos sporto šakos realiais duomenimis, dalį duomenų atskirti testavimui.
2. Sugeneruoti duomenis su nurodytais parametru svoriais ir šiais duomenimis apmokyti dirbtinį neuroninį tinklą, dalį duomenų atskyrus testavimui, atlikti prognozavimą.
3. Naudojantis realiais duomenimis atlikti matematinę regresiją ir nustatyti parametru svorius.
4. Naudojantis realiais duomenimis atlikti prognozavimą pasinaudojant dirbtiniu neuroniniu tinklu ir su gautąja tiesinės regresijos formule.
5. Pasinaudojant lažybų koeficientų duomenimis, nustatyti efektyvumą prognozuojant su dirbtiniu neuroniniu tinklu ir su matematine regresija. Palyginti efektyvumą.

1.4. Dokumento struktūra

Dokumento terminų ir santrumpų dalyje yra supažindinama su specifiniais terminais, kurie vartojami šiame darbe. Įvadiniame skyriuje pateikiama darbo esmė, aptariama probleminė sritis, pateikiamas darbo tikslas ir jam pasiekti reikalingi uždaviniai. Analizės skyriuje apžvelgiamas panašumas tarp biologinio ir dirbtinio neuroninio tinklo, neuroniniais tinklais paremti prognozavimo metodai, DNT architektūra bei šiuo metu populiariausia programinė įranga, kuri skirta darbui su DNT. Specifikacijos skyriuje pateikiama tyrimo eiga, tyrimo metu taikomų metodų realizavimas, tyrimui naudojamų duomenų bazės struktūra bei pateikiamos sukurtos programinės įrangos, skirtos dokumentacijai tirti, nurodomi aparatūrinės ir programinės įrangos konfigūracijos parametrai tam, kad eksperimentą būtų galima sėkmingai atkartoti tam tikromis aplinkybėmis. Rezultatų skyriuje yra apibendrinami gauti tyrimų rezultatai, įvertinamas prognozavimo efektyvumas. Išvadų skyriuje pateikiamos pagal darbo uždavinius suformuluotos išvados. Literatūros šaltiniuose pateikiami moksliniai straipsniai, knygos ir internetiniai šaltiniai, kuriais remiantis buvo rengiamas šis darbas.

2. ANALIZĖ

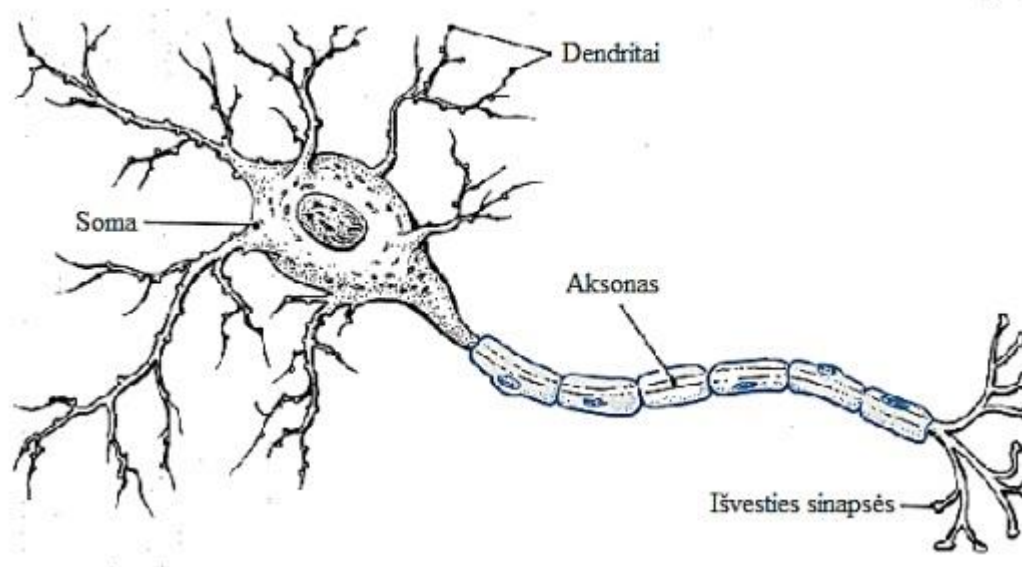
Šiame skyriuje apžvelgiami ir nagrinėjami DNT panašumai su BNT bei DNT veikimo principais, taip pat egzistuojantys prognozavimo metodai, kurie remiasi DNT. Nagrinėjant šį darbą pravartu žinoti jau esamą PĮ, kuri yra naudojama darbui su DNT.

2.1. Biologinis neuronas

Pagrindinis žmogaus nervų sistemos struktūrinis elementas yra nervų ląstelė arba biologinis neuronas. Literatūroje pažymima, kad būtent biologiniais neuronais vyksta informacijos perdavimas iš vienos nervų sistemos dalies į kitą, taip pat informacijos apsikeitimas tarp nervų sistemos ir kitų žmogaus kūno dalių [1].

Biologiniuose neuronuose vyksta itin sudėtingi informacijos apdorojimo procesai. Per juos formuojami viso organizmo refleksai į vidinius ir išorinius dirgiklius.

Biologinis neuronas gali būti skirtingų matmenų ir formų, bet schematiškai jį patogų įsivaizduoti kaip ląstelę su atšakomis (2.1.1 pav.).



2.1.1 pav. Biologinio neurono schema

Biologinis neuronas susideda iš ląstelės kūno (dar vadinamo soma), sudaryto iš branduolio ir atšakų. Biologinio neurono atšakos savo ruožtu yra šios [1]:

- Dendritai – tai biologinio neurono atšakos, kuriomis neuronui perduodami impulsai;
- Aksonas, kuriuo impulsas nuo neurono perduodamas kitoms ląstelėms.

Pagal struktūrą biologinio neurono dentritai yra trumpi ir santykinai platus, sukuria daug sinapsių su kitomis nervų ląstelėmis. Sinapsė (lot. *kontaktas, žinia*) – tai specializuota struktūra, kuri užtikrina kontaktą tarp atskirų biologinio neurono struktūrų. Kiekvieno aksono gale yra sinapsė [1].

Visi impulsai, patenkantys į nervų sistemą, perduodami biologiniam neuronui per tam tikras jo membranos dalis, kurios yra sinapsės kontaktų vietose. Kai signalas pasiekia sinapsę, išsiskiria tam tikri chemikalai (neurotransmiteriai). Skystis (chemikalas) difuzijos būdu skverbiasi į sinapsės terpę ir arba sustiprina, arba susilpnina priimančiojo neurono gebą priimti elektrinius impulsus. Tai priklauso nuo sinapsės. Kuo daugiau sinapsių yra nervų ląstelėje, tuo daugiau impulsų ji gali priimti ir taip dalyvauti didesniame skaičiuje reakcijų, vykstančių žmogaus organizme [2].

Reikėtų pastebėti, kad efektai, kylantys sinapsės aktyvavimo metu, gali būti ir žadinantys, ir slopinantys. Esama situacijų, kai skirtingos sinapsės vienu metu žadina ir slopina neuroną. Šiuo atveju nervų ląstelės reakcija apibūdinama visų sinapsės poveikių suma.

Kartu reikia pabrėžti, kad neuronas sužadinas tik tuo atveju, jei bendras dendritais gautas signalas viršija tam tikrą lygį, vadinamą sužadinimo slenksčiu. Galima teigti, kad neurono sužadinimas – tai jo gebėjimo atsakyti į sinapsių poveikį potencialas. Turint didžiulį skaičių visiškai paprastų elementų, kurių kiekvienas skaičiuoja svorinę įėjimo signalų sumą ir generuoja binarinį signalą, jei suminis signalas viršija tam tikrą lygį, įmanoma atlikti gana sudėtingas užduotis. Būtent šiais pagrindais remiasi dirbtinio neurono idėja. Dirbtinio neurono atveju informacija apdorojama ir paskirstoma apdorojimo elementais (biologiniame neurone tai yra aksonai), sujungtais kartu vienakrypčiais signalų kanalais, t. y. jau minėtomis sinapsėmis. Apdorojimo elementų išėjimo signalas gali būti bet kokio reikalingo matematinio tipo.

Apibendrinant galima teigti, kad biologinis neuronas yra informacijos apie žmogaus nervų sistemos funkcionavimą šaltinis ir kartu tarpininkas tarp skirtingų nervų sistemos elementų skirstant impulsus bei signalus šios sistemos viduje. Biologinio neurono pagrindu funkcionuoja pažangiausios technologijos, pagrįstos signalų arba informacijos srautų apdorojimu ir perdavimu.

2.2. Dirbtinis neuroninis tinklas

Dirbtinis neuroninis tinklas yra masyvus, paralelinis, paskirstytas procesorius, sudarytas iš paprastų skaičiavimo elementų, kurie turi savybę saugoti informaciją ir padaryti ją prieinamą naudoti. Jis panašus į smegenis dviem aspektais:

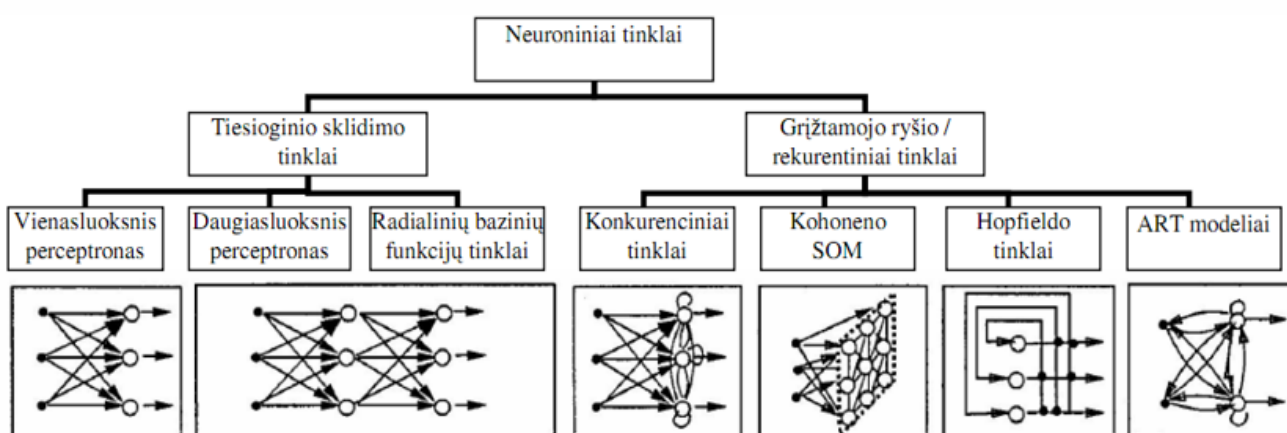
1. Žinios gaunamos iš aplinkos per mokymosi procesą.
2. Žinioms išsaugoti naudojamos neurono reikšmės.

Pasitelkiant IT, yra modeliuojamas tinklas, kuris būtų panašus į biologinį tinklą. Pagrindiniai DNT pranašumai yra šie:

- Masyvus, paralelinis paskirstymas;
- Apibendrinimas – savybė, įgalinanti gauti pakankamai tikslias reikšmes net su duomenimis, kurie buvo nepateikti mokymosi metu.

2.3. Dirbtinių neuroninių tinklų klasifikacija

Dirbtiniai neuroniniai tinklai turi keletą skirtingų modelių (2.3.1 pav.). Jie skiriasi savo struktūra ir galimybėmis. Skiriasi ir jų apmokymo būdai. Tinkamas DNT modelio parinkimas priklauso nuo sprendžiamo uždavinio ir to, kokius rezultatus tikimasi gauti. Ypatingas dėmesys turėtų būti kreipiamas į preliminarų duomenų apdorojimą prieš juos analizuojant, į metodų sudėtingumo, turimų duomenų kiekio ir analizės tikslumo ryšį [3].



2.3.1 pav. Neuroninių tinklų modelių klasifikacija

2.3.1. Tiesioginio sklidimo tinklai

Tiesioginio sklidimo neuroniniuose tinkluose signalai sklinda tik viena kryptimi. Tolimesnio sluoksnio neuronų išėjimai gauna informaciją tik iš prieš tai buvusių sluoksnių. Tokiuose tinkluose nėra nei grįžtamųjų ryšių iš ankstesnių sluoksnių, nei ciklų su tame pačiame sluoksnyje esančiais neutronais. Prie tiesioginio sklidimo neuroninių tinklų priskiriami:

- Vienasluoksnis perceptronas – kai tarp įėjimo ir išėjimo sluoksnių nėra paslėptų neuronų sluoksnių;
- Daugiasluoksnis perceptronas – kai tarp įėjimo ir išėjimo sluoksnių yra vienas ar daugiau paslėptų neuronų sluoksnių;
- Radialinių bazinių funkcijų tinklai – kai paslėptuose sluoksniuose kaip aktyvavimo funkcija yra naudojamos radialinės bazinės funkcijos.

2.3.2. Grįžtamojo ryšio tinklai

Atgalinio ryšio neuroniniuose tinkluose signalai sklinda visomis kryptimis: į priekį, atgal ir tarp to paties sluoksnio neuronų. Dėl to atgalinį ryšį turintys neuroniniai tinklai yra daug sudėtingesni

už tiesioginio sklidimo neuroninius tinklus, bet ir efektyvesni. Tokio tipo tinklai yra dinaminiai, t. y. jų būseną nuolat kinta tol, kol pasiekiamas pusiausvyros taškas ir tinklas jame lieka, kol pasikeičia įvedimo duomenys. Pasikeitus duomenims, yra ieškomas naujas pusiausvyros taškas.

2.4. Dirbtinio neuroninio tinklo mokymas

Kiekvienam DNT modeliui reikalingi skirtingi jų apmokymo algoritmai. Trys pagrindiniai DNT apmokymo algoritmai yra šie:

- Su mokytoju (angl. *Supervised Learning*);
- Be mokytojo (angl. *Unsupervised Learning*);
- Hibridinis (angl. *Hybrid Learning*).

2.4.1. Mokymas su mokytoju

Kai tinklas yra apmokomas su mokytoju, apmokymo metu yra naudojamas išorinis mokytojas, kuris gali būti mokymui skirti apmokymo duomenys arba išorinis stebėtojas, kuris nurodo, koks turėtų būti reikiamas atsakymas. Tinklui apmokyti yra pateikiami įėjimo ir išėjimo duomenų rinkiniai. Apmokymo metu, tinklo neuronų svoriai yra vis koreguojami, kol tinklas susieja įėjimus ir išėjimus su minimalia galima paklaida. Mokymo su mokytoju algoritmams priskiriami mažiausio kvadratinio vidurkio, atgalinio sklaidymo bei radialinės bazės funkcijos algoritmai. Tinklas veiks gerai tik parinkus tinkamą mokymo aibę.

2.4.2. Mokymas be mokytoju

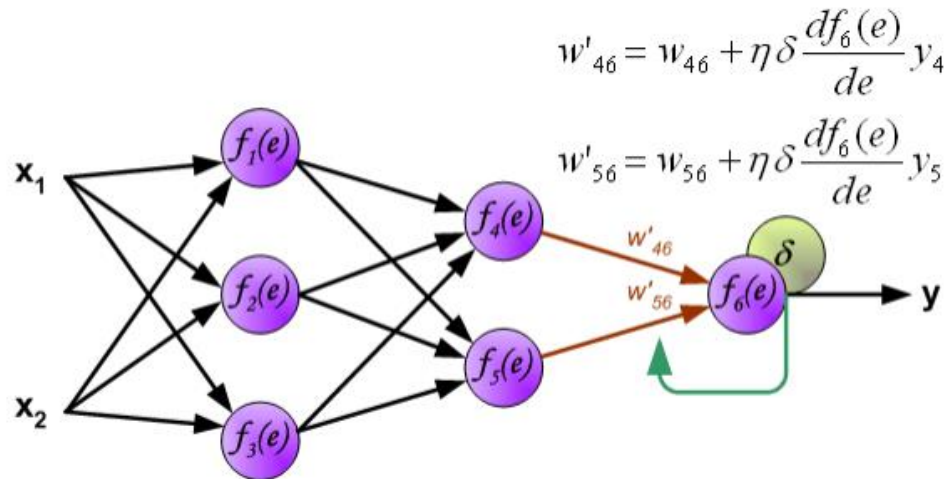
Šis tinklo apmokymo modelis neturi išorinio mokytojo. Tinklas apmokymo metu remiasi vidiniais kriterijais ir tinklo informacija, pagal tai jis pats save suderina. Tinklui yra pateikiami tik įėjimo duomenys ir sistema pati duomenyse suranda panašumus, reguliarumus ir požymius, tokiu būdu suklasifikuodama įėjimus į atitinkamas klases.

2.4.3. Hibridinis mokymas

Hibridinis mokymas jungia mokymo su mokytoju ir be mokytojo tipus: dalis tinklo svorių nustatomi pagal mokymą su mokytoju, kita dalis gaunama iš mokymo be mokytojo [10].

2.4.4. Klaidos atgalinio sklaidimo algoritmas

Klaidos sklaidimo atgal algoritmas (angl. *Error Backpropagation*) **2.4.1 pav.** yra skirtas nustatyti neuronų įėjimų svoriams. Šis algoritmas, pasinaudodamas gradientinio nusileidimo metodu, ieško klaidos funkcijos minimumo svorių erdvėje. Svorių kombinacija, kuriai esant klaidos funkcija yra minimali, laikoma uždavinio sprendiniu. Naudojant šį algoritmą, turi būti skaičiuojamas gradientas, todėl turi būti išpildyta sąlyga, jog klaidos funkcija privalo būti be trūkio taškų ir diferencijuojama. Taigi, naudojant šį algoritmą, negali būti naudojama šuolinė aktyvavimo funkcija, tačiau galima panaudoti loginę sigmoido funkciją.



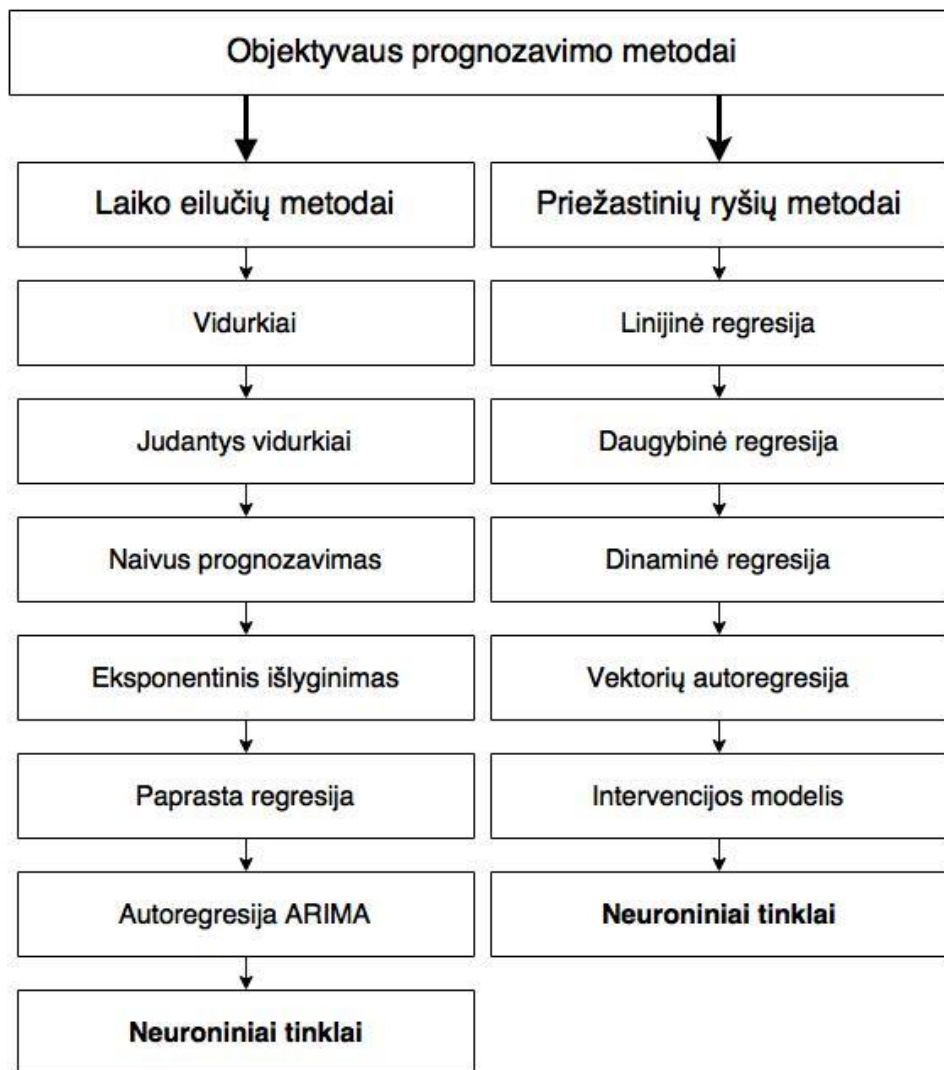
2.4.1 pav. Klaidos atgalinio sklaidimo algoritmas

Klaidos sklaidimo atgal algoritmo žingsniai:

1. Pirmiausia atsitiktinai parenkami pradiniai svoriai;
2. Tinklui paduodami pradiniai duomenys ir gaunamas atsakymas;
3. Tada nustatomos išėjimo sluoksnio neuronų klaidos;
4. Perskaičiuojami svoriai;
5. Suskaičiuojamos paslėpto sluoksnio neuronų klaidos pasiremiant toliau esančio sluoksnio neuronų klaidomis, kurios buvo suskaičiuotos ankstesniame punkte;
6. Randama klaidos išvestinė svorių atžvilgiu;
7. Pakeičiami paslėpto sluoksnio svoriai;
8. Jei dar neviršytas mokymo ciklų limitas, grįžtama į 3 punktą.

2.5. Prognozavimo metodai, paremti DNT

Prognozavimo metodų yra labai įvairių. Tinkamas metodo parinkimas priklauso nuo sprendžiamos užduoties. Šiame darbe aktualūs yra prognozavimo metodai, kurie remiasi DNT veikimo principais, kaip matyti **2.5.1 pav.** neuroniniais tinklais paremtus prognozavimo metodus apima laiko eilučių (angl. *time series*) ir priežastinių ryšių (angl. *causal*) prognozavimo metodai.



2.5.1 pav. Prognozavimo metodai

Esminiai skirtumai tarp laiko eilučių ir priežastinių prognozavimo metodų, paremtų DNT, yra tokie:

- Laiko eilučių prognozavime egzistuoja prielaida, kad duomenų generavimo procesą, kuris sukuria modelius, galima paaiškinti tik iš ankstesnių stebėjimų nuo priklausomo kintamojo.
- Priežastiniame prognozavime duomenų generavimo procesas gali būti paaiškintas pagal sąveikos priežasties (*priežasties ir pasekmės*) nepriklausomus kintamuosius.
- Laiko eilučių (angl. *time series*) ir priežastinių ryšių (angl. *causal*) metodų pranašumai ir trūkumai yra apžvelgiami toliau pateiktoje lentelėje.

Lent. 2.5.1. Prognozavimo modelių pranašumai ir trūkumai

	Pranašumai	Trūkumai
Laiko eilučių modeliai	Efektyvu trumpalaikėms prognozėms	Nereaguoja į išorinius pokyčius, todėl dėl išorinių veiksnių prognozės gali būti klaidingos
	Didelis tikslumas	Reikia labai daug istorijos duomenų, norint efektyviai prognozuoti
	Metodai paprasti ir nesudėtingi	Neefektyvu naudojant ilgalaikėms prognozėms
Priežastinių ryšių modeliai	Gali nustatyti ryšius tarp parametrų	Reikalauja išsamaus nagrinėjimo ir analizės, sugaištama daug laiko
	Gali atskleisti svarbiausius parametrus, kurie turi daugiausia įtakos baigčiai	Net jeigu regresijos modelis geras, gali būti sunku išsiaiškinti esminius kintamuosius, kurie lemia prastą prognozę
	Palaikomas „kas-jeigu“ scenarijus – tai reiškia, kad kaitaliojant parametrus tiesiogiai galima matyti, kaip keičiasi rezultatas	

Iš esmės dirbtiniai neuroniniai tinklai, kurie yra naudojami prognozavimo uždaviniams spręsti, gali būti apibrėžti kaip bet kuris metodas iš **2.5.1 pav.**, priklausomai nuo prognozuotojo sugebėjimų tinkamai suprasti bei interpretuoti turimus duomenis ir parinkti tinkamiausią metodą, kuris būtų pritaikytas prognozuojant ir atitinkamai pagal metodą modeliuojamas neuroninis tinklas.

DNT naudojimas prognozavime turi savų pranašumų, lyginant su kitais metodais. Prognozavime teisingų duomenų generavimo procesas yra sudėtingas ir sunkiai nustatomas, o būtent sporto prognozėse daugelis priklausomybių yra netiesiškos [4], todėl tai labai apsunkina prognozavimo procesą, tačiau čia ir išryškėja tikrieji DNT panaudojimo pranašumai:

1. DNT gali aproksimuoti bet kokią tiesinę arba netiesinę funkciją nepriklausomai nuo jos sudėtingumo laipsnio.
2. DNT gali išmokti bet kokią tiesinę laiko eilučių šabloną (*funkciją*).
3. DNT gali išmokti bet kokią netiesinę laiko eilučių šabloną (*funkciją*).

4. DNT gali ekstrapoliuoti bet kokiais tiesiniais ir netiesiniais laiko eilučių šablonais (*funkcijomis*), tai gali būti apibrėžiama kaip generalizacija.

2.5.1. DNT taikymas prognozavime

Prognozavime svarbiausias aspektas yra tinkamai įvertinti turimų parametrų įtaką galutinei įvykio baigčiai, iš to bus galima nuspręsti, ar prognozės bus efektyvios. Pirmiausia reikia tinkamai parinkti parametras, kurio atžvilgiu bus nagrinėjama kitų parametrų įtaka nuo pasirinkto parametro galutinei įvykio baigčiai. Šis žingsnis yra labai svarbus, nes, netinkamai pasirinkus parametras, rezultatai gali būti labai skirtingi ir tik suklaidinti, o ne duoti naudos. Parametras turėtų būti parenkamas atsižvelgiant į tai, ką tikimasi prognozuoti. Jeigu prognozuojama įvykio baigtis, tai akivaizdu, kad prognozuojamas parametras turėtų būti to įvykio baigties tikimybė, būtent šio parametro aspektu ir turėtų būti analizuojami kiti parametrai.

Jau nuo pat DNT atsiradimo skirtingi modeliai buvo taikomi skirtingiems uždaviniams spręsti, todėl prognozavimo uždaviniams taip pat ne visi modeliai tinkami. Prognozavimo uždaviniuose reikėtų koncentruotis tik į modelius be grįžtamojo ryšio ir geriau naudoti daugiasluksnį perceptroną, o ne vienasluksnį [4]. Kaip anksčiau minėta, sporto prognozėse daugelis požymių yra netiesiški, tad, atsižvelgiant į šį faktą, naudingiausia nagrinėti netiesinį prognozavimą su daugiasluksniu perceptronu.

2.6. Dirbtinių neuroninių tinklų architektūra

Duomenų gavyba yra abstrakti mokslo šaka ir nėra tikslų taisyklių, kaip duomenys turi būti analizuojami. Nors tinkamesnis DNT modelis ir gali būti parenkamas konkrečiam uždaviniui spręsti, tačiau tikslų jo struktūrą iš anksto sumodeliuoti taip, jog rezultatai būtų patys efektyviausi, praktiškai yra neįmanoma. Efektyviausia tinklo struktūra gali būti gaunama eksperimentiniu būdu.

Prieš sprendžiant kokį nors prognozavimo uždavinį, naudojant DNT, reikia gerai įsigilinti į prognozuojamos srities dalyką, kad turimi duomenys būtų suprantami ir suvokti, ką jie reiškia. Tokiu būdu tinklo modeliotojas, turėdamas žinių apie prognozuojamą sritį, geriau nutuoks, kokie parametrai gali būti svarbūs, o kurie išvis gali neturėti jokios įtakos.

Stebint rezultatus, reikia koreguoti tinklo struktūrą, kol rezultatai bus patys efektyviausi. Struktūra gali būti keičiama labai įvairiai, priklausomai nuo DNT modelio. Jeigu tinklas yra daugiasluksnis, tai gali būti didinamas arba mažinamas paslėptų sluoksnių kiekis ir neuronų kiekis juose. Tiek vienasluksniuose, tiek daugiasluksniuose tinkluose gali būti koreguojami įėjimo parametrai pridėdant arba panaikinant kai kuriuos iš jų, priklausomai nuo to, kaip keičiasi išėjimo rezultatai.

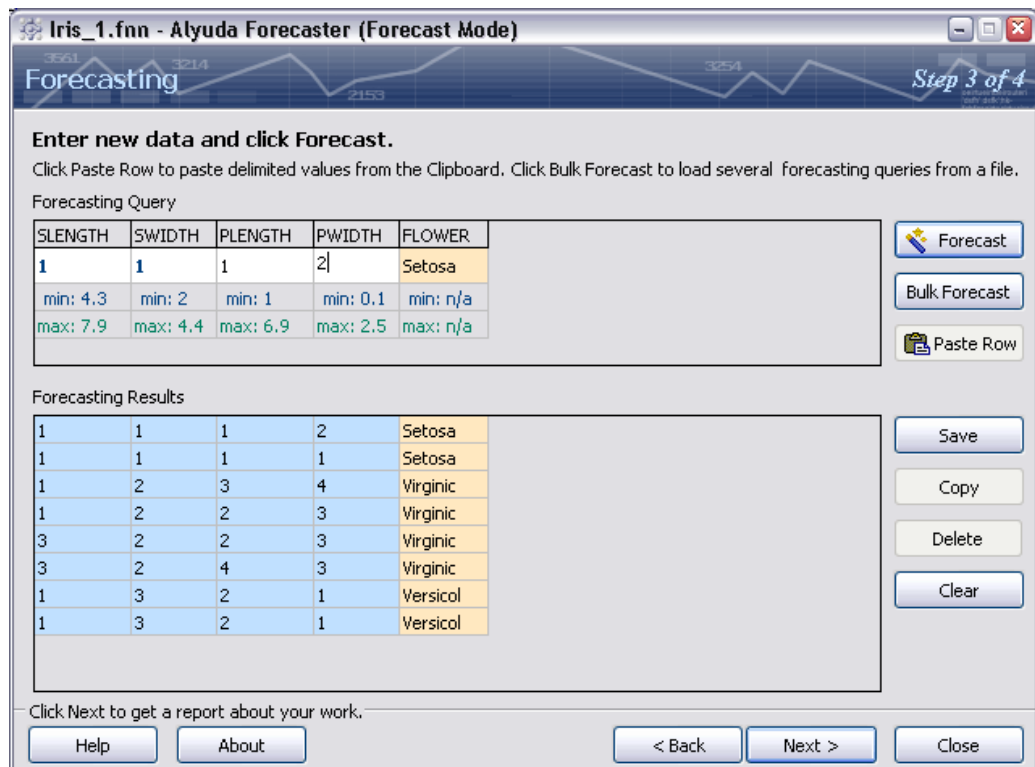
2.7. Prognozavimo PĮ naudojanti DNT

DNT su informatika yra labai glaudžiai susiję, todėl DNT šiuo metu informatikoje yra labai plačiai taikomi, yra gausybė programinės įrangos, skirtos darbui su DNT. Šiame darbe aktualūs prognozavimo uždaviniai, tad būtent šioje dalyje ir apžvelgiama PĮ, kurios kūrėjai tarp programos galimybių siūlo prognozavimo uždavinių sprendimą. Nors praktiškai su bet kuria PĮ, skirta modeliuoti DNT, galima spręsti prognozavimo uždavinius, tačiau vartotojui, kuris neturi daug žinių apie DNT veikimo principus ir modeliavimą, tai gali būti sudėtinga. Prognozavimo uždaviniuose labai svarbus duomenų įvedimas į DNT. Vartotojas turi suprasti pradinių duomenų prasmę, todėl pradiniai duomenys turi būti kruopščiai parinkti. Dėl pastarosios priežasties daugelis komercinių PĮ turi funkcijų, palengvinančių darbą su pradiniais duomenis, kad palengvintų prognozavimo procesą ir padarytų jį kaip įmanoma tikslesnį bei efektyvesnį.

2.7.1. „Aluyda Forecaster“

„Aluyda Forecaster“ [6] PĮ yra vienas iš „Aluyda Research“ kompanijos produktų, skirtų darbui su DNT. Šis produktas yra komercinis. Vartotojas gali pasirinkti vieną iš trijų vartotojo sąsajos režimų:

- Pagrindinį (*Basic*) – tai paprasčiausias režimas, kai vartotojas nori greit gauti jį dominančius rezultatus, kaip įmanoma mažiau atliekant veiksmų. Šiame režime nėra atvaizduojami jokie tarpiniai analizės rezultatai. Šis režimas suteikia galimybę analizuoti duomenis vartotojui, nieko nenusimanančiam apie DNT, todėl gauti rezultatus gali ne tik inžinieriai. Vis dėlto, kad duomenys būtų tinkamai interpretuojami, reikia specialisto žinių.
- Standartinį (*Standart*) – šiame režime vartotojas gali pats keisti nustatymus siekdamas gauti efektyvesnius rezultatus, tačiau tam jau reikia šiek tiek žinių apie DNT veikimo principus.
- Eksperto (*Expert*) – šis režimas skirtas specialistams, jis apima ne tik DNT, bet ir genetinius algoritmus, kuriuos kombinuojant gali būti išgaunami dar efektyvesni rezultatai. Šiame režime vartotojas gali koreguoti visą analizės procesą, t. y. tinklo modelio struktūrą, parametrų parinkimą, apmokymo algoritmą ir t. t.



2.7.1 pav. „Alyuda Forecaster“ PĮ programos langas

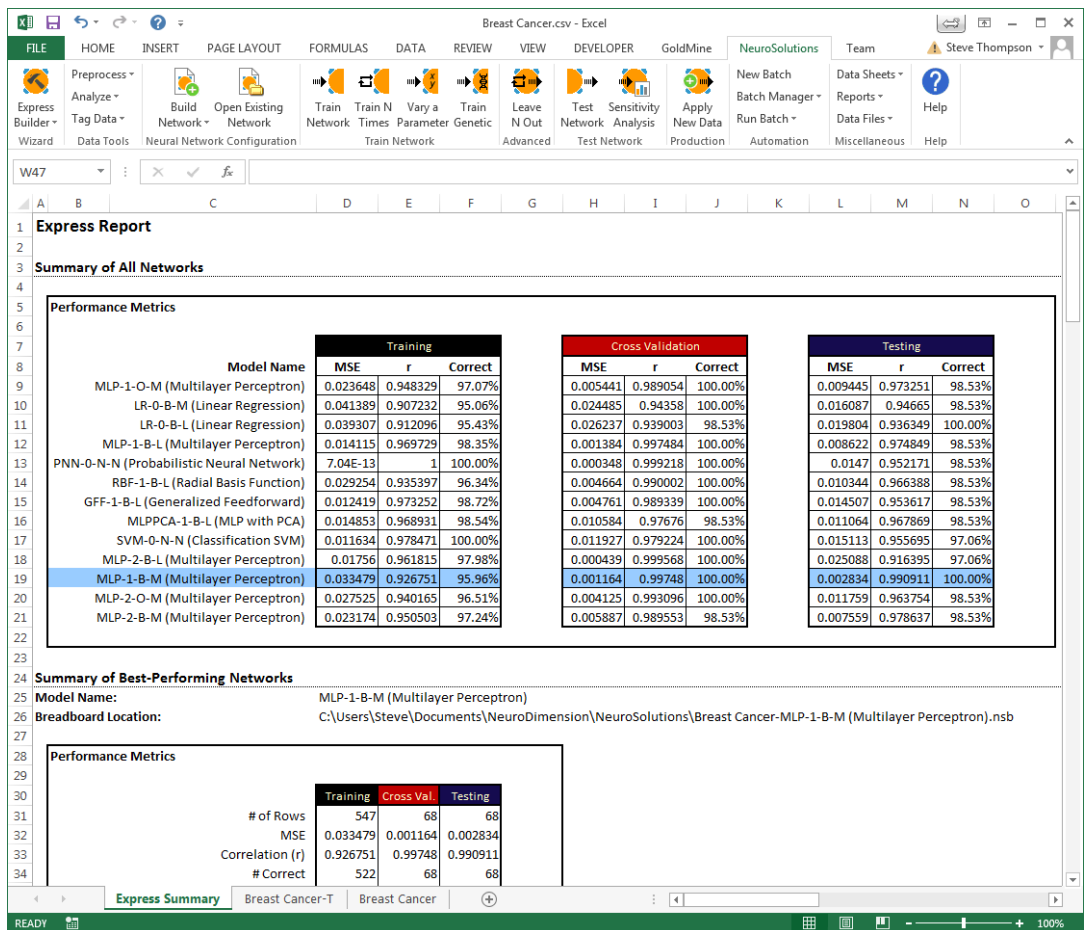
2.7.2. „NeuroSolutions“

„NeuroSolutions“ [7] yra lengvai naudojama PĮ su patogia vartotojo sąsaja. „NeuroSolutions“ programa yra viena iš programinės įrangos lyderių, skirtų darbui su dirbtiniais neuroniniais tinklais, tai – komercinis produktas. Ši programinė įranga gali būti naudojama sprendžiant labai įvairius uždavinius, susijusius su DNT, pvz.: klasterių analizė, pardavimų prognozavimas, sporto prognozės, medicininė klasifikacija ir t. t.

Darbą su šia PĮ iš esmės sudaro trys žingsniai:

1. Neuroninio tinklo analizė;
2. Neuroninio tinklo apmokymas su jūsų duomenimis;
3. Neuroninio tinklo tikslumo testavimas ir prognozių arba klasifikacijos atlikimas.

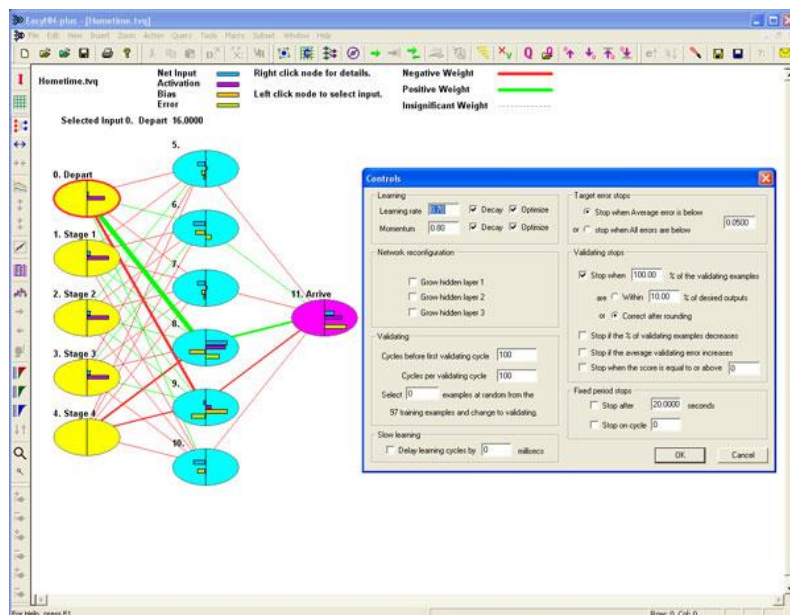
Šios programinės įrangos pranašumas tas, kad ji gali būti integruota kartu su *Microsoft Excel*, *Access*, *Visual Basic*, *MatLab* programomis. Taip pat yra galimybė iš sukurto neuroninio tinklo automatiškai generuoti C++ programinį kodą.



2.7.2 pav. „NeuroSolutions“ programa, integruota Microsoft Excel aplinkoje

2.7.3. „EasyNN“

„EasyNN“ [8] yra komercinis produktas. Pagal savo kainą šis komercinis produktas turi tikrai daug įvairių funkcijų, o ir pati grafinė sąsaja yra paprasta ir patogi, lengvai perprantama eiliniam vartotojui.

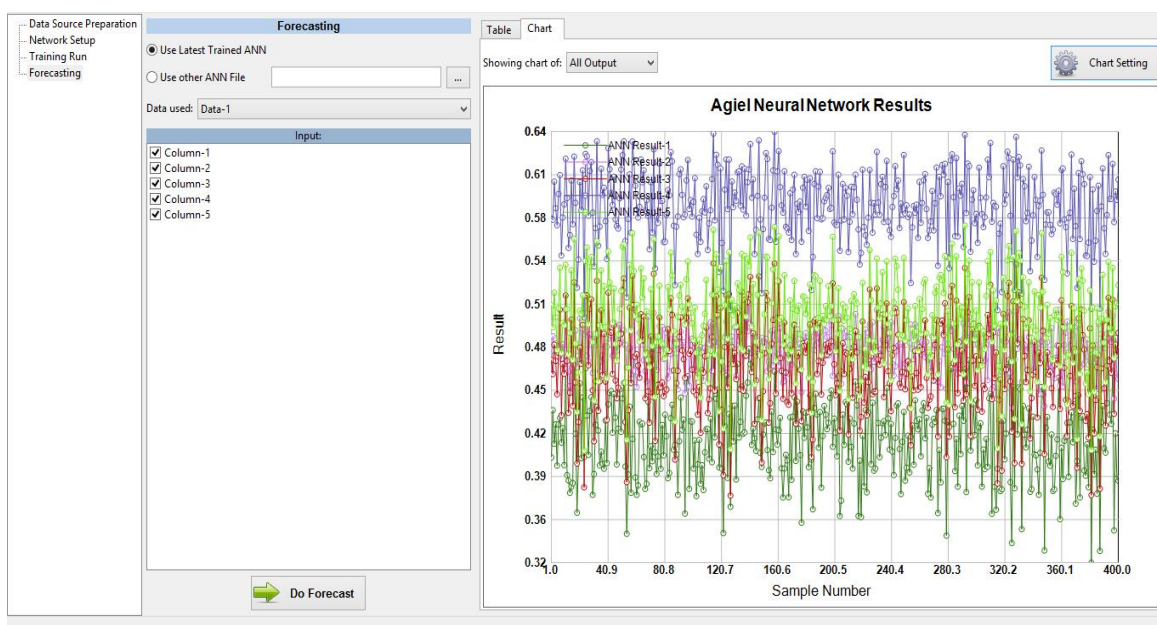


2.7.3 pav. „EasyNN“ programos langas

Su „EasyNN“ [8] programine įranga duomenų analizė yra greita ir paprasta vartotojui, neturinčiam daug žinių apie DNT. Duomenys gali būti importuojami iš tekstinių failų, nuotraukų ar dvejetainių failų. Iš importuotų duomenų automatiškai yra sukuriamas neuroninis tinklas, todėl vartotojui nereikia rūpintis duomenų paruošimu ir detalio jų analize. Praktiškas dalykas – ji turi makrokomandas, kurios gali labai palengvinti darbą su DNT.

2.7.4. „AgielNN“

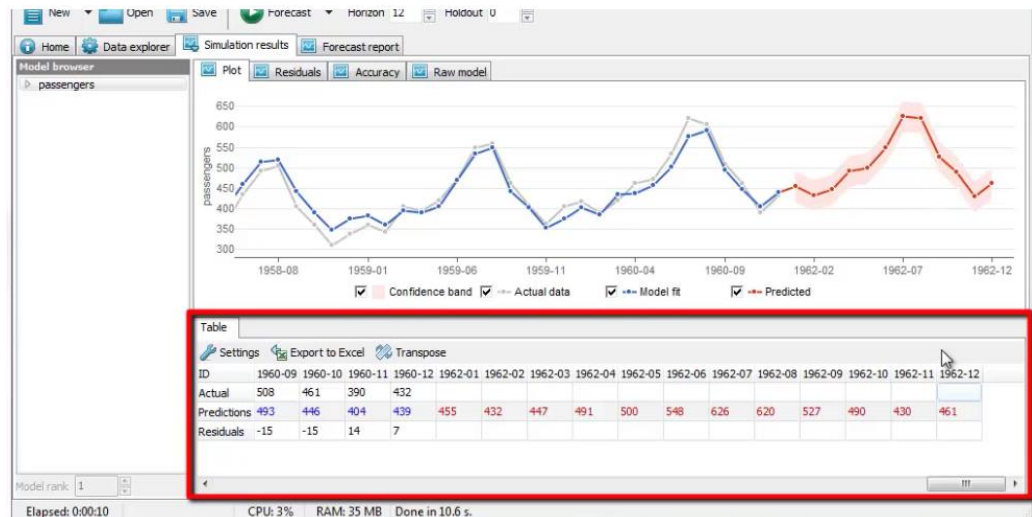
„AgielNN“ [9] – tai atvirojo kodo PI, skirta modeliuoti dirbtinius neuroninius tinklus. Labai paprasta vartotojo sąsaja yra lengvai suprantama vartotojui, neturinčiam daug žinių apie DNT. Didelis pranašumas tas, kad vartotojas gali pats koreguoti programos kodą ir prisitaikyti jį pagal savo reikmes. Programa gali būti sėkmingai naudojama spręsti daugelio sričių prognozavimo uždavinius.



2.7.4 pav. „AgielNN“ programos langas

2.7.5. „GMDH Shell“

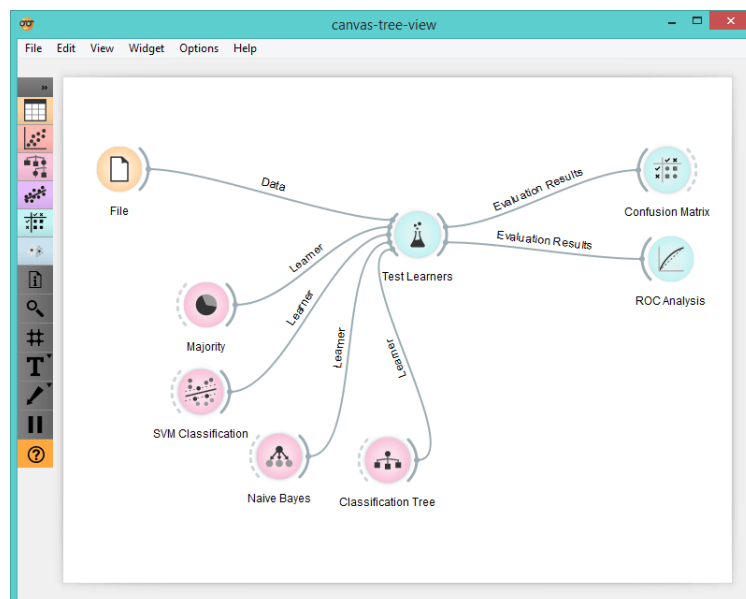
„GMDH Shell“, [11] – vienas iš plačiausiai naudojamų komercinių produktų, kuris yra skirtas įvairioms prognozėms iš turimų duomenų. Duomenys analizuojami pasitelkiant neuroninius tinklus, tačiau naudodamasis šią programine įranga vartotojas neturi galimybės pats koreguoti neuroninio tinklo ir keisti jo parametrus. Ši programinė įranga orientuota į paprastus vartotojus, kurie neturi žinių apie neuroninius tinklus ir jų veikimo principus. Dirbant su šia PI, vartotojui nereikia eksperimentuoti norint rasti tinklo modelį, su kuriuo rezultatai būtų efektyviausi. Visa tai už vartotoją atlieka programinė įranga. Lyginant su kitais tokio tipo komerciniais produktais, šio kaina yra gana didelė.



2.7.5 pav. „GMDH Shell“ programos langas

2.7.6. „Orange“

„Orange“ [12] yra atvirojo kodo programa, parašyta *Python* programavimo kalba. Programa veikia populiariausiose operacinėse sistemose: *Windows, Linux, Mac*. Didelis šios programos pranašumas yra ne tik tas, kad ji visiškai nemokama, bet ir dėl didelės funkcijų gausos, kokios įprastai negali pasiūlyti net ir komerciniai produktai. Šioje programoje yra visi įrankiai, reikalingi duomenų gavybos analizei, tad tai labai palengvina darbą, nes nereikia naudoti daug skirtingų programų. Vartotojo sąsaja yra labai paprasta ir lengvai suprantama vartotojui. Visas analizės procesas gali būti programuojamas loginiais ryšiais.



2.7.6 pav. „Orange“ programos langas

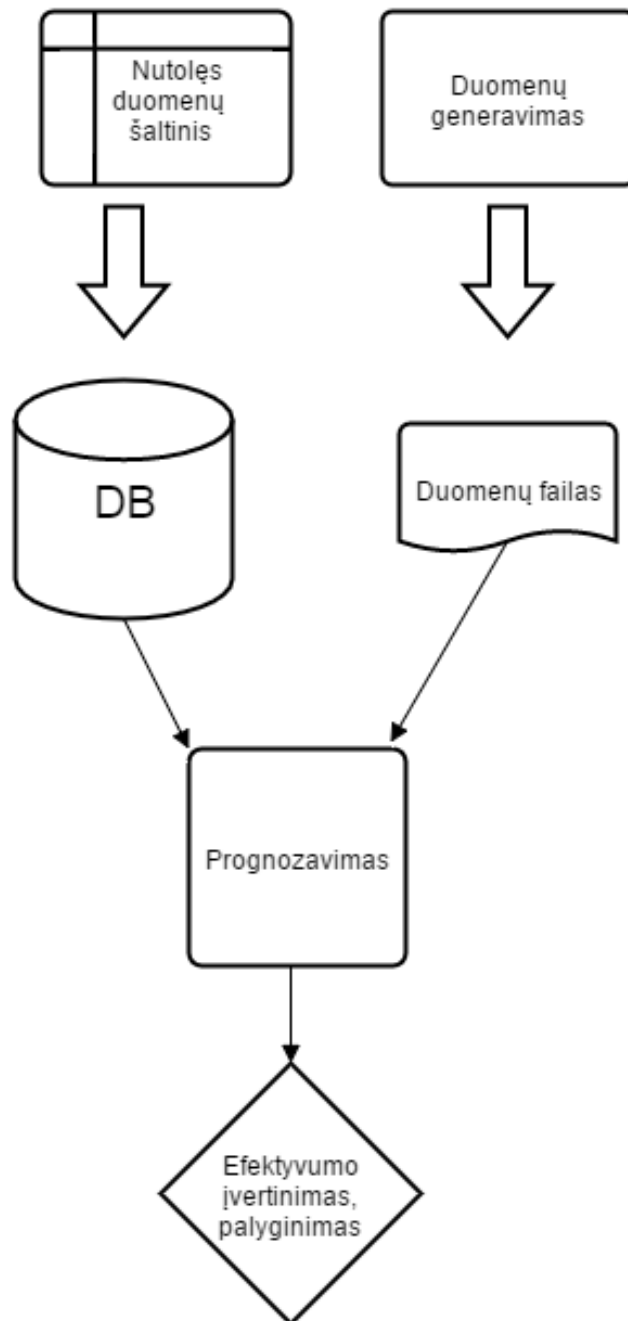
3. SPECIFIKACIJA

Šiame skyriuje apžvelgiama tyrimo eiga, kuri yra reikalinga norint gauti šio tyrimo rezultatus. Įvardijami funkciniai bei nefunkciniai reikalavimai ir darbai keliami kokybės kriterijai. Aprašoma duomenų bazės struktūra ir apžvelgiamos realizuotos programos, skirtos duomenims generuoti bei išgauti iš nutolusių šaltinių ir talpinti juos į duomenų bazę.

3.1. Projekto planas

Eksperimentui atlikti reikia surinkti realius duomenis iš nutolusių šaltinių ir juos tinkamai paruošti bei sugeneruoti pseudoduomenis su iš anksto žinomomis vertėmis, kurios lemia galutinę eksperimento baigtį. Remiantis realiais duomenimis, atliekamas prognozavimas su DNT ir matematine regresija, kad būtų įvertintas efektyvumas ir būtų galima palyginti prognozavimo rezultatus. Iš anksto sugeneruoti duomenys yra naudojami prognozavimui pasinaudojant DNT tam, kad būtų pagrįstos DNT prognozavimo galimybės ir pademonstruota, kaip parametrų svoriai gali paveikti baigtį. Projektas išskaidytas į etapus (žr. **3.1.1 pav.**), kurie yra reikalingi sėkmingai atlikti eksperimentą:

1. Suprojektuoti duomenų bazę duomenims talpinti.
2. Realizuoti programinę įrangą, duomenims išgauti iš nutolusių šaltinių.
3. Išgauti duomenis iš nutolusių šaltinių ir patalpinti juos į vietinę duomenų bazę.
4. Realizuoti programinę įrangą, skirtą generuoti pseudoduomenis pagal vartotojo nustatymus.
5. Sugeneruoti pseudoduomenis, rezultatus įrašyti į failą.
6. Pasirinkus programinę įrangą, skirtą darbui su DNT, atlikti prognozavimą su realiais ir pseudoduomenimis.
7. Pasirinkus programinę įrangą atlikti matematinę regresiją.
8. Prognozavimo rezultatams, gautiems su DNT ir matematine regresija, priskirti baigčių koeficientus ir nustatyti prognozavimo efektyvumą bei palyginti rezultatus.



3.1.1 pav. Projekto plano schema

Testavimo duomenys prieš atliekant prognozavimą turi būti tinkamai paruošti, nes atliekant testavimą yra imami kiekvieno žaidėjo statistikos vidurkiai. Kiekvieno žaidėjo statistiniai rodikliai turi būti apskaičiuojami iš ankstesnių jo susitikimų.

Realizuotos programinės įrangos, naudojamos šiame eksperimente, funkciniai ir nefunkciniai reikalavimai, aparatūrinės įrangos nustatymai, eksperimento metu keliami kokybės reikalavimai yra apibrėžti tolesniuose poskyriuose.

3.2. Funkciniai reikalavimai

Šio tyrimo realizavimui keliami šie funkciniai reikalavimai:

- Parsiunčiami duomenys turi būti talpinami *MySql* tipo duomenų bazėje.
- Duomenys gali būti eksportuojami iš duomenų bazės į tekstinį failą, kuriame duomenys būtų pateikti tinkama forma, kuri reikalinga, kad duomenys galėtų būti pateikti apdorojimui su DNT.
- Vartotojas programoje gali pridėti parametrus su norimu pavadinimu ir svoriu į parametrų sąrašą.
- Vartotojas gali ištrinti iš sąrašo pasirinktą parametą.
- Vartotojas gali sugeneruoti nurodytą susitikimų kiekį naudodamasis parametrais, esančiais sąraše.
- Sugeneruoti pseudoduomenys gali būti eksportuojami į tekstinį failą, kuriame duomenys būtų pateikti tinkama forma, kuri reikalinga, kad duomenys galėtų būti pateikti apdorojimui su DNT.

3.3. Nefunkciniai reikalavimai

Šio tyrimo realizavimui keliami šie nefunkciniai reikalavimai:

- Programinė įranga turi veikti su *Windows* ir *Linux* operacinės sistemomis.
- Jeigu parsijučiant duomenis iš nutolusio šaltinio ryšis nutrūksta arba gaunama kita klaida, dėl kurios sutrinka parsijučimo procesas, tada turi būti išsaugoma būsena, kad, pradėjus siuntimo darbą iš naujo, jis būtų tęsiamas toliau, o ne pradėdamas iš pradžių.
- Kai programa yra bet kokioje aktyvioje proceso vykdymo fazėje, vartotojas negali pradėti kitų veiksmų, kol yra nenutrauktas aktyvaus proceso vykdymas.
- Norint parsijušti susitikimų baigčių koeficientus, duomenų bazėje turi būti susitikimų baigtys, pagal kurias būtų identifikuojamas susitikimas.
- Vartotojo įvedamas parametro svoris gali būti tik skaitinė reikšmė.
- Vartotojo įvesta pradžios data duomenų eksportavimui negali būti vėlesnė už pabaigos datą.

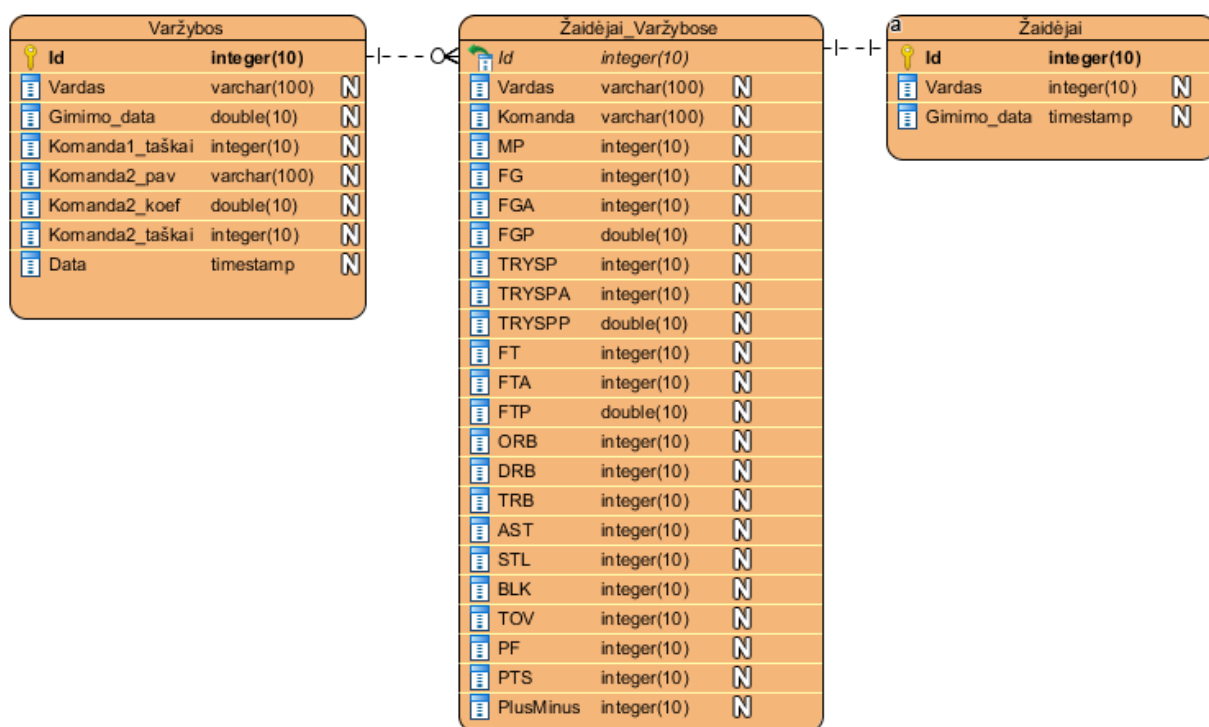
3.4. Kokybės kriterijai

Eksperimentui naudojamas duomenų kiekis yra didelis, todėl, duomenis parsunčiant, eksportuojant juos į failą ir apdorojant, reikia užtikrinti, kad resursai bus tinkamai paskirstyti ir darbas nebus pertrauktas dėl galimo atminties trūkumo arba kitų galimų klaidų. Taip pat turi būti užtikrintas duomenų nuoseklumas ir teisingi parametrų tipai. Jeigu duomenų apdorojimui su DNT naudojama trečiųjų šalių programinė įranga, kuri nėra atvirojo kodo ir vartotojas negali modifikuoti programos, tada tam, kad būtų išvengta duomenų perkrovos, galima duomenų failą skaidyti į daugiau nei vieną ir apdorojimą atlikti su keliais failais.

3.5. Projekto duomenų bazė

Eksperimente naudojamą duomenų bazę **3.5.1 pav.** sudaro trys lentelės:

- „**Varžybos**“ – lentelėje yra saugomi visi analizuojami susitikimai tarp dviejų komandų.
- „**Žaidėjai_Varžybose**“ – lentelėje yra saugomi statistiniai duomenys apie žaidėjus tam tikro susitikimo metu.
- „**Žaidėjai**“ – ši lentelė yra pagalbinių, kurioje tiesiog saugomi žaidėjų vardai ir gimimo datos.



3.5.1 pav. Duomenų bazės schema

Toliau pateiktose lentelėse matomi visų lentelių laukelių paaiškinimai.

Lent. 3.5.1. „Varžybos“ lentelės laukelių paaiškinimai

Laukelio pavadinimas	Paskirtis
Komanda1_vardas	Pirmosios susitikimo komandos pavadinimas
Komanda1_koef	Pirmosios susitikimo komandos koeficientas
Komanda1_tasškai	Pirmosios susitikimo komandos pelnytų taškų kiekis
Komanda2_vardas	Antrosios susitikimo komandos pavadinimas
Komanda2_koef	Antrosios susitikimo komandos koeficientas
Komanda2_tasškai	Antrosios susitikimo komandos pelnytų taškų kiekis
Rezultatas	Loginis kintamasis – kuri komanda nugalėjo: 1 – Pirma komanda 2 – Antra komanda
Data	Susitikimo data

Lent. 3.5.2. „Žaidėjai_Varžybose“ lentelės laukelių paaiškinimai

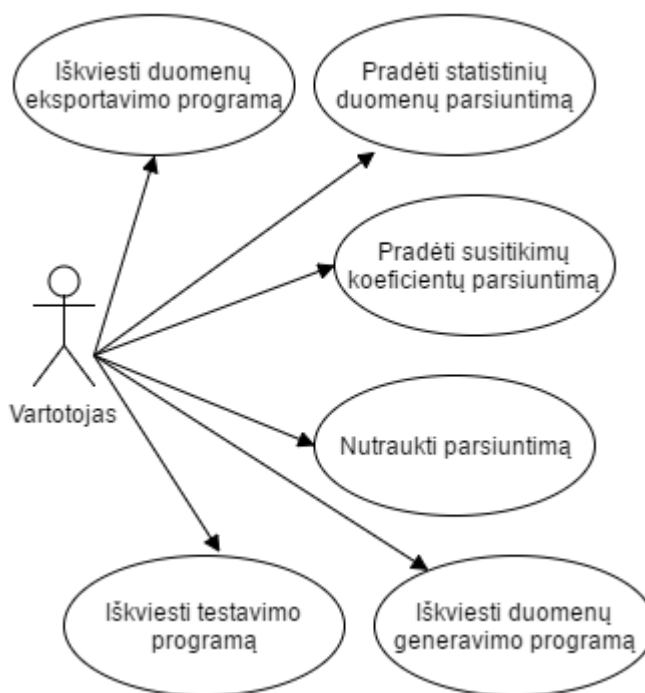
Laukelio pavadinimas	Paskirtis
Id	Susitikimo identifikacinis numeris
MP	Susitikime žaidėjo sužaistas laikas
FG	Pataikytų dvitaškių metimų kiekis
FGA	Iš viso mestų dvitaškių metimų kiekis
FGP	Pataikytų ir nepataikytų dvitaškių metimų santykis
THREEP	Pataikytų tritaškių metimų kiekis
THREPPA	Iš viso mestų tritaškių metimų kiekis
THREPPP	Pataikytų ir nepataikytų tritaškių metimų santykis
FT	Įmestų baudų metimų kiekis
FTA	Mestų baudų metimų kiekis
FTP	Pataikytų ir nepataikytų baudų metimų santykis
ORB	Puolime atkovoti kamuoliai
DRB	Gynyboje atkovoti kamuoliai
TRB	Iš viso atkovotų kamuolių kiekis
AST	Rezultatyvių perdavimų kiekis
STL	Perimtų kamuolių kiekis
BLK	Blokuotų metimų kiekis
TOV	Prarastų kamuolių kiekis
PF	Asmeninių pražangų kiekis
PTS	Per susitikimą žaidėjo pelnytų taškų kiekis
PlusMinus	Žaidėjo naudingumo koeficientas

Lent. 3.5.3. „Žaidėjai“ lentelės laukelių paaiškinimai

Laukelio pavadinimas	Paskirtis
Id	Žaidėjo identifikacinis numeris
Vardas	Žaidėjo vardas ir pavardė
Birthdate	Žaidėjo gimimo data

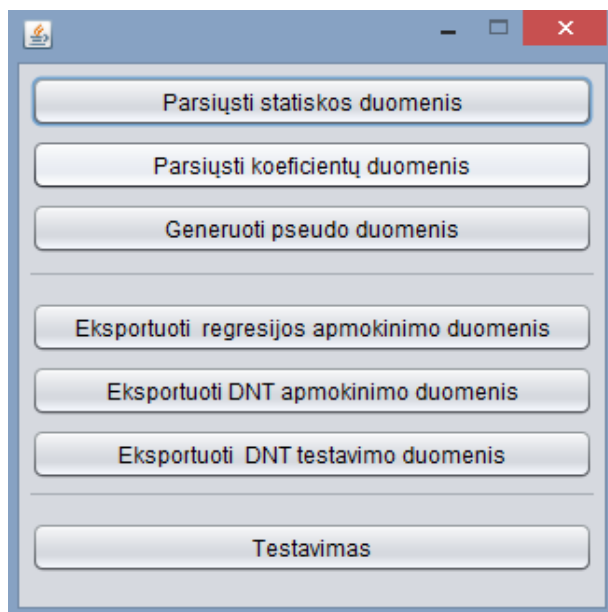
3.6. Duomenų gavyba

Eksperimentui naudojami statistiniai duomenys yra paimti <http://www.basketball-reference.com>. Susitikimų baigčių koeficientų duomenys yra paimti iš kito nutolusio šaltinio – <http://www.oddsportal.com/basketball/usa/nba/results/>. Duomenys iš šių nutolusių šaltinių yra parsiončiami ir patalpinami vietinėje duomenų bazėje, kurios schema aprašyta ankstesniame poskyryje 3.5, su specialiai šiam darbui sukurta programine įranga. Programinė įranga yra realizuota JAVA programavimo kalba. Vartotojo galimi veiksmai naudojantis šia programine įranga pavaizduoti UML panaudos atvejų schemoje 3.6.1 pav.



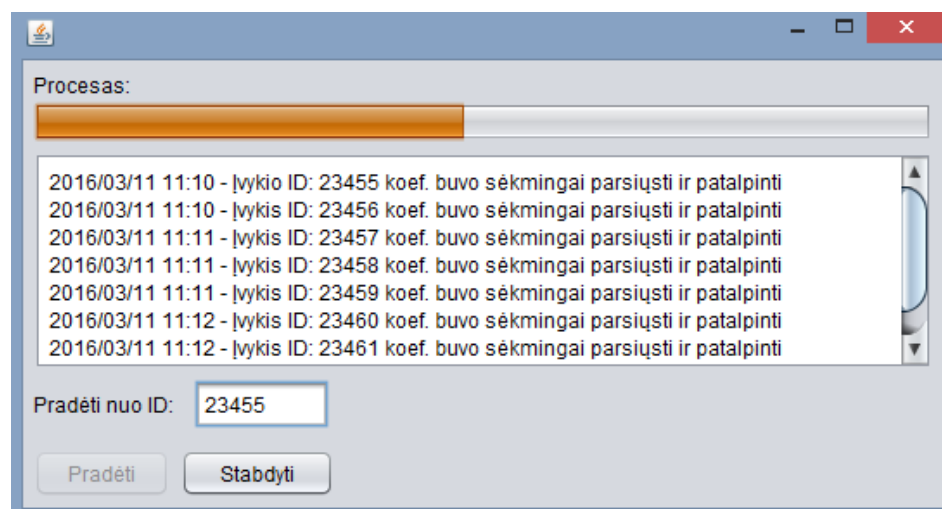
3.6.1 pav. Duomenų parsionimo programos panaudos atvejų schema

Ši programinė įranga turi vartotojo sąsają, pavaizduotą 3.6.2 pav., kurią naudodamas vartotojas mygtukų paspaudimais gali inicijuoti veiksmus iš panaudos atvejų schemos 3.6.1 pav.



3.6.2 pav. Duomenų parsisiuntimo PĮ grafinė sąsaja

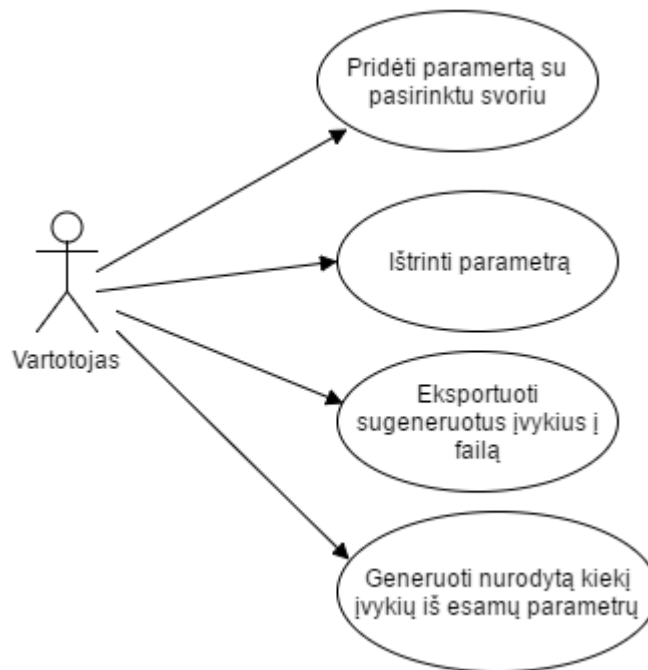
Duomenų parsisiuntimo proceso progresas ir atlikti veiksmai yra atvaizduojami programoje **3.6.3 pav.** Dėl įvairių trikdžių siuntimas gali bet kada nutrūkti, todėl vartotojas turi galimybę pradėti siuntimą nuo pasirinktojo įvykio, kad nereikėtų siųsti visų duomenų iš naujo.



3.6.3 pav. Duomenų parsisiuntimo procesas

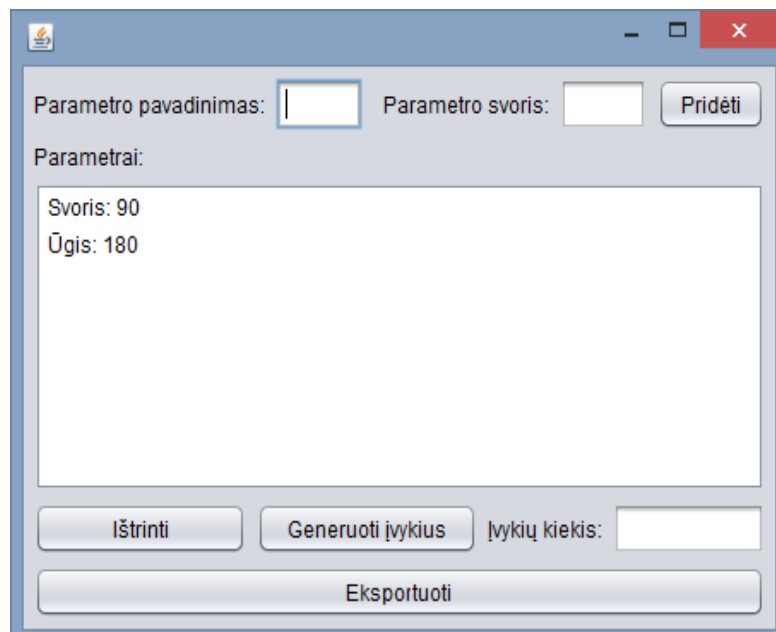
3.7. Duomenų generavimas

Specialiai šiam darbui sukurta pseudoduomenų generavimo programa, kad vartotojas galėtų susigeneruoti įvykių duomenis su savo norimomis parametų reikšmėmis. Programa, generuodama įvykių baigtis, atsižvelgia į parametų svorių įtaką ir pagal tai nustato įvykio baigtį. Šiame eksperimente naudojami pseudoduomenys yra skirti patikrinti naudojamo DNT modelio veikimą ir pademonstruoti, kaip parametrai daro įtaką baigčiai. Programos panaudos atvejai yra pavaizduoti UML panaudos atvejų schemoje **3.7.1 pav.**



3.7.1 pav. Duomenų generavimo programos panaudos atvejų schema

Vartotojo patogumui programa turi grafinę vartotojo sąsają, pavaizduotą **3.7.2 pav.** Ji suteikia galimybę įvesti duomenis į programą ir mygtukų paspaudimais inicijuoti panaudos atvejų veiksmus.

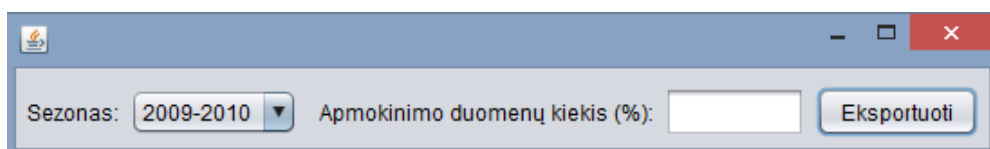


3.7.2 pav. Duomenų generavimo programos grafinė sąsaja

3.8. Duomenų eksportavimas

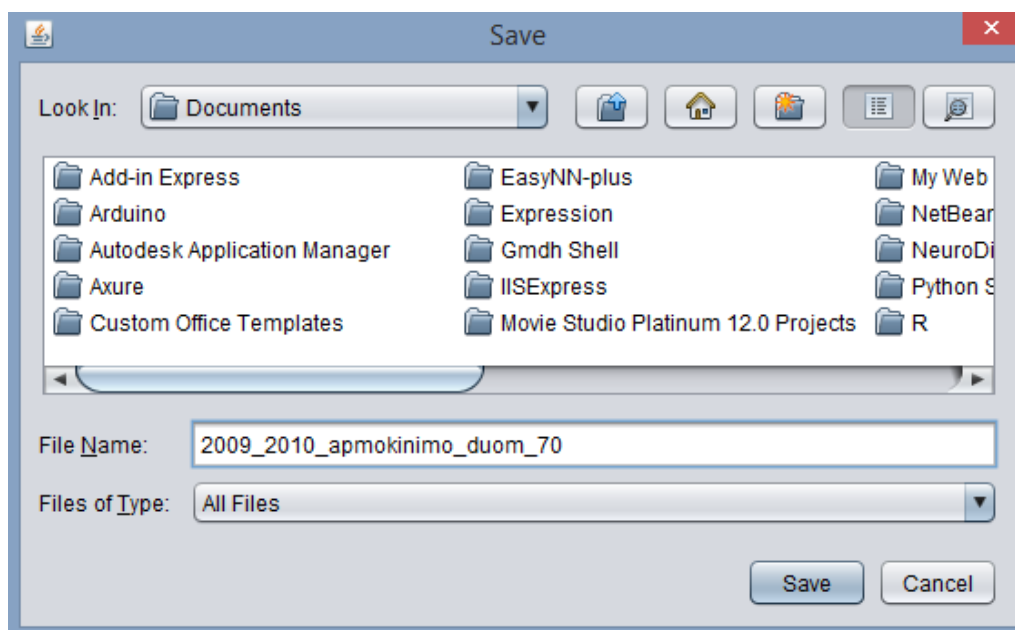
Duomenų eksportavimo programa yra skirta iš duomenų bazės eksportuoti duomenis į failą, paruoštus tinkama forma, kad būtų galima įvesti į programą darbui su DNT. Ši programa iškviečiama mygtuko „Testavimas“ paspaudimu iš pagrindinės programos, kuri aprašyta 3.6 poskyryje.

Programoje 3.8.1 pav. parenkamas konkretus sezonas arba visi sezonai, kurių duomenis norima eksportuoti. Nurodomas apmokymo duomenų kiekis, kuris bus skirtas apmokymui, o kita likusi duomenų dalis bus skirta testavimui.



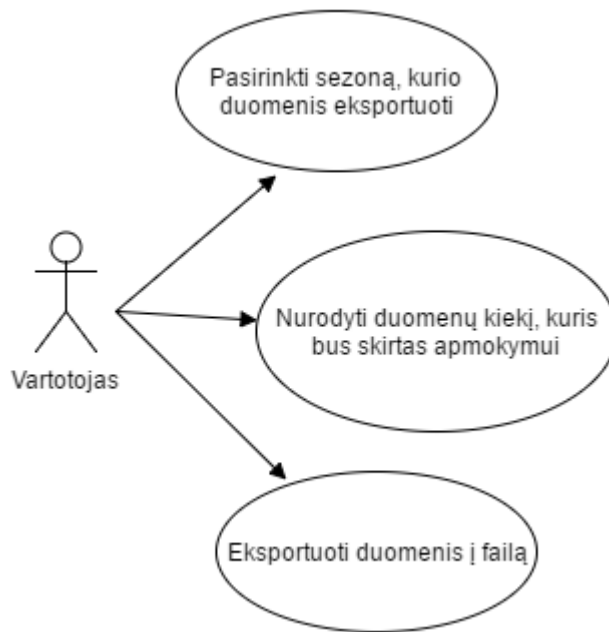
3.8.1 pav. Duomenų eksportavimo programa

Kai eksportuojamas apmokymo failas yra saugojamas 3.8.2 pav., tada automatiškai sugeneruojamas ir tekstinis duomenų failas, kurio pavadinime yra priedas `_test`. Pastarajame faile talpinami testavimo duomenys, sugeneruoti atitinkama tvarka (žr. 4.2. poskyrį), reikalinga teisingam testavimui.



3.8.2 pav. Apmokymo duomenų failo išsaugojimas

Programos galimi veiksmai pavaizduoti panaudos atvejų schemeje 3.8.3 pav.

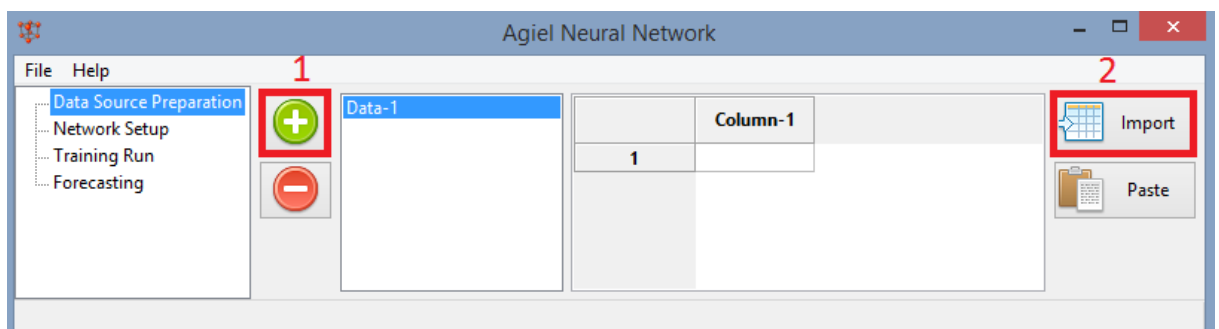


3.8.3 pav. Duomenų eksportavimo panaudos atvejų schema

3.9. PĮ DNT prognozavimui

Šio projekto prognozavimui su DNT yra naudojama „AgielNN“ programa, kuri buvo pasirinkta nagrinėjat jau esamas PĮ, skirtas darbui su DNT. Ši programa apžvelgta 2.7.4 poskyryje.

Prieš atliekant prognozavimą su DNT, programoje užkraunami apmokymo ir testavimo duomenys. Programos grafinėje sąsajoje 3.9.1 pav. paspaudus 1 numeriu pažymėtą mygtuką sukuriamas kintamasis, kuriame bus saugoma norima duomenų matrica. Viena matrica turėtų būti skirta apmokymo duomenims, o kita – testavimo. Duomenys į matricą užkraunami paspaudus 2 numeriu pažymėtą mygtuką.



3.9.1 pav. Duomenų importavimas „AgielNN“ programoje

Kai duomenys yra užkrauti, galima suformuoti DNT ir jį apmokyti. DNT tinklas suformuojamas pasirinkus „*Network Setup*“ skiltį, kuri pavaizduota 3.9.2 pav. Šioje skiltyje pažymėjus duomenų matricą su apmokymo duomenimis, yra parenkami stulpeliai, kurie bus DNT įėjimai bei išėjimai, ir parinkus paslėptų sluoksnių kiekį bei nurodžius neuronų kiekį juose, paspaudžiamas „*Build Network*“ mygtukas.

Training Testing (Optional)

Data used: Data-1

Input:

- Column-1
- Column-2
- Column-3
- Column-4
- Column-5
- Column-6
- Column-7
- Column-8
- Column-9
- Column-10

Output:

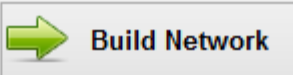
- Column-22
- Column-23
- Column-24
- Column-25

Hiddens:

Nodes of each hidden layer:

Hidden layers: 1


	Number of Node
1	10



3.9.2 pav. DNT suformavimas „AgielNN” programoje

Suformavus DNT, jis apmokomas pasirinkus skiltį „Training Run“. Šioje skiltyje, kurią galima matyti **3.9.2 pav.**, parenkami nustatymai, pagal kuriuos DNT bus apmokytas. Pasirinkus norimus tinklo nustatymus, paspaudžiamas „Do Training“ mygtukas ir prieš tai suformuotas DNT bus apmokytas ir paruoštas prognozavimui.

Maximum Epochs	1000
Maximum Error	0.0001
<input type="checkbox"/> Custom Initial Random Weights	Min <input type="text" value="-1"/> Max <input type="text" value="1"/>
Learning Algorithm	FANN_TRAIN_RPROP
Learning Rate:	0.7
Learning Momentum	0
Activation Function for Hidden	FANN_SIGMOID
Activation Function for Output	FANN_SIGMOID
Activation Steepness for Hidden	0.5
Activation Steepness for Output	0.5
Training Error Function	FANN_ERRORFUNC_TANH
Training Stop Function	FANN_STOPFUNC_MSE
Bit Failed Limit	0.35
QUICKPROP Decay	-0.0001
QUICKPROP Mu	1.75
RPROP Increase Factor	1.2
RPROP Decrease Factor	0.5
RPROP Delta Min	0
RPROP Delta Max	50
RPROP Delta Zero	0.1
SARPROP Weight Decay Shift	-6.644
SARPROP Step Error Threshold Factor	0.1
SARPROP Step Error Shift	1.385
SARPROP Temperature	0.015

 **Do Training**

3.9.3 pav. DNT apmokymas „AgielNN“ programoje

Su paruoštu DNT prognozavimą galima atlikti pasirinkus „Forecasting“ skiltį, kuri pavaizduota **3.9.4 pav.** Šioje skiltyje pasirenkamas testuojamų duomenų rinkinys ir pažymimi stulpeliai, kurių duomenys bus DNT įėjimai. Prognozavimas atliekamas paspaudus „Forecasting“ mygtuką. Išprognuozuotos reikšmės pasirodys dešinėje pusėje.

Forecasting

Use Latest Trained ANN

Use other ANN File ...

Data used: Data-1 ▼

Input:

- Column-1
- Column-2
- Column-3
- Column-4
- Column-5
- Column-6
- Column-7
- Column-8
- Column-9
- Column-10
- Column-11
- Column-12
- Column-13
- Column-14
- Column-15
- Column-16
- Column-17
- Column-18
- Column-19
- Column-20
- Column-21
- Column-22
- Column-23
- Column-24
- Column-25

➔ Do Forecast

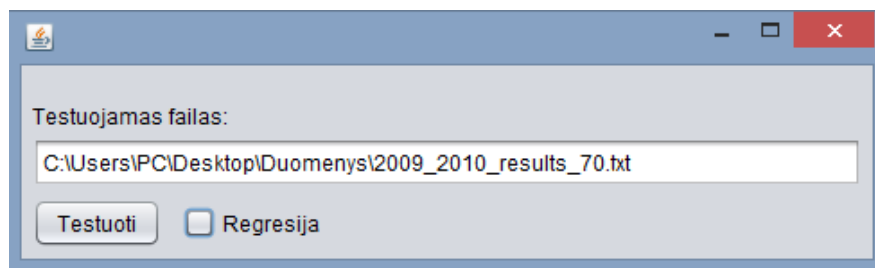
Table
Chart

ANN Result-1	
1	1.7391092225909
2	1.7715761727747
3	1.8231858271502
4	1.8050240213849
5	1.7856934507942
6	1.808910309342
7	1.7939582842349
8	1.7873414433825
9	1.7860259327044
10	1.7950786952614
11	1.7894026726886
12	1.8341544108169
13	1.8302198186036
14	1.8310839445874
15	1.8341524786466
16	1.8138391765449
17	1.7951919997062
18	1.8001759089896
19	1.7756886732787
20	1.8142194672829
21	1.8065752568326
22	1.8352874057214
23	1.7609202538874
24	1.810343979662
25	1.7777293423117
26	1.8005019007894
27	1.7836469357076
28	1.794411749728
29	1.8331857990115
30	1.7916377477737
31	1.8354577835025

3.9.4 pav. Prognozavimas su DNT „AgielNN“ programoje

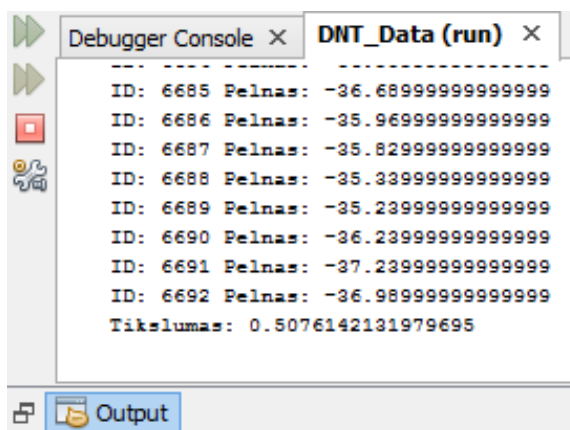
3.10. Testavimas

Išprognuotų reikšmių efektyvumo testavimas atliekamas su šiam tyrimui sukurta PĮ. Testavimo programą galima iškviesti iš 3.6 poskyryje apžvelgtos programos. Šioje programoje pasirenkamas rezultatų failas su išprognuotomis reikšmėmis, kurios buvo gautos atlikus prognozę su DNT arba regresija. Jeigu rezultatų failas yra gautas regresijos būdu, tokiu atveju ties „Regresija“ turi būti pažymėta varnelė. Jeigu tai DNT prognozės rezultatai – pažymėti nieko nereikia. Pasirinkus testuojamą failą, paspaudžiamas „Testuoti“ mygtukas ir testavimo rezultatai bus atvaizduoti programos valdymo skydelyje.



3.10.1 pav. Prognozės rezultatų testavimas

Vykstant testavimui, rezultatuose 3.10.2 pav. atspausdinamas testuojamo įvykio numeris ir bendra pelno reikšmė po to įvykio prognozės. Gale atspausdinamas viso testavimo proceso tikslumas.



3.10.2 pav. Rezultatų testavimo procesas

3.11. Naudingumo balai

Daugelyje komandinių sporto šakų naudojamas matas – „naudingumo balas“, kuris įvertina žaidėjo kokybę varžybose ir indėlį į komandos rezultatus. Kiekvienoje sporto šakoje ir lygoje šis balas yra apskaičiuojamas skirtingai. Atliekant šį tyrimą, naudoti NBA lygos žaidėjų naudingumo balai, kurie apskaičiuojami pagal toliau pateiktą formulę, kuri buvo išvesta regresijos būdu vertinant 2001–2014 m. sezonų duomenis. Šiame tiriamajame darbe neuroninio tinklo įėjimai yra būtent šie balai.

Naudingumo balo formulė [15] buvo išvesta naudojantis regresija ir remiantis individualia žaidėjo statistika bei komandos statistika – tam, kad žaidėjo statistikos vertės būtų sulygintos su komandos. Naudingumo balo formulė yra tokia:

$$\begin{aligned}
 BPM = & a * ReMPG + b * ORB\% + c * DRB\% + d * STL\% + e * BLK\% + f * AST\% - \\
 & g * USG\% * TO\% + h * USG\% * (1 - TO\%) * [2 * (TS\% - TmTS\%) + i * AST\% + j * \\
 & (3PAr - Lg3PAr) - k] + l * sqrt(AST\% * TRB\%)
 \end{aligned}$$

(1)

Tikslinis kintamasis kuriant šią regresiją yra 14 metų laikotarpio RAPM vidurkis, gautas iš 961 žaidėjų. Tam, kad ši formulė būtų efektyvi, buvo apskaičiuota kiekvieno sezono BPM ir išvestas kiekvieno žaidėjo indėlio vidurkis (įvertinant ir jo žaidimo laiką), o BPM kintamieji buvo keičiami ir skaičiavimai kartojami, kol paklaida sumažinta iki minimalios reikšmės.

Kintamieji buvo parenkami rankiniu būdu, kiekvieną sykį kartojant tuos pačius skaičiavimus, parenkant skirtingus modelius bei sąveikas tarp jų. Mažiausiai įtakos turintys statistiniai duomenys buvo išmesti siekiant surasti formulę, kurią naudojant paklaida būtų mažiausia bei reikėtų mažiausio kintamųjų skaičiaus. Išvesta daug formulių, tačiau galutine buvo parinkta ta, kuri turėjo mažiausią galimos paklaidos tikimybę.

ReMPG (žaistų minučių skaičius), ORB% (atkovotų kamuolių puolime procentas), DRB% (atkovotų kamuolių gynyboje procentas), STL% (perimtų kamuolių procentas), AST% (rezultatyvių perdavimų procentas) – tai yra paprasčiausi pagrindiniai kintamieji, kurie priklauso nuo žaidėjo turėtų galimybių rungtynėse, pavyzdžiui, ORB% yra žaidėjo atkovotų kamuolių puolime skaičius, padalytas iš žaidėjo turėtų galimybių atkovoti kamuolį.

Sudarant šią formulę buvo nemažai ginčų, ar teisinga įtraukti žaistų minučių skaičių. Akivaizdu, jog tai turi daug įtakos apskaičiuojant tikslų BPM, taigi šie duomenys buvo įtraukti į formulę nepaisant to, kad minutes reguliuoja treneris. Teigiamas (ir didelis) MPG koeficientas reiškia, jog treneriai gali matyti tai, ko nemato statistikos lentelė ir kad jie yra teisesni skirstydami minutes žaidėjams taip, kaip jiems atrodo geriausia, nors duomenys gali rodyti kitaip. Be to, žaistų minučių skaičius gali pasitarnauti priešininkams – iš duomenų galima labai lengvai suprasti, jog žaidėjas, aikštelėje praleidžiantis 5 minutes, ko gero, žaidžia prieš antros arba trečios sudėties žaidėjus, o tas, kuris aikštėje žaidžia 35 minutes, ko gero, kovoja su aukščiausio lygio oponentais. Tam, kad būtų sumažintas formulės netikslumas MPG statistika yra regresuojama – dėl to formulėje ji ir pavadinta ReMPG.

Atkovoti kamuoliai yra įprastinis krepšinio statistinis parametras. Buvo svarstoma daug variantų, kaip jis turėtų atsispindėti galutinėje formulėje, tačiau visi jie pasirodė netinkami. Taip pat svarstyta ir apie TRB% (visų atkovotų kamuolių per rungtynes procentą), tačiau buvo suskaičiuota, kad, išskaidant šią statistiką į dvi dalis, gaunama mažesnė paklaida. Blokuoti metimai bei perimtų kamuolių skaičius taip pat yra standartiniai krepšinio rodikliai. Perimtas kamuolys yra vertingesnis matas nei blokuotas metimas, ir tai atsispindi formulėje.

Rezultatyvūs perdavimai – taip pat tiesiškai priklausantis parametras, tačiau jie sąveikauja abiejose formulės dalyse (tiek tiesinėje, tiek netiesinėje), dėl to jų koeficientas šioje dalyje didelės reikšmės neturi. Klaidų procentas imamas nuo 100 komandos atakų. Visi šie paminėti parametrai yra įprasti, tačiau įmestų taškų skaičiavimas formulėje gana unikalus. Pelnyti taškai formulėje žymimi: $[USG\% * (\text{taškai, pelnyti per ataką} - \text{nustatyta bendra taškų riba atakos metu})]$.

Jei žaidėjo naudingumas yra aukštesnis už nustatytą ribą, jo buvimas aikštelėje komandai padeda; jei žemesnis, komanda nuo to kenčia. $USG\%$ yra padaugintas iš $[1 - TO\%$ (TO – klaidos)] Taip iš bendro jo naudos skaičiaus atimamas klaidų skaičius, nes klaidos nėra naudingos. Pats $USG\%$ yra žaidėjo atakų, kuriose jis metė, skaičius, padalytas iš visų komandos atakų.

- $2*TS\%$ – žaidėjo pelnyti taškai per ataką, kurioje jis metė į krepšį.
- $-2*TmTS\%$ – rodo komandos metimo vidurkį be žaidėjo, kurio BPM skaičiuojamas. Priežastis, kodėl tai skaičiuojama, paprasta – žaidėjas, mėtantis 50 % taiklumu blogoje komandoje, yra naudingesnis nei mėtantis tokiu pačiu 50% taiklumu geroje komandoje.
- $i*AST\%$ – duoda teigiamą vertę rezultatyviems perdavimams padauginus juos iš žaidėjo naudingumo – paprastai tariant, žaidėjo rezultatyvūs perdavimai yra vertesni, jei jis, be to, dar ir užbaigia atakas.
- $j*(3Par - Lg3Par)$ – duoda teigiamą vertę žaidėjo sugebėjimui susikurti erdvę, atsitverti bei atkovoti kamuolius puolime po tritaškio. Tam, kad formulė geriau veiktų istorinėje perspektyvoje, čia yra lyginamas žaidėjo ir lygos vidurkis.
- k – pastovi riba.

Galiausiai į formulę įtraukta teigiama išraiška tarp atkovotų kamuolių bei rezultatyvių perdavimų. Tai gali būti interpretuojama įvairiai: atletiškas, sąveikaujantis su krepšinio IQ, sudėjimas, sąveikaujantis su krepšinio įgūdžiais ir t. t. Ši dalis taip pat yra labai svarbi, nes, įtraukus ją į formulę, ji tapo šiek tiek visapusiškesnė. Toliau pateiktoje lentelėje pateikiami visi formulės parametrai ir jų koeficientų vertės.

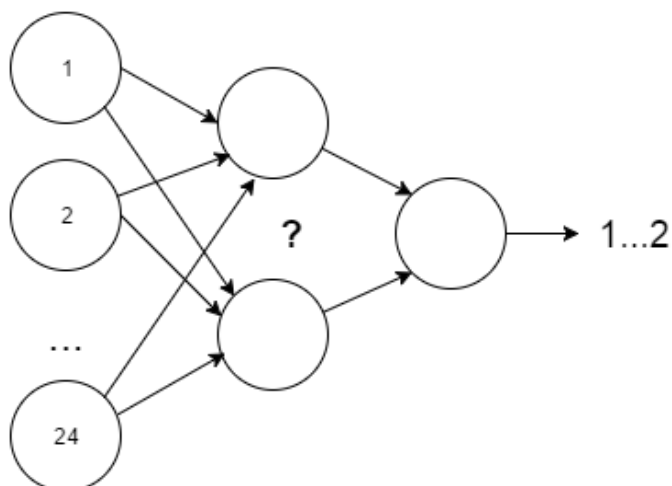
Lent. 3.11.1. Naudingumo balo formulės parametrai ir jų vertės

Koeficientas	Terminas	Vertė
a	Regr. MPG	0,123391
b	ORB%	0,119597
c	DRB%	-0,151287
d	STL%	1,255644
e	BLK%	0,531838
f	AST%	-0,305868
g	TO%*USG%	0,921292
h	Pelnyti taškai	0,711217
	USG%	
	TO%	
	S% & TmTS%	
i	AST sąveika	0,017022
j	3PAr sąveika	0,297639
k	Pelnytų taškų riba	0,213485
l	$\sqrt{\text{AST%} * \text{TRB\%}}$	0,725930

3.12. Neuroninio tinklo modelis

Remiantis esamais tyrimais ir literatūros šaltiniais [4], teigiama, jog prognozavimo uždaviniams spręsti efektyviausias daugiasluoksnio perceptrono be grįžtamojo ryšio modelis, todėl šiame darbe būtent toks ir naudojamas.

Supaprastinta šio DNT modelio schema yra pavaizduota **3.12.1 pav.** Šis modelis iš viso turi 24 įėjimus. Pagal NBA lygos reglamentą kiekvienoms rungtynėms daugiausia gali būti registruota 12 žaidėjų, todėl atitinkamai kiekvienam žaidėjui iš abiejų komandų po vieną įėjimą. Žaidėjų naudingumo balai į DNT įėjimus paduodami visada tokia pat tvarka: pirmosios komandos 12 žaidėjų naudingumo balai ir tada toliau eina kitos komandos 12 žaidėjų naudingumo balai.



3.12.1 pav. Eksperimente naudojamo DNT modelio schema

Jeigu žaidėjas tam tikrose rungtynėse nežaidė, tada apmokymo etape jo parametrų reikšmės yra lygios 0. Jeigu rungtynėms buvo registruota mažiau negu 12 žaidėjų, tokiu atveju tušti įėjimai yra užpildomi 0 reikšmėmis.

Kiek modelis turi turėti paslėptų sluoksnių ir kokį kiekį neuronų juose, reikia nustatyti eksperimentiniu keliu kaitaliojant sluoksnių ir neuronų kiekius ir stebint, kaip keičiasi rezultatai.

Šiame eksperimente naudojamo neuroninio tinklo modelis turi vieną išėjimą. Išėjimo reikšmių diapazonas yra nuo 1 iki 2. Jeigu reikšmė lygi 1, tada prognozuojama, kad nugalėjo pirmoji komanda, jeigu lygi 2 – tada antroji. Išprognozuotos reikšmės yra suapvalinamos, kad jos būtų lyginamos su realiais rezultatais, kurių baigtys yra apibrėžtos 1 arba 2 reikšme.

4. EKSPERIMENTINĖ DALIS

Eksperimento metu analizuoti 2009–2015 metų NBA sezonų duomenys. Šiame skyriuje aprašomas eksperimento atlikimas, kuris leidžia pagrįsti, ar įmanomas efektyvus krepšinio varžybų baigčių prognozavimas. Sėkmingiems tyrimo rezultatams gauti yra patikrinamas naudojamo modelio veikimas su pseudoduomenimis, tada aprašomas realių duomenų paruošimas prognozavimui su DNT ir atliekamas prognozavimas naudojant patikrintą DNT modelį bei matematinę regresiją, o tada palyginami ir apibendrinami gauti rezultatai.

4.1. Neuroninio tinklo modelio patikrinimas

Prieš prognozavimą su realiais duomenis yra atliekamas neuroninio tinklo patikrinimas, kuris leidžia nustatyti, ar šio DNT architektūros modelis su tam tikrais apmokymo parametrais geba tinkamai prognozuoti. Naudojamo DNT modelio veikimas tikrinamas su pseudoduomenimis. Pseudoduomenų parametrai – tai kompiuterio atsitiktinai sugeneruotos vertės, kurios patenka į intervalą nuo -50 iki 50. Toks intervalas pasirinktas todėl, kad naudingumo balų vertės daugiausia patenka į šį intervalą. Kiekviena vertė padauginama iš indekso, kuris yra iš anksto atsitiktinai sugeneruotas. Šis indeksas apibrėžia parametro įtaką galutinei baigčiai, kuri yra prognozuojama. Galutinė baigtis apskaičiuojama tokiu būdu:

$$A = X_1 0.54 + X_2 0.45 + X_3 0.26 + X_4 0.74 + X_5 0.91 + X_6 0.44 + X_7 0.35 + X_8 0.61 + X_9 0.54 + X_{10} 0.35 + X_{11} 0.76 + X_{12} 0.78 \quad (2)$$

$$B = X_{13} 0.32 + X_{14} 0.68 + X_{15} 0.86 + X_{16} 0.21 + X_{17} 0.54 + X_{18} 0.75 + X_{19} 0.46 + X_{20} 0.84 + X_{21} 0.44 + X_{22} 0.86 + X_{23} 0.34 + X_{24} 0.65 \quad (3)$$

$$\begin{cases} A \geq B = 1 \\ A < B = 2 \end{cases} \quad (4)$$

Pirmosios dvylika verčių yra padauginamos iš atsitiktinių indeksų ir susumuojamos tarpusavyje, tas pats atliekama ir su kitomis dvylika verčių. Abu gautieji rezultatai palyginami tarpusavyje, jeigu pirmųjų dvylikos verčių suma yra didesnė už kitų dvylikos verčių sumą arba lygi, tada galutinė baigtis yra 1, priešingu atveju baigtis bus lygi 2.

Lent. 4.1.1. DNT modelio testavimo rezultatai su pseudoduomenimis

Apmokymo duomenų kiekis	Testavimo duomenų kiekis	Paslėptų sluoksnių kiekis	Neuronų kiekis paslėptuose sluoksniuose	Prognozavimo tikslumas (procentais)
1000	500	0	0	89 %
2000	1000	0	0	86 %
5000	2500	0	0	87 %
10000	5000	0	0	96 %
1000	500	1	10	62 %
2000	1000	1	10	85 %
5000	2500	1	10	85 %
10000	5000	1	10	98 %
1000	500	2	20	62 %
2000	1000	2	20	50 %
5000	2500	2	20	70 %
10000	5000	2	20	85 %
1000	500	3	30	48 %
2000	1000	3	30	50 %
5000	2500	3	30	51 %
10000	5000	3	30	49 %

Didinant paslėptų sluoksnių skaičių, nesvarbu, kiek juose yra paslėptų neuronų, prognozuojama vertė artėja prie pastovios reikšmės **4.1.1 pav.** ir taip mažėja prognozavimo tikslumas, tai galima pastebėti iš **Lent 4.1.1.** duomenų, kai kiekvieną kartą pridėjus vieną papildomą sluoksnį prognozavimo tikslumas vis prastėja, kol pasiekiamas 50 % arba prastesnis tikslumas.

	ANN Result-1
1	1.6352469921112
2	1.6352472305298
3	1.6352473497391
4	1.635246694088
5	1.6352470517159
6	1.6352477669716
7	1.6352472305298
8	1.6352474093437
9	1.6352474689484
10	1.635247528553
11	1.6352469325066
12	1.6352474689484
13	1.6352472305298
14	1.6352470517159
15	1.6352472901344
16	1.6352480053902
17	1.6352471113205
18	1.6352476477623
19	1.6352469921112
20	1.6352472901344

4.1.1 pav. Prognozavimo rezultatų dalis „AgielINN“ programoje naudojant 3 paslėptus sluoksnius

Daugiausia prognozavimo tikslumą lemia empirinių duomenų kiekis, didėjant duomenų kiekiui, prognozavimo tikslumas taip pat didėja. Šiame poskyryje aprašytas bandymas patvirtina, kad tikrinamas DNT modelis sėkmingai (t. y. 98 %) geba prognozuoti įvykių baigtis, kai parametru vertės tiesiogiai daro įtaką baigčiai ir empirinių duomenų kiekis yra pakankamas. Didžiausias prognozavimo tikslumas buvo pasiektas naudojant vieną paslėptą sluoksnį, todėl būtent toks tinklo modelis yra naudojamas prognozavimui su realiais duomenimis šiame tyrime.

4.2. Duomenų paruošimas

Atliekant prognozavimo testavimą su realiais duomenimis, juos reikia atitinkamai paruošti. Apmokymo duomenys išlieka tokie patys, kokie yra saugomi duomenų bazėje, o testavimo duomenys turi būti verčių vidurkiai iš prieš tai jau buvusių įvykių.

Šiame tyrime duomenys yra paruošiami su specialiai šiam tikslui sukurta PĮ, kuri aptarta specifikacijos skyriuje.

Atliekant testavimą yra imituojamas realus prognozavimas, kuris būtų buvęs atliekamas anksčiau su tuo momentu turimais duomenimis. Prognozavimo scenarijus prognozuojant realius įvykius, kurie dar tik įvyks, būtų toks:

- 1) DNT yra apmokomas su turimais duomenimis.

- 2) Apskaičiuojami visų žaidėjų, kurie dalyvaus prognozuojame įvykyje, naudingumo balų vidurkiai.
- 3) Apskaičiuoti vidurkiai yra DNT įėjimai, kurie paduodami į tinklą ir išprognozuojama tikėtina baigtis.

Pagal pateiktą scenarijų yra atliekamas ir testavimas, o duomenys paruošiami tokia tvarka:

- 1) Testuojamo laikotarpio duomenų dalis (70 % arba 85 %) yra eksportuojama iš duomenų bazės į tekstinį failą. Išrenkami visų žaidėjų, kurie buvo registruoti varžybose, naudingumo balai ir įrašomi į failą. Pirmiausia įrašomas įvykio identifikacinis numeris, toliau eina pirmos komandos dvylika naudingumo balų ir tada kitos komandos dvylika reikšmių. Jeigu reikšmių mažiau nei dvylika, trūkstamos reikšmės įrašomos kaip 0. Paskutinė reikšmė eilutėje įrašoma 1 arba 2, jeigu laimėjo pirmoji komanda – 1, kitu atveju 2. Duomenys, esantys duomenų bazėje, ir tie patys duomenys, jau eksportuoti į apmokymo failą, yra pavaizduoti **4.2.1 pav.** ir **4.2.2 pav.** Šie išvesti duomenys yra naudojami tinklo apmokymui.

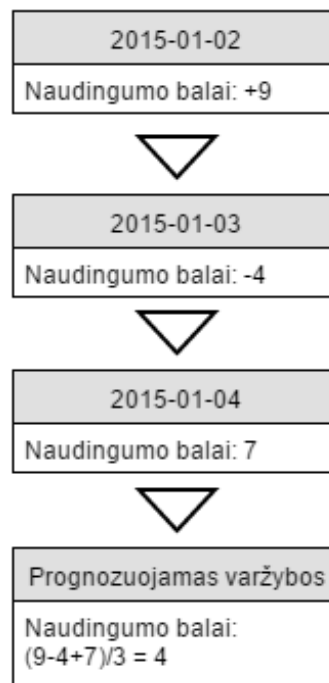
Id	Vardas	Komanda	PlusMinus
1	Ray Allen	Boston Celtics	4
1	Paul Pierce	Boston Celtics	1
1	Rajon Rondo	Boston Celtics	-3
1	Kevin Garnett	Boston Celtics	-4
1	Kendrick Perkins	Boston Celtics	-2
1	Rasheed Wallace	Boston Celtics	10
1	Marquis Daniels	Boston Celtics	11
1	Shelden Williams	Boston Celtics	7
1	Eddie House	Boston Celtics	6
1	LeBron James	Cleveland Cavaliers	0
1	Anthony Parker	Cleveland Cavaliers	3
1	Mo Williams	Cleveland Cavaliers	-10
1	Anderson Varejao	Cleveland Cavaliers	-6
1	Shaquille O'Neal	Cleveland Cavaliers	2
1	Zydrunas Ilgauskas	Cleveland Cavaliers	-2
1	Daniel Gibson	Cleveland Cavaliers	-4
1	Jamario Moon	Cleveland Cavaliers	-9
1	J.J. Hickson	Cleveland Cavaliers	-4

4.2.1 pav. Varžybose registruotų žaidėjų naudingumo balai duomenų bazėje

2009-2010_leaming_70.txt	
1	1 4 1 -3 -4 -2 10 11 7 6 0 0 0 3 -10 -6 2 -2 -4 -9 -4 0 0 0 0 1
2	2 1 1 12 15 9 16 -1 -6 0 0 0 0 -3 -9 -17 -4 7 -11 -12 -9 3 0 0 0 1
3	3 -3 5 0 14 -17 -9 -12 -3 -6 0 0 0 14 8 9 8 5 2 -1 -4 -1 3 -8 0 2
4	4 -10 -22 -8 -8 -3 -1 8 1 -9 0 0 0 -1 6 4 14 -7 15 1 9 4 0 0 0 2
5	5 -19 -16 -7 -11 -4 -4 3 12 2 0 0 0 17 23 15 12 16 -7 -5 -8 1 -9 0 0 2

4.2.2 pav. Eksportuoti apmokymo duomenys faile

- 2) Testuojamo laikotarpio duomenys yra suformuojami apskaičiuojant visų žaidėjų, kurie buvo registruoti toms varžyboms, naudingumo balų vidurkius tuo momentu, kai įvyko varžybos. Individualaus žaidėjo naudingumo balo apskaičiavimas prognozuojamoms varžyboms pavaizduotas 4.2.3 pav.



4.2.3 pav. Individualaus žaidėjo naudingumo balo apskaičiavimas testavimui

4.3. Prognozavimas naudojant DNT

Šiame poskyryje aprašomas prognozavimas su realiais duomenimis. Siekiant išlaikyti didesnę prognozavimo tikslumą, pirmiausia atliekamas prognozavimas kiekvieno sezono atskirai, o tada atliekamas paskutinio sezono prognozavimas apmokymui panaudojant visus ankstesnius sezonų duomenis. Prognozės atliekamos 2009–2015 m. sezonams išskiriant 70 % ir 85 % sezono duomenų tinklo apmokymui, kad būtų galima stebėti apmokymo duomenų kiekio poveikį prognozavimo rezultatams.

Kaip aprašyta ankstesniame 4.1 poskyryje, atlikus DNT prognozavimo efektyvumo bandymą su pseudoduomenimis, nustatyta, kad efektyviausiai prognozuoja DNT su vienu paslėptu

sluoksniu ir 10 neuronų jame, todėl, prognozavimui su realiais duomenimis naudojamas tas pats tinklo modelis.

Lent. 4.3.1. Kiekvieno sezono prognozavimo rezultatai su DNT naudojant 70 % duomenų apmokymui

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009–2010 m.	70 %	65 %
2010–2011 m.	70 %	48 %
2011–2012 m.	70 %	61 %
2012–2013 m.	70 %	39 %
2013–2014 m.	70 %	62 %
2014–2015 m.	70 %	70 %

Bendras prognozavimo tikslumas – 57 %

Lent. 4.3.2. Kiekvieno sezono prognozavimo rezultatai su DNT naudojant 85 % duomenų apmokymui

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009 – 2010 m.	85 %	59 %
2010 – 2011 m.	85 %	50 %
2011 – 2012 m.	85 %	66 %
2012 – 2013 m.	85 %	61 %
2013 – 2014 m.	85 %	58 %
2014 – 2015 m.	85 %	59 %

Bendras prognozavimo tikslumas – 59 %

Lent. 4.3.3. Prognozavimo rezultatai su DNT naudojant visų sezonų duomenis

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009 – 2015 m.	70 %	49 %
2009 – 2015 m.	85 %	67 %

Bendras prognozavimo tikslumas – 58 %

4.4. Prognozavimas naudojant regresiją

Šiame poskyryje aprašomas prognozavimas naudojant tiesinę regresiją su tais pačiais duomenimis, kaip ir 4.3 poskyryje aprašytame prognozavime su DNT. Esminis skirtumas tarp prognozavimo, kuris buvo atliktas su DNT, ir prognozavimo su regresija yra tas, kad, naudojantis DNT, galutinė baigtis buvo apibrėžta intervalu tarp 1 ir 2, o naudojant regresijos būdą ši baigtis nėra tinkama vertinti galutinį rezultatą, nes regresijoje nėra vertinami kintamųjų ryšiai nei tarp tos pačios komandos, nei su priešininkų komandos kintamaisiais, todėl toks galutinės baigties apibrėžimas, koks buvo naudotas DNT atveju, čia neturi prasmės. Regresijoje galutinė baigtis įvertinama pagal numanomą pelnytą taškų skaičių, kuriam įtakos turi žaidėjų naudingumo balai. Kuriai komandai prognozuojamas didesnis pelnytų taškų skaičius, ta komanda ir numatoma susitikimo nugalėtoja. Iš apmokymui išskirtų duomenų regresijos būdu nustatomas nepriklausomas kintamasis ir kiekvieno

įėjimo svoris – kiekvienam sezonui atskirai ir visiems sezonams kartu. Regresijos būdu apskaičiuoti indeksai ir nepriklausomi kintamieji yra pateikti tolesnėse lentelėse.

Lent. 4.4.1. Kiekvieno sezono kintamųjų reikšmės, gautos iš regresijos naudojant 70 % apmokymo duomenų

	2009–2010	2010–2011	2011–2012	2012–2013	2013–2014	2014–2015
Nepr. kint.	100,581	100,4932	95,749	98,417	99,150	100,093
X ₁	-0,032	0,008	-0,065	-0,210	0,068	0,002
X ₂	-0,090	-0,147	0,325	0,205	0,031	0,085
X ₃	-0,014	0,193	-0,477	0,010	-0,099	-0,117
X ₄	0,148	-0,191	0,243	0,044	-0,085	0,136
X ₅	0,138	0,1513	-0,079	0,070	0,141	-0,022
X ₆	-0,237	0,205	0,261	-0,115	-0,056	-0,258
X ₇	0,309	-0,089	-0,116	0,141	0,233	0,287
X ₈	-0,070	-0,072	-0,037	-0,178	-0,029	-0,171
X ₉	0,223	-0,297	-0,013	0,265	-0,044	0,112
X ₁₀	0,110	0,313	0,024	-0,245	0,123	-0,201
X ₁₁	-0,008	-0,066	0,176	0,152	-0,094	-0,016
X ₁₂	-0,181	0,076	-0,173	-0,045	-0,085	0,157

Lent. 4.4.2. Kiekvieno sezono kintamųjų reikšmės, gautos iš regresijos naudojant 85% apmokymo duomenų

	2009–2010	2010–2011	2011–2012	2012–2013	2013–2014	2014–2015
Nepr. kint.	100,782	100,325	95,465	98,442	99,9721	100,010
X ₁	0,005	0,041	-0,065	-0,161	0,029	0,003
X ₂	-0,153	-0,181	0,305	0,104	0,005	0,116
X ₃	0,030	0,146	-0,293	0,083	-0,054	-0,148
X ₄	-0,004	-0,151	0,087	-0,014	-0,056	0,125
X ₅	0,123	0,136	-0,061	0,088	0,111	-0,013
X ₆	-0,161	0,229	0,229	-0,067	0,001	-0,185
X ₇	0,219	-0,086	-0,126	0,090	0,181	0,178
X ₈	-0,045	-0,050	-0,003	-0,162	-0,019	-0,179
X ₉	0,187	-0,241	-0,002	0,253	-0,107	0,208
X ₁₀	0,073	0,260	0,018	-0,151	0,190	-0,290
X ₁₁	-0,019	-0,097	0,088	0,045	-0,044	0,097
X ₁₂	-0,125	0,078	-0,126	-0,022	-0,102	0,086

Lent. 4.4.3. Visų sezonų kintamųjų reikšmės, gautos iš regresijos naudojant 70 % ir 85 % apmokymo duomenų

	2009–2015 (70 %)	2009–2015 (85 %)
Nepr. kint.	98,731	98,5202
X ₁	-0,034	-0,016
X ₂	0,030	0,011
X ₃	0,002	-0,005
X ₄	-0,026	-0,028
X ₅	0,078	0,076
X ₆	-0,019	-0,020
X ₇	0,097	0,083
X ₈	-0,070	-0,091
X ₉	0,023	0,040
X ₁₀	0,082	0,085
X ₁₁	-0,037	-0,024
X ₁₂	-0,044	-0,036

Turint regresijos indeksus ir nepriklausomus kintamuosius, prognozuojamas komandos pelnytų taškų skaičius yra apskaičiuojamas pagal šią formulę:

$$\begin{aligned}
 \text{Taškų kiekis} = & \text{Nep. kint.} + X_1 \text{indeksas}_1 + X_2 \text{indeksas}_2 + X_3 \text{indeksas}_3 + X_4 \text{indeksas}_4 \\
 & + X_5 \text{indeksas}_5 + X_6 \text{indeksas}_6 + X_7 \text{indeksas}_7 + X_8 \text{indeksas}_8 + X_9 \text{indeksas}_9 \\
 & + X_{10} \text{indeksas}_{10} + X_{11} \text{indeksas}_{11} + X_{12} \text{indeksas}_{12}
 \end{aligned}
 \tag{5}$$

Kurios komandos išprognozuotas taškų skaičius, priklausantis nuo komandos narių naudingumo balų įtakos, yra didesnis, ta komanda ir numatoma nugalėtoja.

Prognozavimas atliktas tokia pat tvarka kaip ir su DNT. Toliau lentelėse yra pateikti prognozavimo rezultatai.

Lent. 4.4.4. Kiekvieno sezono prognozavimo rezultatai su regresija naudojant 70 % duomenų apmokymui

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009–2010 m.	70 %	65 %
2010–2011 m.	70 %	54 %
2011–2012 m.	70 %	51 %
2012–2013 m.	70 %	55 %
2013–2014 m.	70 %	54 %
2014–2015 m.	70 %	51 %

Bendras prognozavimo tikslumas – 55 %

Lent. 4.4.5. Kiekvieno sezono prognozavimo rezultatai su regresija naudojant 85 % duomenų apmokymui

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009–2010 m.	85 %	60 %
2010–2011 m.	85 %	48 %
2011–2012 m.	85 %	54 %
2012–2013 m.	85 %	59 %
2013–2014 m.	85 %	55 %
2014–2015 m.	85 %	51 %

Bendras prognozavimo tikslumas – 54%

Lent. 4.4.6. Prognozavimo rezultatai su regresija naudojant visų sezonų duomenis

Sezonas	Apmokymo duomenų kiekis	Prognozavimo tikslumas
2009–2015 m.	70 %	56 %
2009–2015 m.	85 %	57 %

Bendras prognozavimo tikslumas – 56%

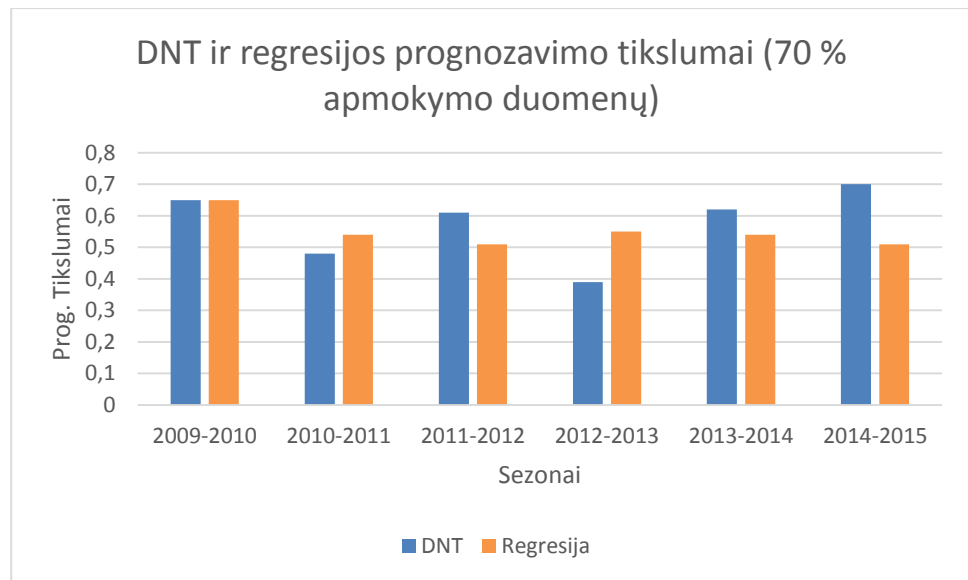
4.5. Rezultatų analizė

Prognozuojant su DNT, kai buvo naudojama 70 % ir 85 % kiekvieno sezono duomenų apmokymui, taip pat ir su visų sezonų duomenimis, rezultato atspėjimo santykis buvo didesnis negu naudojant regresiją. Toliau pateiktoje lentelėje nurodyti visų prognozavimo rezultatų tikslumai ir jų santykis procentais. Padidinus duomenų apmokymo kiekį, DNT prognozavimo rezultatai pagerėjo, o regresijos rezultatai pablogėjo, nes regresija nėra tokia jautri duomenų kiekiui kaip DNT ir jos tikslumas mažiau priklauso nuo duomenų kiekio. Prognozuojant, kai buvo naudojami visų sezonų duomenys, regresijos rezultatai pagerėjo, o DNT pablogėjo – skirtumas tarp prognozių tikslumų sumažėjo, o tai rodo, kad didėjant duomenų kiekiui tikslumai susilygintų.

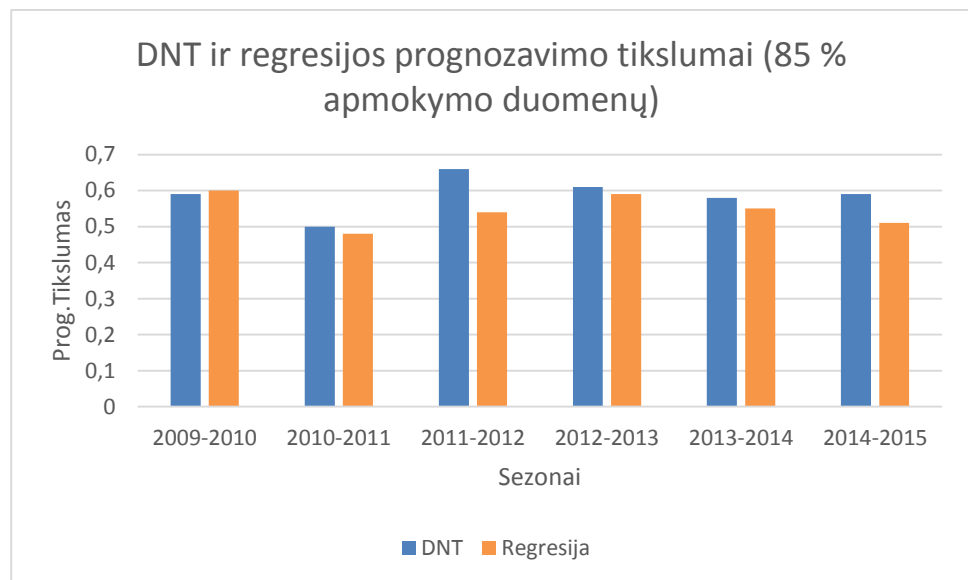
Lent. 4.5.1. Prognozavimo rezultatų palyginimas

	Kiekvieno sezono atskirai (70 % apmokymo duomenų)	Kiekvieno sezono atskirai (80 % apmokymo duomenų)	Visų sezonų bendrai (70 % ir 85 % apmokymo duomenų)
DNT	57 %	59 %	58 %
Regresija	55 %	55 %	56 %
Skirtumas	2 %	4 %	2 %

Kiekvieno sezono prognozavimo tikslumai naudojant 70 % ir 85 % apmokymo duomenų yra palyginti toliau pateiktuose grafikuose.



4.5.1 pav. DNT ir regresijos prognozavimo tikslumo palyginimas (70 % apmokymo duomenų)



4.5.2 pav. DNT ir regresijos prognozavimo tikslumo palyginimas (85 % apmokymo duomenų)

Kaip galima pastebėti iš anksčiau pateiktų grafikų, regresija prognozavimo tikslumu pranoko DNT tik vienintelį sezoną ir labai menkai. Bendras prognozavimo tikslumas yra didesnis naudojant DNT.

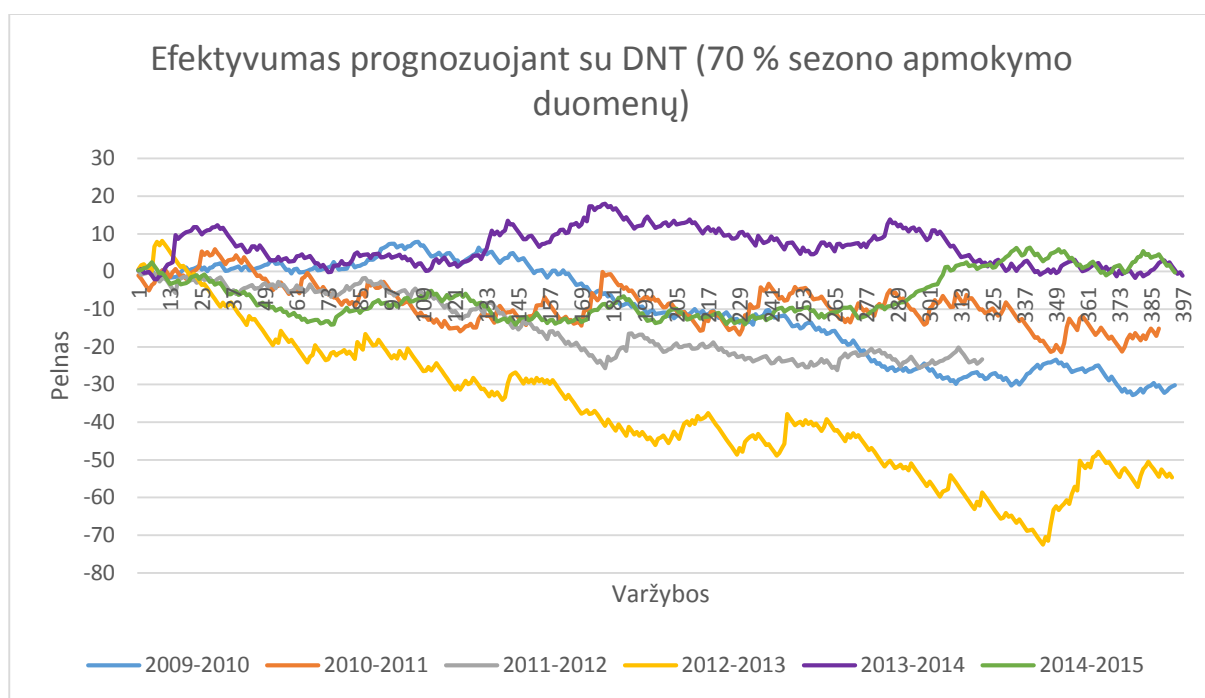
Prognozavimo efektyvumui įvertinti yra naudojami lažybų bendrovių koeficientai, kurie buvo pasiūlyti už įvykių baigtis. Lažybų bendrovės apskaičiuoja tikimybes, kokie yra vienos ir kitos komandos šansai laimėti. Pagal apskaičiuotą tikimybę nustatomi lažybų koeficientai, pagal kuriuos yra apskaičiuojama laimėjimo suma atspėjus rezultatą. Koeficientai lažybų bendrovių yra šiek tiek padidinami, kad ilguoju statymo laikotarpiu turėtų pranašumą pelno atžvilgiu prieš lažybų statytoją.

Efektyvumas įvertinamas tokia tvarka:

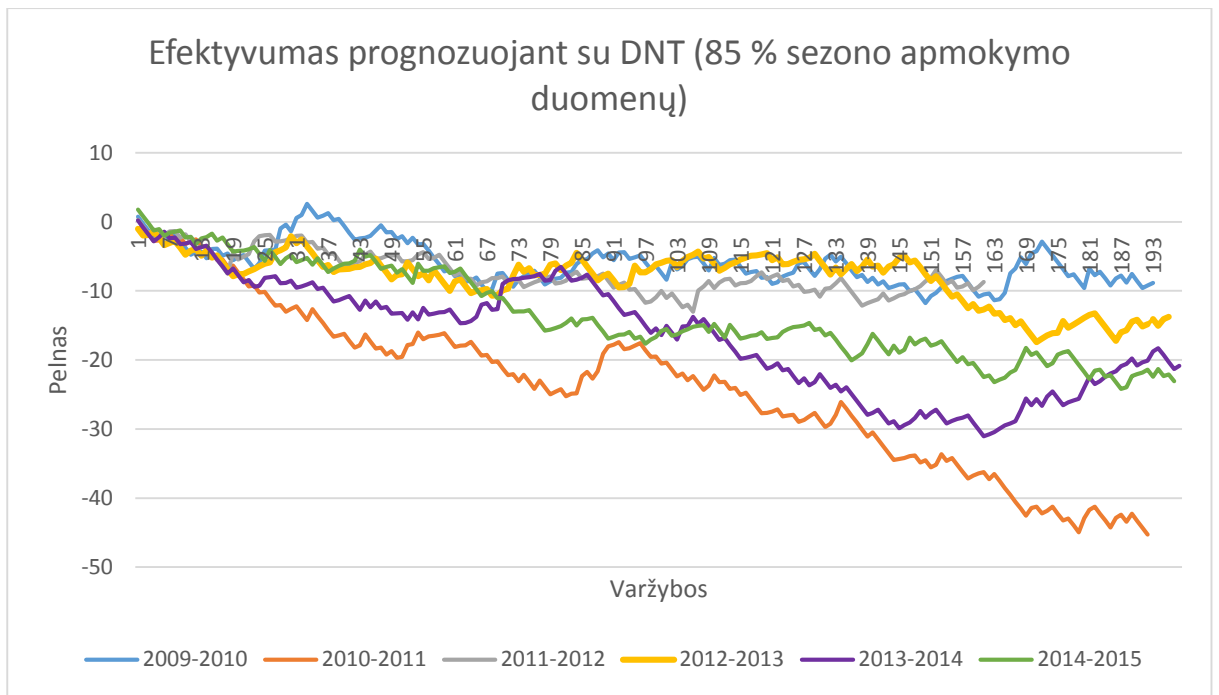
Jeigu išprognuotas rezultatas buvo teisingas, tada prie bendros pelno sumos pridedamas pelnas gautas iš atspėtų varžybų, kurių rezultatas buvo prognozuojamas. Vienų atspėtų varžybų pelnas apskaičiuojamas tokiu būdu:

$$Pelnas = (statoma\ suma * koeficientas) - statoma\ suma \quad (6)$$

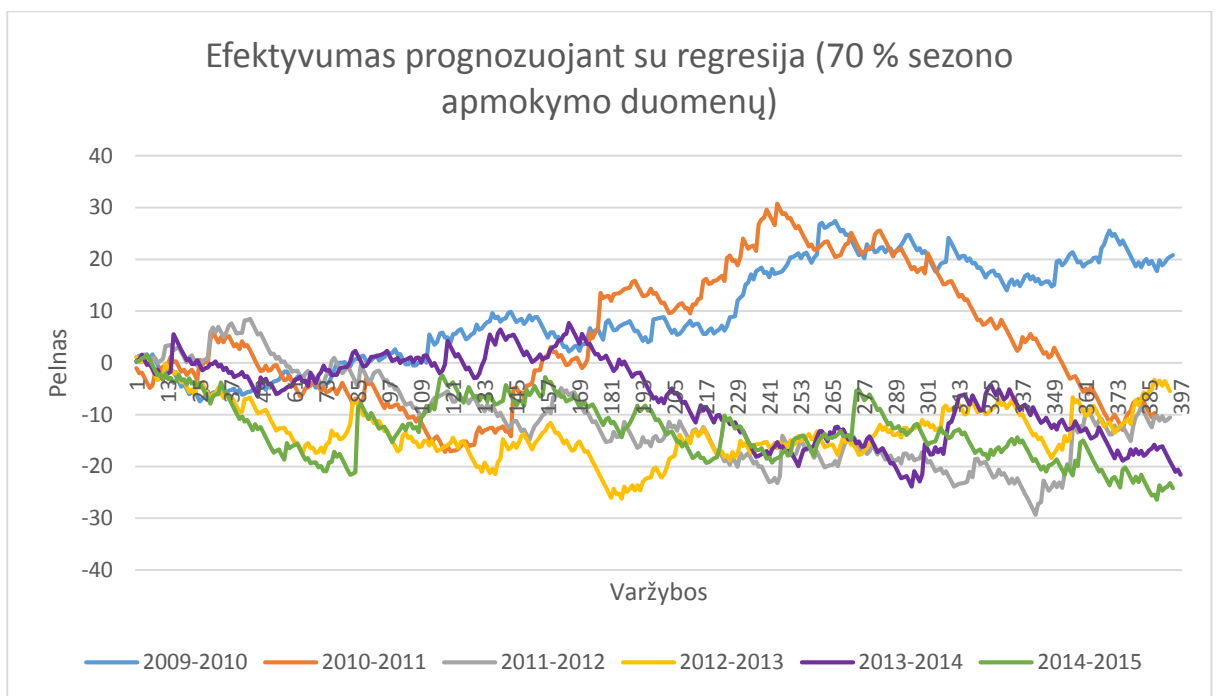
Jeigu išprognuotas rezultatas yra neteisingas, tada iš bendros pelno sumos atimama statoma suma. Kuo funkcijos kreivė labiau auganti, tuo rezultatas yra efektyvesnis. Tokiu būdu įvertinti prognozavimo rezultatai, kurie buvo gauti pasitelkus DNT ir regresiją, grafiškai pateikti toliau.



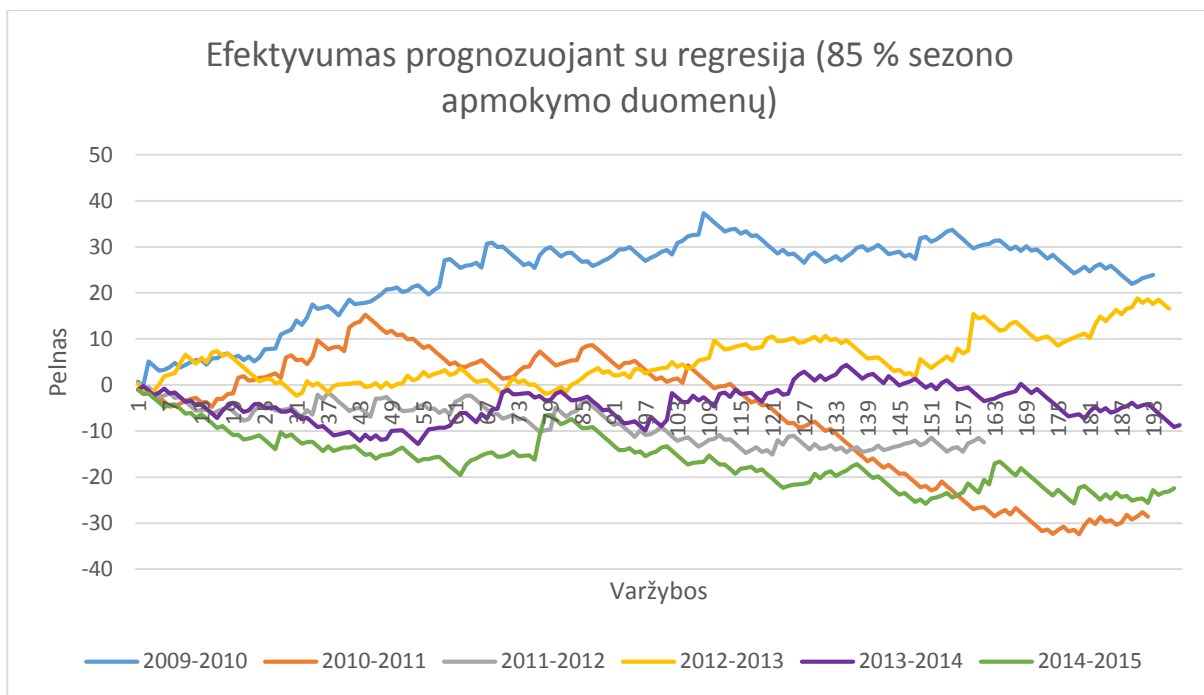
4.5.3 pav. Efektyvumo įvertinimas prognozuojant su DNT (70 % apmokymo duomenų)



4.5.4 pav. Efektyvumo įvertinimas prognozuojant su DNT (85 % apmokymo duomenų)

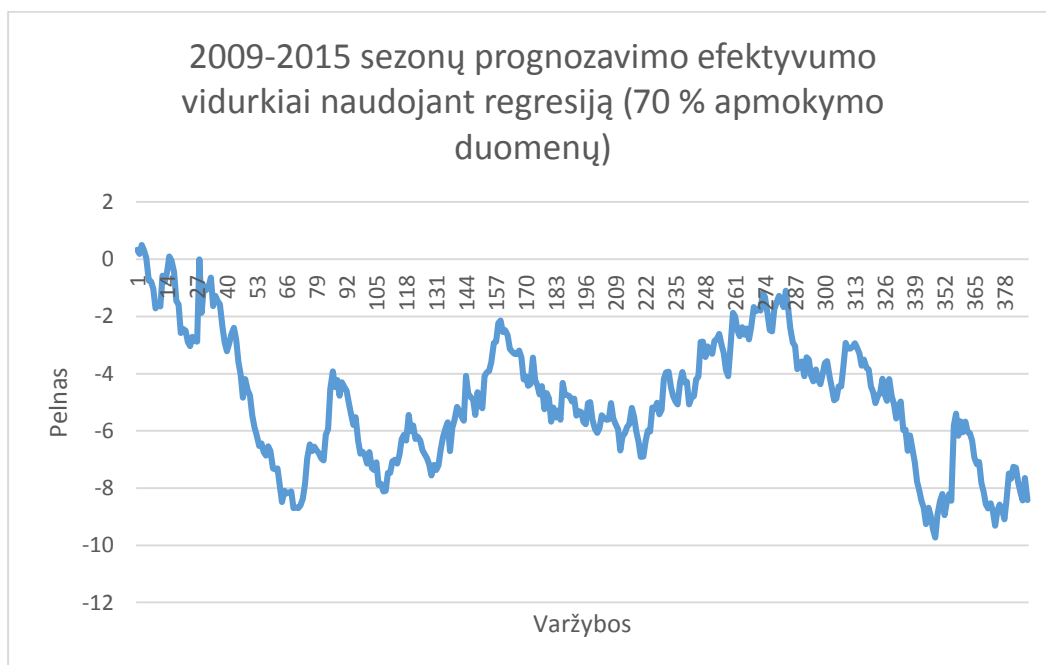


4.5.5 pav. Efektyvumo įvertinimas prognozuojant su regresija (70 % apmokymo duomenų)

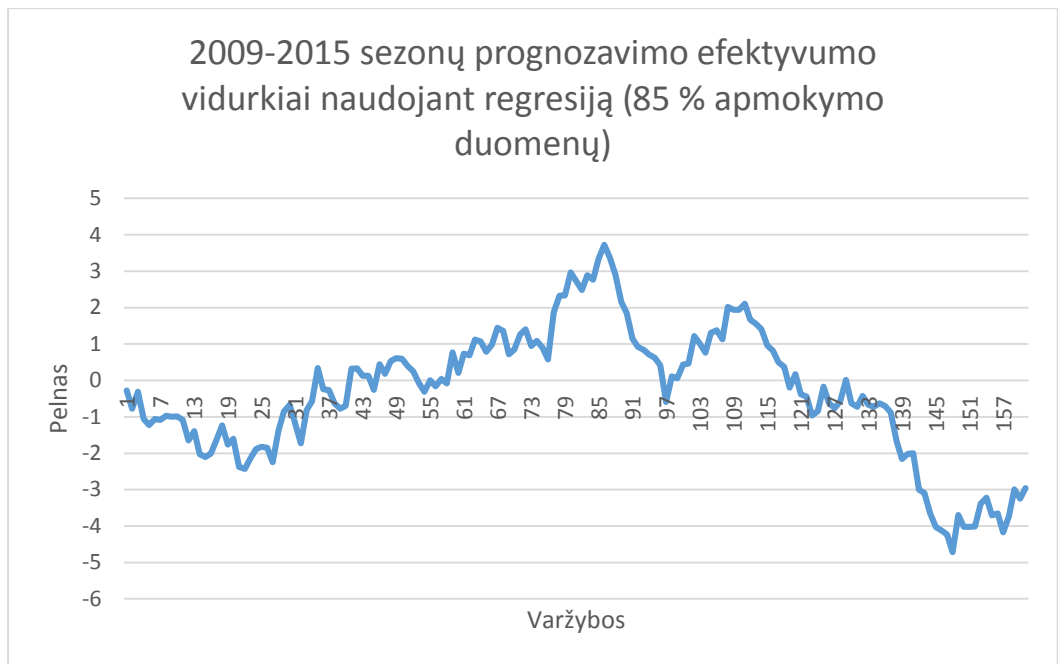


4.5.6 pav. Efektyvumo įvertinimas prognozuojant su regresija (85 % apmokymo duomenų)

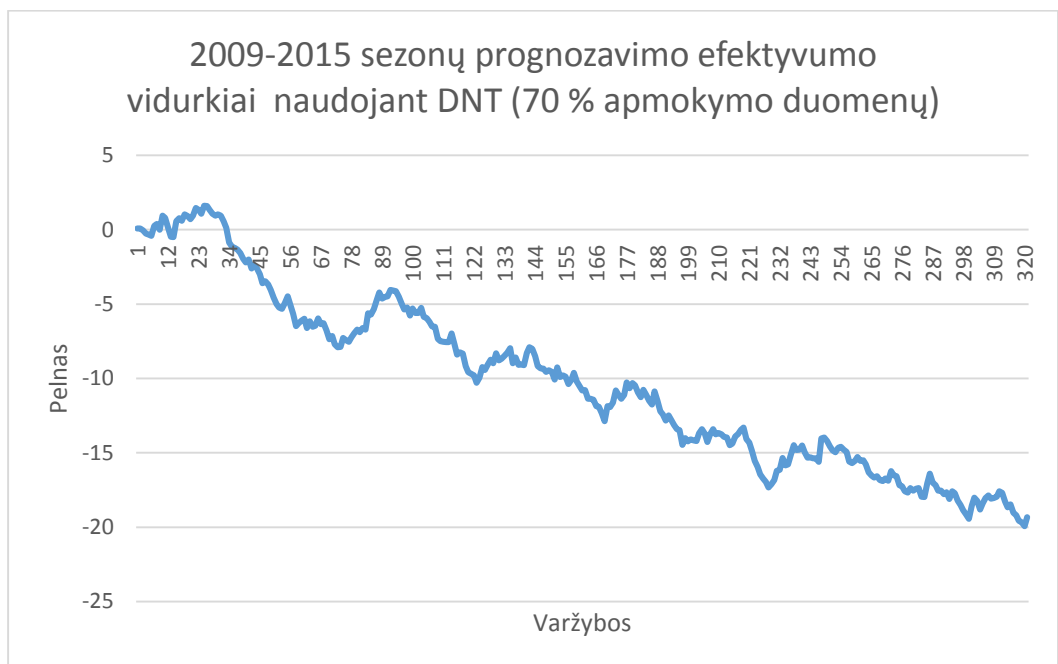
Išvedus visų sezonų rezultatų vidurkį, gaunama funkcijos kreivė, kuri atspindi visų sezonų efektyvumą, ir galima pastebėti, ar ilguoju laikotarpiu funkcija yra didėjanti ar mažėjanti, bei įvertinti, ar prognozės yra efektyvios.



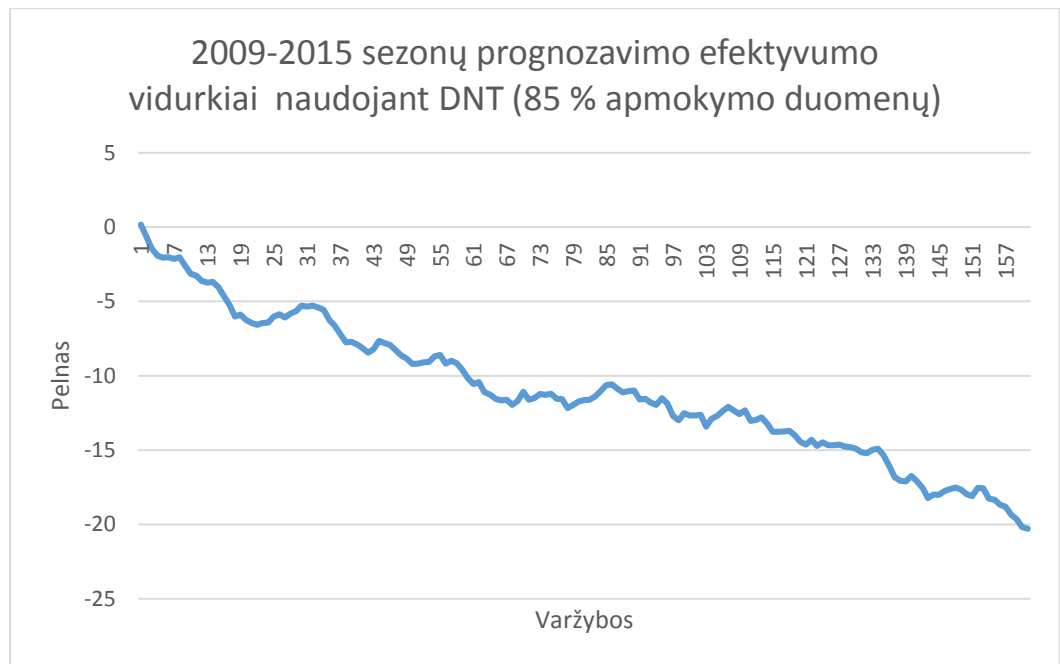
4.5.7 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant regresiją (70 % apmokymo duomenų)



4.5.8 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant regresiją (85 % apmokymo duomenų)



4.5.9 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant DNT (70 % apmokymo duomenų)

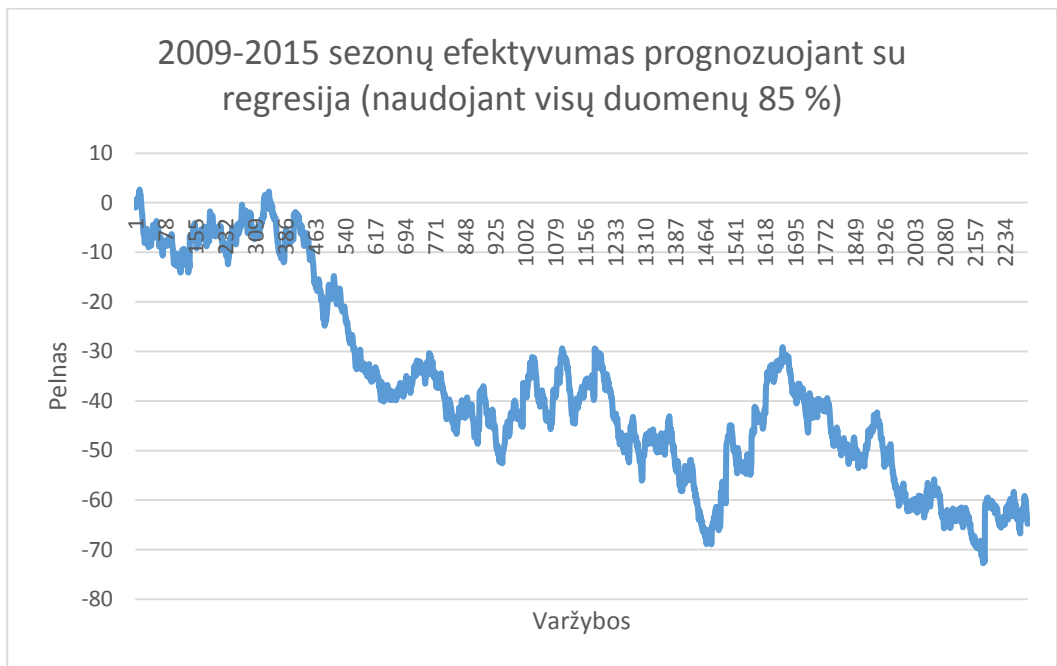


4.5.10 pav. 2009–2015 sezonų bendras efektyvumo įvertinimas naudojant DNT (85 % apmokymo duomenų)

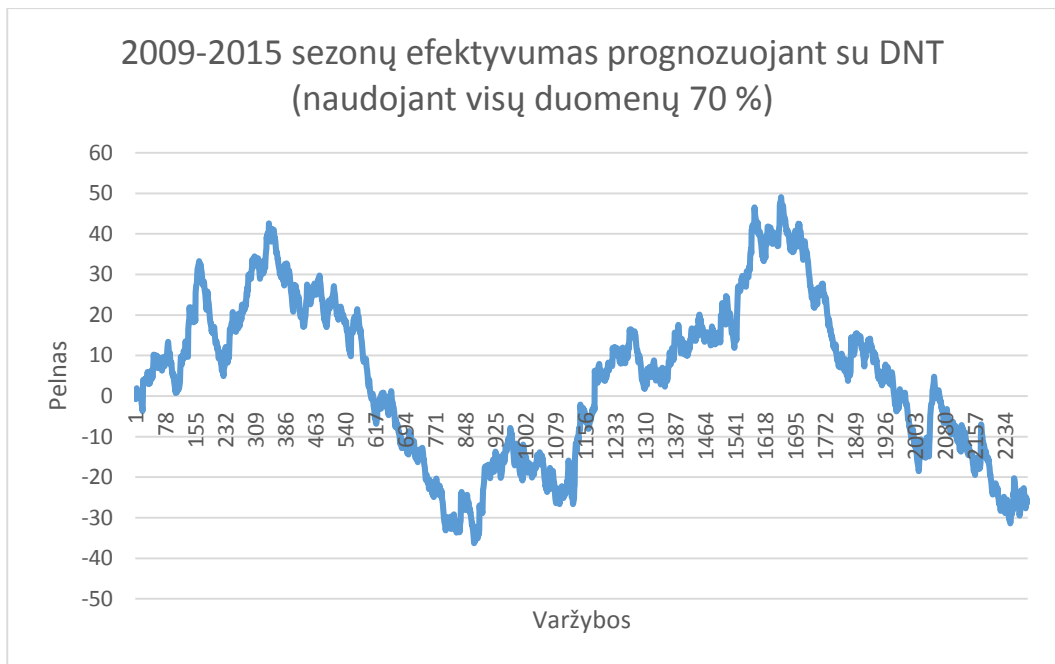
Nors prognozuojant su DNT tikslumas buvo didesnis negu su regresija, bet iš anksčiau pateiktų efektyvumo grafikų matyti, jog, naudojant DNT, gautos efektyvumo funkcijų kreivės yra akivaizdžiai mažėjančios, o prognozuojant regresijos būdu – efektyvumo kreivės yra tolygios. Regresijos tikslumas buvo mažesnis už DNT, tačiau prognozuojant regresijos būdu tiksliau išprognozuoti įvykių rezultatai, už kuriuos buvo pateikti didesni koeficientai, todėl regresijos prognozės šiuo atžvilgiu yra efektyvesnes nei DNT. Efektyvumo rezultatai pasikeičia DNT naudai, kai apmokymo duomenų kiekis yra didelis (nuo 10000 apmokymo pavyzdžių). Iš toliau pateiktų grafikų matyti, jog, apmokymui naudojant visų sezonų duomenis, regresijos būdu gaunamos efektyvumo kreivės tampa mažėjančiomis, o prognozuojant su DNT jos tampa tolygios.



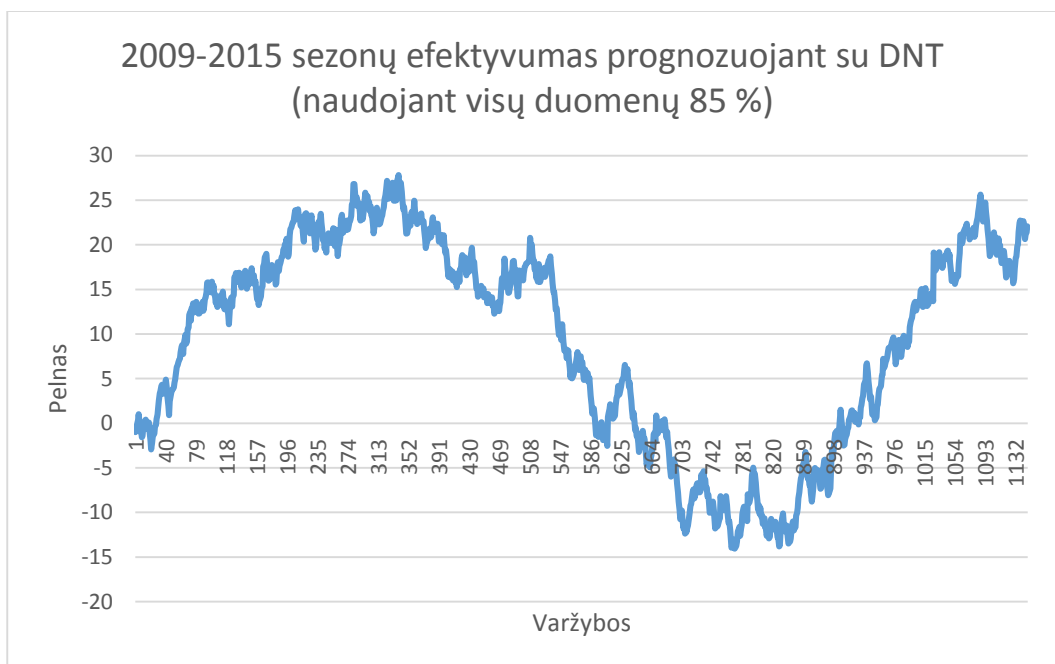
4.5.11 pav. 2009–2015 sezonų efektyvumas prognozuojant su regresija (naudojant visų duomenų 70 %)



4.5.12 pav. 2009–2015 sezonų efektyvumas prognozuojant su regresija (naudojant visų duomenų 85 %)



4.5.13 pav. 2009–2015 sezonų efektyvumas prognozuojant su DNT (naudojant visų duomenų 70 %)



4.5.14 pav. 2009–2015 sezonų efektyvumas prognozuojant su DNT (naudojant visų duomenų 85 %)

Apibendrinus visus rezultatus, galima teigti, kad, esant nedaug apmokymo duomenų, efektyviau naudoti regresiją, nors tikslumas ir bus mažesnis, bet rezultatai efektyvesni. Jeigu duomenų daug ir prognozuojama trumpu periodu, geriau rinktis DNT, nes iš gausių duomenų jis geba gerai įvertinti ryšius su kitais kintamaisiais, o dėl to efektyvumas tik gerėja. DNT pranašesnis už regresiją ir tuo atžvilgiu, kad yra mažiau jautrus neteisingai duomenų interpretacijai.

5. IŠVADOS

1. DNT optimaliai veikia, kai jo apmokymui yra išskiriama 70 % visų turimų duomenų, naudojamų tyrimui. Didesnis nei 70 % duomenų dalies išskyrimas apmokymui ženkliai nepadidina dirbtinio neuroninio tinklo tikslumo.
2. Tam tikro DNT modelio veikimą galima patikrinti imituojant prognozavimą, kai prognozuotojas iš anksto jau žino tam tikrų parametrų įtaką galutinei baigčiai. Jeigu baigtis tiesiškai priklauso nuo parametrų ir apmokymo duomenų kiekis yra pakankamas, bei tinklo modelis yra tinkamas, tada prognozavimo tikslumas visada bus aukštesnis nei 95 %. Atskirų parametrų įtaką galutinei baigčiai galima išsiaiškinti, kai apmokymo duomenyse yra pakeičiami vieno stulpelio duomenys konstanta ir stebima, ar prognozavimo tikslumas pakito. Jeigu rezultatas nepakito, tai reiškia, kad parametras yra nereikšmingas ir gali būti pašalintas iš duomenų.
3. Pasinaudojant matematine regresija naudinga išsiaiškinti, kokią įtaką galutinei baigčiai turi kiekvienas parametras, ir kokius nedarančius įtakos ir nereikšmingus parametrus galima pašalinti iš tyrimo duomenų. Naudojantis regresija, būtina gerai suvokti tyrimo duomenų reikšmę, nes jų interpretavimas regresijoje ir prognozavime su DNT gali skirtis ir suklaidinti tyrėją.
4. Sporte galutinė įvykio baigtis nėra visiškai tiesiškai priklausoma nuo parametrų. Daugelio parametrų įtaka galutiniam rezultatui yra priklausoma ir nuo kitų parametrų. Dėl šios priežasties regresijos prognozavimo tikslumas vidutiniškai yra 3 % prastesnis nei DNT, kuris geba įvertinti parametrų tarpusavio ryšius ir atlikti tikslesnę prognozavimą. DNT yra pranašesnis už regresiją, nes jam gali būti pateikiamas neribotas kiekis parametrų, ir apmokymo proceso metu esant pakankamam kiekiui duomenų jis atitinkamai įvertins parametrų įtaką. Tuo tarpu regresijai kuo daugiau parametrų pateikiama, tuo tikslumas mažėja, o tada prasčiau įvertinama parametrų įtaka.
5. Nors regresijos prognozavimo tikslumas mažesnis negu DNT, bet yra efektyvesnis. Regresija, lyginant su DNT, geba tiksliau prognozuoti daugiau įvykių, už kuriuos buvo siūlomas didesnis koeficientas ir taip gaunamas didesnis pelnas. Regresija efektyvumu DNT pranoksta tik tada, kai apmokymo duomenų kiekis nėra didelis, imtinai iki 5000 apmokymo pavyzdžių. Šis kiekis gali svyruoti priklausomai nuo parametrų kiekio.

Naudoti DNT sporto įvykių baigtims prognozuoti nėra efektyvu dėl per didelio kiekio parametrų, kurie lemia baigtį. Vien žaidėjų statistinių parametrų įtaka baigčiai nėra pakankama, kad ją būtų galima tiksliai numatyti. Šis tyrimas gali būti toliau plėtojamas tirti individualių sporto šakų baigtis – čia susitikimų kiekis nėra didelis, o baigtį lemia tik vieno sportininko parametrai.

6. LITERŪRA

1. Š. Raudys. Žinių išgavimas iš duomenų. Klaipėda, 2008. ISBN 978-9955-18-345-7.
2. A. Verikas, A. Gelžinis. Neuroniniai tinklai ir neuroniniai skaičiavimai. Kaunas, 2008. ISBN 978-9955-591-53-5
3. V. Medvedev. Tiesioginio sklidimo neuroninių tinklų taikymo daugiamačiams duomenims vizualizuoti tyrimai. Vilnius, 2007. ISBN 978-9955-28-211-2.
4. „Forecasting with Artificial Neural Networks“ [Tinkle]. [žiūrėta 2015 m. sausio 14 d.]. Prieiga per internetą: <http://www.neural-forecasting.com/Downloads/EVIC05_tutorial/EVIC'05%20Slides%20-%20Forecasting%20with%20Neural%20Networks%20Tutorial%20SFCrone.pdf>
5. P. J. Braspenning, F. Thuijsman, A.J.M.M.Weijters. 1995: Artificial Neural Networks: An Introduction to ANN Theory and Practice. Germany, Springer.
6. „Alyuda forecasting“ [Tinkle]. [žiūrėta 2015 m. gegužės 8 d.]. Prieiga per internetą: <<http://www.alyuda.com/forecasting-software.htm>>.
7. „NeuroSolutions“ [Tinkle]. [žiūrėta 2015 m. balandžio 5 d.]. Prieiga per internetą: <<http://www.neurosolutions.com/neurosolutions/>>.
8. „EasyNN“ [Tinkle]. [žiūrėta 2015 m. balandžio 19 d.]. Prieiga per internetą: <<http://www.easynn.com/>>.
9. „AgielNN“ [Tinkle]. [žiūrėta 2015 m. balandžio 28 d.]. Prieiga per internetą: <<http://firagiel.com/tbq/index.php?cID=agiel-neural-network/>>.
10. G.Dzemyda. Dirbtiniai neuroniniai tinklai. [Tinkle]. [žiūrėta 2015 m. birželio 12 d.]. Prieiga per internetą: <<http://www.estudijos.vpu.lt/course/view.php?id=31>>
11. „GMDH Shell“ [Tinkle]. [žiūrėta 2015 m. birželio 14 d.]. Prieiga per internetą: <<https://www.gmdhshell.com/neural-network-software>>.
12. „Orange“ [Tinkle]. [žiūrėta 2015 m. lapkričio 18 d.]. Prieiga per internetą: <<http://orange.biolab.si/>>.
13. „NBA & ABA Basketball Statistics & History“ [Tinkle]. [žiūrėta 2016 m. sausio 1 d.]. Prieiga per internetą: <<http://orange.biolab.si/>>.
14. „NBA Results & Historical Odds, Basketball USA Archive“ [Tinkle]. [žiūrėta 2016 m. sausio 1 d.]. Prieiga per internetą: <<http://www.oddsportal.com/basketball/usa/nba/results/>>.
15. „About Box Plus/Minus (BPM)“ [Tinkle]. [žiūrėta 2016 m. kovo 12 d.]. Prieiga per internetą: <<http://www.basketball-reference.com/about/bpm.html>>

7. PRIEDAI

Prie projekto pateikiamas kompaktinis diskas, kuriame įrašyta:

- NetBeans IDE 8.1 RC projektas su sukurta PĮ naudota šiame tyrime.
- Tyrimo duomenų failai.
- MySql duomenų bazė .sql formatu.