



VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS
KOMPIUTERINIO IR DUOMENŲ MODELIAVIMO KATEDRA

Magistro baigiamasis darbas

Ateities sandorių vykdymo analizė

Atliko:

Linas Žvirblis

parašas

Vadovas:

Saulius Blažiūnas

Vilnius
2019

Turinys

Sutartinis terminų žodynas	4
Santrauka	5
Summary	6
Ivydas	7
1. Bendrosios sąvokos	9
1.1. Vertybinių popierių biržos veikimo modelis	9
1.1.1. Atidėtas pavedimas	9
1.1.2. Rinkos įsakymas	9
1.1.3. Rinkos knyga	9
1.2. Ateities sandoriai	10
2. Sandorių vykdymo problema	12
2.1. Fiksuoti ir kintami kaštai	12
2.1.1. Praslydimas	12
2.2. Susijusių darbų apžvalga	13
3. Dirbtiniai neuroniniai tinklai ir jų taikymas finansiniuose tyrimuose	15
3.1. Dirbtinio neurono modelis	15
3.2. Dirbtinis neuroninis tinklas	16
3.3. Dirbtinio neuronų tinklo mokymas	17
3.4. Taikymas finansiniuose tyrimuose	18
4. Sprendimų medžiai ir jų taikymas finansiniuose tyrimuose	19
4.1. Sprendimų medžio sąvoka	19
4.2. Sprendimo medžio mokymas	20
4.3. Taikymas finansiniuose tyrimuose	20
5. Duomenys	22
5.1. Bendras aprašymas	22
5.2. Prognozavimui naudojamos įvestys	23
5.3. Sandorių simuliacijų aprašymas	26
5.4. Prognozuojamų įvykių aprašymas	28
6. Neuroninio tinklo taikymo praktinis tyrimas	30
6.1. Neuronų skaičius paslėptame sluoksnyje	30
6.2. Mokymosi ciklų skaičius	32
7. Sprendimo medžio taikymo praktinis tyrimas	34
7.1. Optimalių parametrų sprendimo medžiui tyrimas	34
8. Dirbtinio neuroninio tinklo ir sprendimo medžio rezultatų palyginimas	38
Išvados ir rekomendacijos	42
Ateities tyrimų planas	43

Literatūros šaltiniai	44
Priedai	47
A. Sandorių vykdymo simuliacijų algoritmas pirkimo atvejui	47

Sutartinis terminų žodynas

1. Ateities sandoris (angl. future contract) – finansinis instrumentas, įpareigojantis pirkti ar parduoti tam tikrą kiekį finansinio turto (akcijų, valiutų, žaliavų, metalų ir daug kitų rūšių) iš anksto sutartą dieną ateityje, sutarta kaina.
2. Rinkos agentas (angl. market agent) – spekuliantas, investicinis fondas, bankas ar bet koks kitas rinkoje dalyvaujantis subjektas.
3. Rinkos įsakymas (angl. market order) – sandorio tipas, kuris įvykdomas momentaliai, geriausia rinkoje siūloma kaina.
4. Atidėtas pavedimas (angl. limit order) – sandorio tipas, kuriam nurodoma norima kaina ir galiojimo laikas. Gali būti atšauktas.
5. Rinkos knyga (angl. limit order book) – sąrašas pirkimo ir pardavimo atidėtųjų pavedimų pateiktų rinkai.
6. Praslydimas (angl. slippage) – sandorio kaštų sudedamoji dalis, atsirandanti nepavykus įvykdyti sandorio norima kaina.
7. Komisiniai (angl. commissions) – fiksuota sandorio kaštų dalis, tenkanti brokeriui.
8. Kintamumas (angl. volatility) – dydis nurodantis vertybinio popieriaus kainos tendenciją keistis.
9. DNT – dirbtinis neuroninis tinklas.
10. MSE (angl. mean square error) – vidutinė kvadratinė paklaida. Matas skaičiuoti prognozavimo modelio tikslumą.
11. SM – sprendimo medis (angl. decision tree).
12. MKJ – minimalus kainos judesys. Dydis per kurį mažiausiai gali keistis finansinio instrumento kaina biržoje.

Santrauka

Šiame darbe nagrinėjama sandorių vykdymo problema ateities sandorių biržose. Darbe iškeliamas tikslas ištirti ar galima sandorių kaštus sumažinti pritaikant duomenų prognozavimo metodus. Tyrinėjama dirbtinio neuroninio tinklo ir sprendimo medžio pritaikymo galimybė. Simuliacijos su realiais ateities sandorių rinkos duomenimis rodo, kad prognozuojant kainos kryptį su dirbtiniais neuroniniais tinklais arba sprendimų medžiais metodus, galima sumažinti sandorių kaštus. Pasiekiamas 6-7% kaštų sumažėjimas, o kainos klasifikavimo tikslumas svyruoja nuo 49% iki 51%. Darbe nagrinėtas sandorių kaštų mažinimo metodas gali būti naudingas įrankis nedideliems investiciniams fondams ar pavieniams investuotojams.

Summary

Analysis of Order Execution in Futures Markets

This paper analyzes the order execution problem in futures markets. Order execution is one of the problems researched in algorithmic trading field. The main goal of this paper is to research a possibility of decreasing the trading costs using either artificial neural network or decision tree. This is done by employing artificial neural network using multilayer perceptron structure. While decision tree uses CART method. The strategy is implemented and simulated by using 2.5 years of real world historical futures level 1 data. Several distinctive multilayer perceptron architectures are compared with various number of inputs. The results show that proposed method for order execution can sometimes outperform naive strategy. The cost are reduced about 6-7%. The prediction accuracy of price classification is between 49% and 51%. This research could be beneficial for individual investors and other smaller market participants.

Iyadas

Sparčiai besivystančios kompiuterių technologijos lemia daugelio sričių perėjimą į skaitmeninę erdvę. Viena iš tokių sferų yra prekyba finansų rinkose. Technologiniam progresui leidus biržas iš realios fizinės vietos perkelti į elektroninę erdvę, buvo suskubta kurti įvairius instrumentus, algoritmus, strategijas ir kitus įrankius palengvinančius investuotojo darbą arba net išvis užimančius jo vietą. Taip gimė algoritminės prekybos mokslo sritis, ji apjungia investavimo, programavimo, statistikos, duomenų analizės žinias į bendrą visumą. Gavusi pradžią prieš maždaug du dešimtmečius, algoritminė (arba skaitmeninė) prekyba pamažu tapo pagrindiniu prekybos būdu finansų rinkose. Prekybos robotai kasmet atsiriekia vis didesnę apyvartos dalį. Algoritminės prekybos dalis nuo 15% 2003 metais kilo iki 60% 2008 metais ir 2012 metais pasiekė net 85% visos apyvartos (M. Glantz et al. [7]). Panašius vertinimus pateikia J. Nagel [19], teigdamas, jog 2015 metais ateities sandorių biržose jau 90% apyvartos sudarė elektroninė prekyba.

Viena svarbesnių algoritminės prekybos nagrinėjamų problemų yra sandorių vykdymas. Pasirinkta sandorių vykdymo strategija turi didelę įtaką prekybos kaštams. Neoptimali taktika gali gerokai sumažinti prekybos pelną ar net privesti iki tol pelningą prekybos sistemą iki nuostolio. Remiantis M. Borkovec et al. [4] net 40% finansų rinkos dalyvių sandorių kaštus kaltina dėl tikėtinų ir realios grąžos nesutapimų (negauto pelno).

Sandorių vykdymą galima apibrėžti kaip veiksmą ar strategiją, kuomet gavus signalą iš algoritminės prekybos sistemos perkama (parduodama) nurodytas kiekis finansinio instrumento, stengiantis kuo arčiau pasiekti signalo iniciatoriaus pasirinktą kainą. Sandoriui įvykdyti signalo iniciatorius skiria tam tikrą toleruojamą laiko periodą. Turint tokias pradines sąlygas: fiksuotą vykdymo laiką, siekiamą kainą bei pasinaudojus skirtingais biržos sandorių tipais (rinkos įsakymu ir atidėtu pavedimu) galima kurti įvairias vykdymo strategijas.

Paprasčiausia sandorių vykdymo strategija yra vos tik gavus signalą pirkti (parduoti), siųsti į biržą rinkos įsakymą taip momentaliai įvykdant prikimą (pardavimą). Tokios strategijos privalumai yra paprastumas, trumpas vykdymo periodas, tačiau ši taktika vidutiniu atveju atneš pastovius fiksuotus kaštus.

Sudėtingesnės sandorių vykdymo strategijos paprastai išnaudoja tuo metu rinkoje pasiekiamą ir istorinę informaciją. Pasiekiamai informacijai pritaikius pasirinktus prognozavimo modelius (pavyzdžiui tiesinę regresiją ar neuroninį tinklą) gaunami mažesni sandorių vykdymo kaštai.

- Aprašyti vertybinių popierių biržos veikimo modelį ir ateities sandorius;
- Apibrėžti sandorių vykdymo problemą;
- Aprašyti praktinėje dalyje naudotus dirbtinį neuroninį tinklą ir sprendimo medį;
- Apibrėžti sandorių vykdymo strategiją;
- Suprogramuoti biržos simuliaciją sandoriams vykdyti;
- Simuliuoti sandorių vykdymo strategijas;
- Rasti optimalius parametrus dirbtiniam neuroniniam tinklui ir sprendimų medžiui;
- Palyginti prognozavimo tikslumą naudojant dirbtinį neuroninį tinklą ir sprendimo medį;

Darbo teorinėje dalyje nuosekliai apibrėžtas metodas, siūlantis sandorių vykdymui pasitelkti neuroninius tinklus. Praktinėje dalyje pasiūlytas metodas realizuotas, naudojantis juo atliktos simuliacijos su realiais istoriniais ateities sandorių rinkos duomenimis. Gauti rezultatai buvo palyginti su praktikoje sutinkama nesudėtinga sandorių vykdymo strategija. Tam tikrais atvejais pasiūlytas metodas veikia neblogai, aplenkdamas lyginamąjį metodą, tačiau dažnu atveju jo tikslumas prastesnis arba panašus. Dėl nepastovaus prognozavimo tikslumo negalima teigti, kad siūlomas metodas yra universaliai geras būdas vykdyti sandoriams, tačiau atiktas tyrimas suteikia pagrindą tolimesniam darbui.

Pirmajame darbo skyriuje pateikiamas vertybinių popierių biržos veikimo modelis, skirtingi sandorių tipai. Aprašomas darbe tiriamas finansinis instrumentas – ateities sandoriai.

Antrasis skyrius skirtas apibrėžti sandorių vykdymo problemą, tuo metu atsirandančius kaštus. Apžvelgiami susiję moksliniai darbai, kurie pabrėžia tokio tyrimo, kaip šis, svarbą.

Trečiame darbo skyriuje aprašomi dirbtiniai neuroniniai tinklai. Pradedant nuo dirbtinio nerono modelio, privedama prie praktiniame tyrime naudojamo daugiasluoksnio perceptrono apibrėžimo ir taikymo principų. Skyrius baigiamas dirbtinių neuroninių tinklų taikymo panašiuose finansiniuose tyrimuose apžvalga.

Ketvirtajame skyriuje supažindinama su sprendimo medžiais, jų galimybe prognozuoti finansinius duomenis.

Penktajame skyriuje aprašomi naudoti duomenys. Pateikiamas sandorių vykdymo algoritmas, taikytos prielaidos ir tyrimo apribojimai. Aprašomas tyrime nagrinėtas kainos krypties prognozavimo principas.

Šeštame ir septintame skyriuose ieškoma optimalių neuroninio tinklo ir sprendimo medžio parametrų.

Aštuntame skyriuje apibendrinami abiejų metodų prognozavimo rezultatai.

Šis magistro baigiamasis darbas yra testinis, pradėtas praėjusio semestro metu. Didžioji 1–3 skyrių, 5.1 poskyrių bei iš dalies 5.3 poskyrio, darbo dalis paimta iš to paties autoriaus mokslo tiriamojo darbo projekto [31].

1. Bendrosios sąvokos

1.1. Vertybinių popierių biržos veikimo modelis

Beveik visos pasaulio elektroninės vertybinių popierių biržos veikia pagal vienodą modelį. Tai komas dvigubo aukciono principas, kuomet tiek pirkėjai, tiek pardavėjai skelbia norimą pirkti ar parduoti vertybinių popierių kiekį ir kainą. Šie nurodymai pirkti arba parduoti, vadinami atidėtaisiais pavedimais (angl. limit orders), nugula ir atsispindi rinkos knygoje (angl. order book). Atidėtieji pavedimai lieka rinkos knygoje iki atšaukimo arba kol atsiranda pirkėjas (analogiškai pardavėjas) sutinkantis mokėti siūlomą kainą. Toliau pateiktas rinkos vertybinių popierių biržos veikimo modelis, paaiškinant skirtingus sandorių tipus ir rinkos knygos principą. Pasitelkiamas pavyzdys.

1.1.1. Atidėtas pavedimas

Atidėtasis pavedimas (angl. limit order) tai toks sandorio tipas, kuomet pirkėjas (pardavėjas) nurodo biržai norą pirkti (parduoti) vertybinių popierių kiekį ir norimą kainą. Šie sandoriai nugula rinkos knygoje. Tokių sandorių iniciatoriai paprastai nesiekia kuo greičiau nusipirkti (ar parduoti) vertybinių popierių taip bandydami gauti jiems palankesnę kainą.

1.1.2. Rinkos įsakymas

Rinkos įsakymas (angl. market order) tai toks sandorio tipas, kuomet pirkėjas (analogiškai pardavėjas) nurodo biržai tik norą pirkti (parduoti) ir vertybinių popierių kiekį. Biržoje toks sandoris įvykdomas momentaliai už geriausią tuo metu siūlomą kainą. Rinkos įsakymą pateikęs pirkėjas gaus mažiausią tuo metu siūlomą pardavimo atidėtųjų sandorių kainą. Analogiškai pardavėjas parduos už didžiausią tarp pirkimo atidėtųjų pavedimų kainą. Šio tipo sandorių privalumas yra įvykdymo greitis, prarandant galimybę gauti šiek tiek geresnę kainą.

1.1.3. Rinkos knyga

Lentelėje 1 pateiktas hipotetinis vertybinio popieriaus rinkos knygos pavyzdys. Sugeneruoti pseudo duomenys. Šios lentelės pagalba bus paaiškinta rinkos knygos veikimo principas.

Kaip jau buvo minėta atidėtųjų pavedimų atvejais jie atsiduria rinkos knygoje. Rinkos knygoje būna tik neįvykdyti ir neatšauti atidėtieji pavedimai. Tarkime prekeivis nori pirkti 1200 vertybinio popieriaus vienetų už nedidesnę kaip 109,85 kainą. Tokiu atveju prekeivis siunčia į biržą atidėtajį pavedimą ir jo sandoris atsiduria rinkos knygoje. Analogiškai tarkime atsiranda dar vienas prekeivis, kuris nori parduoti 400 vienetų už 110,55 arba daugiau, jo sandoris taip pat atsiduria rinkos knygoje. Tokiu principu pildosi rinkos knyga. Ji skiriama į pardavimo pavedimus ir pirkimo pavedimus. Abi rinkos knygos dalys rūšiuojamos taip, kad aukščiau pirkimo eilėje atsiduria didesnę, o pardavimo eilėje – mažesnę kainą siūlantys pavedimai. Atidėtieji pavedimai išliks rinkoje iki kol bus atšaukti arba atsiras pirkėjas (analogiškai pardavėjas) ir sandoris bus įvykdytas.

Jei skirtingi rinkos agentai pateikia pavedimus ta pačia kaina, tai tame pačiame kainos lygyje jie būtų rūšiuojami pagal pateikimo biržai laiką. Tarkime jei atsiratų dar vienas pardavėjas, norintis parduoti už 110,55 kainą, tai jo norimas parduoti kiekis prisidėtų prie esamų rinkos knygoje 400. Tačiau vėlesnio pardavėjo vertybiniai popieriai būtų parduoti tik po to, kai rinkos knygoje nebeliktų prieš jį pateiktų 400 vienetų.

1 lentelė. Rinkos knygos pavyzdys. Užpildyta pseudo duomenimis.

Pirkimo pavedimai		Pardavimo pavedimai	
Kaina	Kiekis	Kaina	Kiekis
110,50	2245	110,55	400
110,40	200	110,60	700
110,10	3220	110,61	810
109,85	1200	110,85	85
109,70	425	110,90	300
109,55	925	111,05	100
109,50	7250	111,20	1300
109,30	1800	111,55	325
109,00	910	111,80	445
108,30	210	112,55	540
106,95	180	114,00	100

Rinkos įsakymo atveju, jei pateiktas pirkimo sandoris, tai jis suporuojamas su tuo metu rinkos knygoje aukščiausiai esančiu (mažiausia kaina) pardavimo atidėtuojų pavedimu. Jei pateiktas rinkos įsakymas parduoti, tai jis poruojamas su aukščiausiai esančiu (didžiausia kaina) pirkimo atidėtuojų pavedimu. Tarkime į biržą pateiktas 300 vienetų pirkimo rinkos įsakymas, o rinkos knyga atrodo analogiškai pateiktai lentelėje 1. Tokiu atveju pirkėjas nusipirktų 300 vienetų už 110,55 kainą, o rinkos knygoje ties šiuo lygiu liktų 100 vienetų. Jei būtų pateiktas rinkos įsakymas, kurio vienetų dydis viršytų siūlomą kiekį ties aukščiausiu kainos lygiu, tai likutis būtų paimtas iš antro geriausio kainos lygio ir t.t. kol rinkos įsakymas būtų pilnai įvykdytas.

Tokiu atveju jei į biržą būtų pateiktas atidėtasis pavedimas pirkti vertybinį popierių už didesnę nei tuo metu aukščiausiai eilėje esanti pardavimo kainą, tai toks sandoris būtų automatiškai paverstas į rinkos įsakymą. Analogiškai pardavimo atveju.

Svarbu paminėti, kad praktikoje paprastai išskiriama geriausia pirkimo (angl. best bid) ir geriausia pardavimo (angl. best ask) siūlomos kainos, bei skirtumas tarp jų (angl. bid ask spread). Geriausios biržoje pasiekiamos kainos ir skirtumas tarp jų parodo kokį nuostolį prekeivis patirtų, jei rinkos įsakymų būdu įsigytų 1 vienetą vertybinio popieriaus ir iškart jį parduotų.

1.2. Ateities sandoriai

Ateities sandoris (būsimasis sandoris) tai išvestinis finansinis instrumentas, įpareigojantis už sutartą kainą pirkti arba parduoti iš anksto apibrėžtą kiekį finansinio turto. Rinkoje kotiruojami ateities sandoriai būna standartizuoti t.y. apibrėžta paskutinė prekybos diena, pristatymo vieta ir data, kiekis, kokybė daug kitų parametru. Būsimieji sandoriai buvo sukurti apsidraudimui nuo netikėtų kainų šuolių ir naudojami ūkininkų, žaliavų išgavėjų, įmonių perdirbėjų. Dabartiniais laikais be originalios paskirties ateities sandoriai dar naudojami ir investavimui. Gera bazė susipažinti su ateities sandoriais yra pateikta Hull [9] knygoje.

Šiame darbe bus apsiribota tik ateities sandorių (angl. futures) vykdymo problema. Ateities sandoriai suteikia galimybę prekiauti įvairiu finansiniu turtu, populiariausi jų: metalai, agrokultūros, valiutos, energetikos priemonės, akcijų indeksai, obligacijos. Būsimųjų sandorių privalumai yra didelis likvidumas, maži komisiniai, galimybė naudoti svertą. Nors ateities sandoriai ir nėra

nišinis instrumentas, o populiarius plačiai pasaulyje naudojamas investavimo ir apsidraudimo įrankis, tačiau plati skirtingų investavimo instrumentų pasiūla leidžia tiriant tik ateities sandorius darba padaryti reikšmingu ir kitoms finansinio turto klasėms (akcijoms, obligacijoms, valiutų rinkai).

Prekyba ateities sandoriais atkartoja nulinės sumos žaidimą (angl. zero sum game). Pavyzdžiui nusipirkus naftos ateities sandorį už 60\$ ir kainai pakilus pardavus jį už 70\$, uždirbtą pelną turės kompensuoti priešingoje pusėje esantis pirkėjas praradęs atitinkamą sumą.

2. Sandorių vykdymo problema

Šiame skyriuje bus plačiau apžvelgtas pagrindinis darbo tyrimo objektas – sandorių vykdymo problema.

2.1. Fiksuoti ir kintami kaštai

Vykdamas sandorį t.y. perkant ar parduodant finansinį instrumentą visada patiriami kaštai. Sandorio vykdymo kaštus galima išskirti į dvi dalis. Pirmiausia tai brokerio komisiniai ir mokestis biržai, kurie dažniausiai būna fiksuoti arba priklauso nuo apyvartos. Juos sumažinti galima tik derantis su brokeriu, šios rūšies kaštų sumažinimo nagrinėjamas nėra prasmingas ir tai darbe nebus daroma. Antroji sandorio vykdymo kaštų sudedamoji dalis yra kintama ir vadinama praslydimu (angl. slippage). Praslydimas bus nagrinėjamas šiame darbe, todėl jis yra išskirtas ir plačiau apžvelgtas atskirame poskyryje.

2.1.1. Praslydimas

Praslydimas yra skirtumas atsirandantis tarp kainos kuria norėta ir tikėtasi įvykdyti sandorį ir realiai rinkoje gautos kainos. Paprastai vidutiniu atveju praslydimas prekeiviui atneša nedidelį nuostolį. Tokie kaštai atsiranda natūraliai dėl rinkoje egzistuojančio skirtumo tarp geriausios pardavimo ir geriausios pirkimo kainos (bid ask spread). Kitos priežastys lemiančios praslydimo efektą:

- Kintamumo rizika. Finansinio instrumento kaina rinkoje gali greitai pasikeisti ir pajudėti nepalankia prekeiviui kryptimi.
- Uždelsimas (angl. lag). Kainai judant labai greitai, net mažiausias uždelsimas dėl pavyzdžiui prekeivio fizinio nutolimo nuo biržos serverių, gali lemti kainos pasikeitimą nepalankia kryptimi.
- Likvidumo trūkumas. Prekeivis pasirenka per didelį vertybinių popierių kiekį lyginant su situacija rinkos knygoje.
- Kiti retesni atvejai, dėl kurių sandoris gali būti neįvykdytas ar uždelstas.

Praslydimas skaičiuojamas toliau pateikta formule:

$$Praslydimas = (P_{gauta} - P_{laukta}) * \mathbb{1}_{kryptis}. \quad (2.1)$$

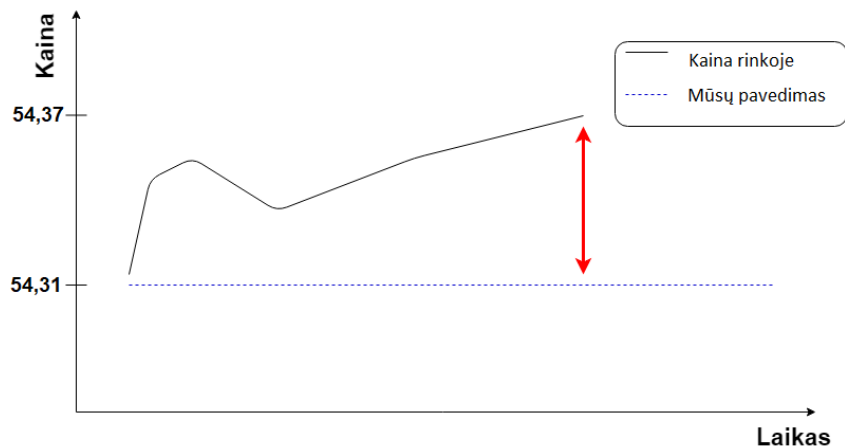
Kur P_{gauta} yra galutinė vykdymo kaina. P_{laukta} yra kaina, kurią tikėtasi gauti vykdant sandorį ir $\mathbb{1}_{kryptis}$ yra sandorio krypties indikatorius, kuris šifruojasi taip:

$$\mathbb{1}_{kryptis} = \begin{cases} 1, & \text{jei pirkimo sandoris} \\ -1, & \text{jei pardavimo sandoris.} \end{cases} \quad (2.2)$$

Taip pat svarbu paminėti, kad šiame darbe $Praslydimas > 0$ yra laikomas nuostoliu prekeiviui, o $Praslydimas < 0$ yra pelnas. Tai tikslinga apsibrėžti, kadangi literatūroje sutinkamas ir atvirkštinis variantas.

Praslydimą galima sumažinti taikant skirtingas sandorių vykdymo strategijas taip pat išnaudojant turimus rinkos duomenis. Sumažinti praslydimą yra pagrindinis darbo praktinės dalies tikslas. Toliau pateiktas praslydimo skaičiavimo pavyzdys.

1 pavyzdys. Tarkime prekeivis naudoja tam tikrą prekybos strategiją, pagal kurią jis pasiruošęs pirkti vieną naftos ateities sandorį už 54,31\$. Kainai pasiekus šią ribą, prekeivis siunčia į biržą sandorį pirkti vieną naftos ateities sandorį už 54,31\$, tačiau instrumento kaina staigiai kyla ir reali įvykdymo kaina yra 54,37\$. Naudojantis formule 2.1 gauname, kad praslydimas lygus 0,06\$. Tai reiškia, kad prekeivis patyrė 0,06\$ neplanuotų kaštų dėl praslydimo. Situaciją iliustruoja 1 pav.



1 pav. Nesėkmingai susiklostęs pirkimo sandoris, kuomet patirta 0,06 praslydimo kaštų

2.2. Susijusių darbų apžvalga

Prognozavimas finansų rinkose yra aktuali ir nelengva problema. Nors kurį laiką buvo ginčijamasi dėl finansinių duomenų chaotiškumo S. Yang et al. [35], E. Peters [22], tačiau pastaruoju metu prieita išvada, kad chaoso duomenyse nėra A. Bensaida [3], B. Adrangi et al. [1]. Iššūkliai tiriant finansines laiko eilutes atsiranda dėl duomenų nestacionarumo T. Schmitt et al. [25] ir netiesiškumo M. Small et al. [28], B. LeBaron [14]. Dėl šių priežasčių paprasti tiesiniai prognozavimo metodai finansiniams duomenims neveikia arba veikia su didesne paklaida už netiesinius metodus D. McMillan [18].

Nors efektyvios rinkos teorija E. Fama [6] teigia, kad visa rinkos agentams pasiekiamą informacija yra įskaičiuota į kainą ir neefektyvumo rinkose nėra, tačiau ypač trumpame periode, kuris ir bus nagrinėjamas praktinėje dalyje ši teorija neveikia. Tokį teiginį grindžia darbai nagrinėjantys rinkos knygos dinamiką trumpame periode. Vienas iš pavyzdžių Zheng et al. [36] darbas, kuriame autoriai naudodami realius Prancūzijos akcijų biržos duomenis parodo, jog kainos judesius galima nuspėti iš rinkos knygos informacijos. Tyrime taip pat teigiama, kad prognozei pakanka geriausios siūlomos kainos duomenų. Kercheval ir Zhang [12] panaudoja rinkos knygos duomenis ir pritaikę atraminių vektorių klasifikatorių empiriškai parodo, jog jų pasiūlytas metodas tinkamas trumpo periodo kainos judesiui prognozuoti.

Dauguma darbų nagrinėjančių sandorių vykdymą koncentruojasi į kaštų mažinimą dėl likvidumo trūkumo rinkoje. Tiriama problema, kuomet bandoma įvykdyti didelius sandorius, kurie išaugina prekybos kaštus dėl savo masyvumo arba mažos pasiūlos rinkoje. Šiame darbe pasirinkta kita kryptis, mažesni sandoriai ir trumpesnis vykdymo periodas. Darbo tema aktualesnė mažesne

apyvarta prekiaujantiems rinkoms agentams, pavyzdžiui Lietuvos investiciniams fondams, kurie išskyrus pensijinius fondus, savo dydžiu gerokai nusileidžia pasaulio standartams. Vienas iš darbų atitinkantis šio tyrimo temą Lehalle et al. [15], kur autoriai siūlo metodą sandorių vykdymui pranokstantį nesudėtingas strategijas. Pasitelkus rinkos knygos modelį, tyrime tiek teoriškai, tiek praktiškai įrodomas pasiūlytos strategijos veikimas. Raudys et al. [24] aprašo ir empiriškai tiria paprastą sandorių vykdymo strategiją, kuri anot autorių veikia geriau, negu prekyba naudojant vien tik rinkos įsakymus. Jų darbe nagrinėjami ateities sandoriai.

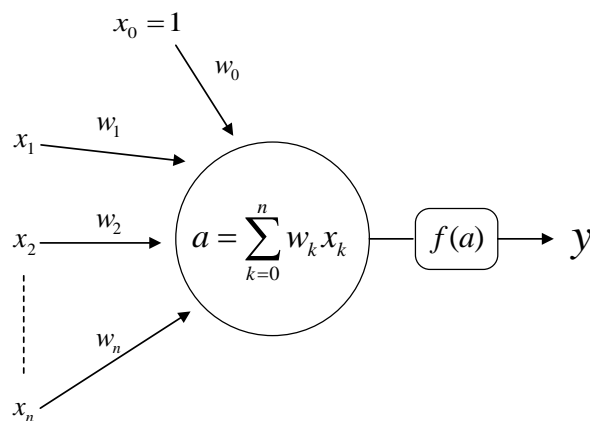
Apibendrinant aprašytus susijusius darbus galima daryti išvadas, jog prognozės finansų rinkose gali duoti teigiamus rezultatus, tačiau dėl duomenų ypatybių, norint pasiekti didesnę tikslumą, prioritetą reikėtų taikyti netiesinės analizės metodams. Tyrimai patvirtina, jog naudojantis rinkos knygos duomenimis įmanoma prognozuoti trumpo periodo kainos judesius. Darbai sandorių vykdymo srityje patvirtina, jog naudojantis pasiekiamą informacija įmanoma aplenkti naivias sandorių vykdymo strategijas. Paminėtos išvados leidžia teigti, jog šio darbo praktinėje dalyje atliekamas tyrimas turi mokslinę prasmę ir pagrindimą.

3. Dirbtiniai neuroniniai tinklai ir jų taikymas finansiniuose tyrimuose

Dirbtiniai neuroniniai tinklai (DNT, angl. artificial neural networks) arba tiesiog neuroniniai tinklai tai modeliavimo aparatas, atkartojantis biologinių neuronų veikimo modelį. DNT kaip būdą modeliuoti duomenims pasiūlė W. McCulloch ir W. Pitts [17] 1943 metais. Neuroniniai tinklai apmokomi iš imčių pavyzdžių ir prie jų prisitaikę, vėliau geba prognozuoti nematytus duomenis. DNT yra taikomi vaizdų ir signalų atpažinime, diagnostikoje, medicinoje, finansuose, modeliavime ir daug kitų sričių.

3.1. Dirbtinio neurono modelis

DNT sudaro paskirų dirbtinių neuronų (2 pav.), sujungtų jungtimis, aibė. Todėl pirmiausia bus apžvelgtas šis elementas.



2 pav. Dirbtinis neuronas

Neurono užduotis yra gavus įėjimo reikšmes x_1, x_2, \dots, x_n , grąžinti išėjimo reikšmę y . Išėjimo reikšmė gaunama tokiu principu (žr. 2 pav.):

- Neuronas gauna baigtinį skaičių įėjimo reikšmių x_1, x_2, \dots, x_n , kiekviena iš gautų reikšmių turi svorį w_1, w_2, \dots, w_n . Taip pat neuronas dar turi laisvąjį narį w_0 (ang. bias).
- Skaičiuojama įėjimo reikšmių ir svorių sandauga $a = \sum_{k=1}^n w_k x_k + w_0$ arba paprasčiau $a = \sum_{k=0}^n w_k x_k$, jei įvestume pastovų $x_0 = 1$.
- Naudojant ankstesniu žingsniu gautą reikšmę a , pritaikius aktyvacijos funkciją f , apskaičiuojama išėjimo reikšmė $y = f(a)$.

Laisvojo nario w_0 prasmė yra pastumti aktyvacijos funkciją, tai padeda DNT pasiekti didesnę tikslumą. Aktyvacijos funkcijų yra įvairių, populiariausios yra sigmoidinė (3.1), slenkstinė, tiesinė, tangento sigmoidas ir kitos. Keletas dažniau naudojamų aktyvacijos funkcijų apžvelgta [11]. Šio darbo praktinėje dalyje nebus išsivelta į palyginimus kuri aktyvacijos funkcija geriausiai veikia

konkrečiu atveju. Universaliai gerai įvairiems atvejams tinka sigmoidinės aktyvacijos funkcija K. Hornik [8], ji dažniausiai sutinkama tyrimuose, todėl bus apsiribota šios funkcijos taikymu:

$$f(a) = \frac{1}{1 + e^{-a}}. \quad (3.1)$$

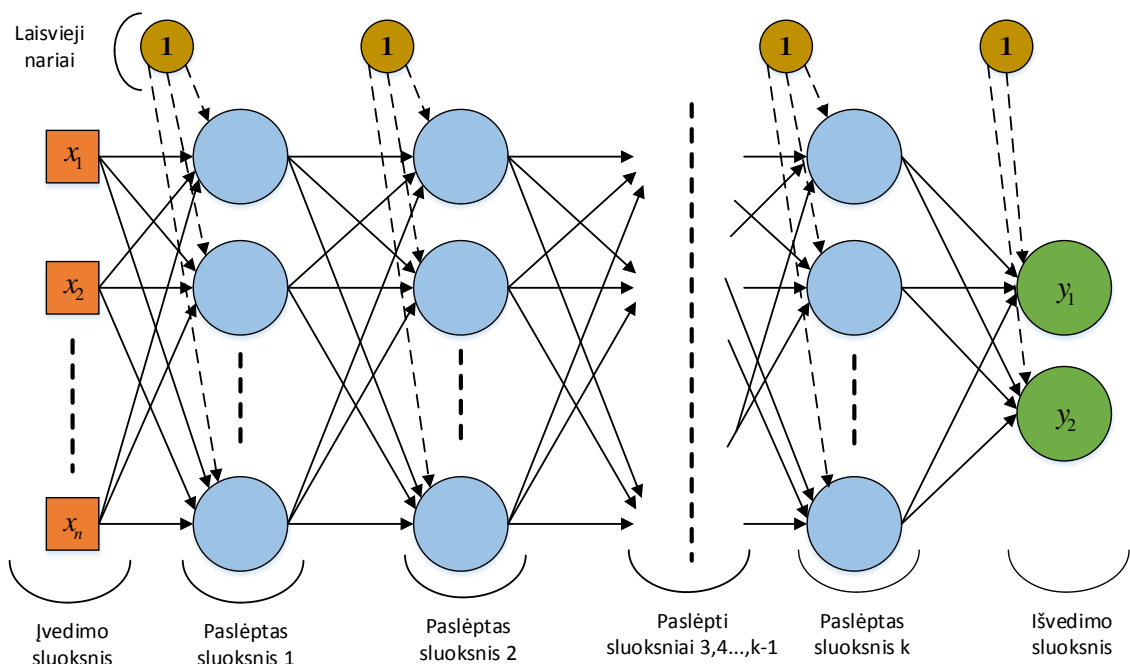
3.2. Dirbtinis neuroninis tinklas

Dirbtiniai neuronai sujungiami į tinklą ir taip gaunamas DNN. Neuroninis tinklas gali būti pavaizduotas kaip grafas, su viršūnėmis neuronais, o šakos svoriai tarp neuronų su svoriais. Pagal neuronų sujungimo būdą skiriami du neuroninių tinklų tipai: tiesioginio sklido (angl. feed forward) ir grįžtamo ryšio (feedback) tinklai. Tiesioginio sklido DNT atveju signalai sklinda nuosekliai, viena kryptimi. Grįžtamo ryšio DNT sudėtingesni, signalai tarp neuronų gali sklirti įvairiomis kryptimis. Šiame darbe apsiribota tiesioginio sklido neuroniniais tinklais.

Dažniausiai praktikoje pasitaikantys DNT sudaryti iš trijų sluoksnių:

- Įvedimo sluoksnis – duomenys paduodami į dirbtinį neuroninį tinklą.
- Paslėptas sluoksnis – pasluoksniui vyksta skaičiavimai, pereinant per visus paslėptus sluoksnius.
- Išvedimo sluoksnis – gauti neuroninio tinklo rezultatai.

Praktikoje dažniausiai sutinkamas modelis yra daugiasluoksnis perceptronas (angl. multilayer perceptron) [30]. Šis modelis išpildo tiesioginio sklido DNT savybes yra gana nesudėtingas ir universalus, todėl bus taikomas praktinėje dalyje. Jo savybės ir parametrai tinka tiriamai problemai spręsti. Tokios DNT struktūros pavyzdys pateiktas 3 paveiksle.



3 pav. Daugiasluoksnio perceptrono pavyzdys, su n įėjimo reikšmių, k paslėptų sluoksnių ir 2 išvesties reikšmėmis, svoriai diagramoje nepažymėti

3.3. Dirbtinio neuronų tinklo mokymas

Skiriami trys dirbtinių neuroninių tinklų mokymo būdų tipai:

- Mokymas su mokytoju – tinklui pateikiama aibė $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, kur $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ i -ojo mokymo imties elemento įvedimo reikšmių vektorius, o $y_i = (y_{i1}, y_{i2}, \dots, y_{ik})$ i -ojo mokymo imties elemento tikslinė išvedimo reikšmė. DNT mokosi keisdamas neuronų jungčių svorius, bandydamas pasiekti, kad jo išvedimo reikšmės kuo labiau priartėtų prie duotųjų. Formaliai tariant neuroninis tinklas ieško tokios funkcijos $f : X \rightarrow Y$, kur $X = (x_1, x_2, \dots, x_n)$ įvedimo reikšmių vektorius, o $Y = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ DNT išvedimo reikšmės (prognozė). Rastoji funkcija f turi kaip galima labiau sumažinti skirtumus tarp prognozių \hat{y}_i ir tikslinių išvedimo reikšmių y_i , $i = \overline{1, n}$.
- Mokymas su paskatinimu – panašiai kaip ir mokyme su mokytoju, bet užuot davus teisingas reikšmes tinklui pasakoma ar jo rezultatas teisingas, ar neteisingas.
- Mokymas be mokytojo – tinklui neduodama jokios informacijos apie išvedimo reikšmes. DNT mokymosi metu remiasi vidiniais kriterijais ieškodamas panašumų tarp mokymo aibės įvesčių. Pavyzdžiui tinklui padavus įvairių nuotraukų, kur vienoje iš jų pavaizduoti žmonės, o kitose šunys, apmokius mokymo be mokytojo principu, DNT tikėtina atskirtų nuotraukas į dvi klases.

Tyrime naudojamam daugiasluoksnio perceptrono tipo DNT apmokyti taikomas taip vadinamas „klaidos sklidimo atgal“ algoritmas (angl. backpropagation algorithm). Šis algoritmas naudoja mokymo su mokytoju strategiją, jo esmė minimizuoti dirbtinio neuroninio tinklo klaidą naudojantis gradientinio nusileidimo metodu. Algoritmas pavadintas taip, nes pradedant nuo išėjimo, skleidžia klaidą atgal link įvedimo sluoksnio.

„Klaidos sklidimo atgal“ algoritmas pilnai išvestas P. Werbos [33] straipsnyje. Pagrindiniai žingsniai taikant „klaidos sklidimo atgal“ metodą, kurie tai pat bus naudojami šiame darbe yra tokie:

1. Atsitiktiniu ar kitu būdu parenkami pradiniai DNT svoriai ir naudojant įvedimo reikšmes apskaičiuojamos neuroninio tinklo išvedimo reikšmės;
2. Nustatoma paklaida taikant paklaidos funkciją, jei paklaidos funkcija nelygi nuliui tęsiame;
3. Naudojamos formulės, išvestos pasiremiant dalinėmis išvestinėmis. Pradedant sluoksniu, esančiu tolimiausiu nuo įvedimo sluoksnio ir pasluoksniui artėjant prie įvedimo sluoksnio perskaičiuojami svoriai kiekvienam paslėptam sluoksniui;
4. Jeigu nepasiektas sustojimo kriterijus (pvz. fiksuotas iteracijų skaičius ar paklaidos reikšmė), tinklui vėl paduodami duomenys, jis persiskaičiuoja ir grįžtama į 2 žingsnį;

Paklaidos funkcija gali būti įvairi, dažniausiai sutinkami pavyzdžiai: kvadratinių paklaidų vidurkis (angl. mean squared error, MSE), kvadratinių paklaidų suma (angl. sum of squared error, SSE), absoliučių paklaidų vidurkis (angl. mean absolute error, MAE), kvadratinių paklaidų vidurkio šaknis (angl. root mean square error, RMSE). Šiame darbe DNT apmokyti apsiribota kvadratinių paklaidų vidurkio funkcija (MSE):

$$E(W) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (3.2)$$

čia W atitinka DNT svorių matricą, y_i reali, žinoma i -ojo n dydžio imties elemento išvedimo reikšmė, o \hat{y}_i yra prognozė tam pačiam imties elementui. Žvelgiant į $E(W)$ formą, akivaizdu, jog $E(W) \geq 0$ bet koku atveju. Todėl mokymo metu stengiamasi kuo labiau minimizuoti paklaidos funkcijos reikšmę ją artinant prie nulio.

3.4. Taikymas finansiniuose tyrimuose

Dirbtiniai neuroniniai tinklai sugeba atlikti sudėtingas klasifikacijos ir atpažinimo problemas [21]. DNT yra universalus įrankis, galintis aproksimuoti bet kokią funkciją [16], tai ypač svarbu prisimenant poskyryje 2.2 pasiektas išvadas dėl finansinių duomenų sudėtingumo.

Ankstesnę teiginį patvirtina mokslinė literatūra. Neuroniniai tinklai dažnai naudojami tyrimuose prognozuojant finansinių popierių kainas. Vienas iš pavyzdžių Isfan et al. [10] darbas. Autoriai pritaiko DNT nuspėti Portugalijos akcijų biržos kainoms. Panašių empirinio pagrindo darbų galima rasti daug, todėl plačiau jie nebus nagrinėjami. Bendras principas prognozuoti akcijų kainas išdėstyta Lawrence [13].

Silva et al. [26] pritaiko daugiasluksnį perceptroną įvairaus periodo Brazilijos akcijų rinkoje prekiavimų instrumentų kainai prognozuoti. Autoriai teigia, jog DNT geriau atspėja kainos dinamiką trumpesniuose laiko perioduose. Sirignano [27] darbe sukuria naują neuroninio tinklo architektūrą, kurią panaudoja geriausių siūlomų pirkimo ir pardavimo kainų rinkos knygoje prognozavimui. Apmokyti DNT autorius naudoja 500 akcijų rinkos knygų duomenis. Darbe pateikiamas prognozavimo palyginimas tarp sukurtos architektūros neuroninio tinklo, tradicinio DNT ir logistinės regresijos modelio. Abiejų architektūrų DNT tikslumu lenkia logistinės regresijos modelį.

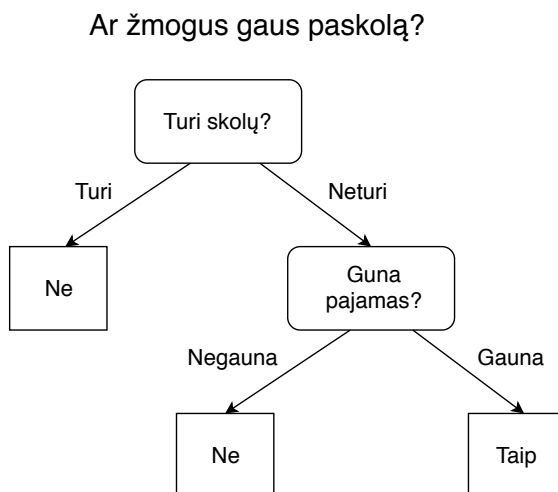
Remiantis tuo, jog DNT yra galingas įrankis, sugebantis spręsti įvairaus pobūdžio klasifikacijos ir atpažinimo problemas, taip pat yra sėkmingai taikomi panašiuose tyrimuose, autoriaus nuomone, tikslinga naudoti daugiasluksnį perceptroną praktinės dalies tyrimui realizuoti.

4. Sprendimų medžiai ir jų taikymas finansiniuose tyrimuose

4.1. Sprendimų medžio sąvoka

Sprendimų medis (SM, angl. decision tree) tai populiarus duomenų modeliavimo įrankis, atkartojantis medžio struktūrą. Metodo esmė yra sukurti daugiapakopę sistemą, kurioje galimi sprendiniai būtų nuosekliai atmetami remiantis žinomais įvesties kintamaisiais iki kol gaunamas galutinis sprendinys.

Sprendimų medį sudaro mazgų aibės ir juos jungiančios briaunos (šakos). Medis prasideda nuo vienintelio mazgo neturinčio į jį vedančios briaunos, vadinamu šaknimi (angl. root). Šaknį 4 pav. atitinka „Turi skolų?“ klausimas. Visi kiti medžio mazgai turi po lygiai vieną į juos vedančią šaką. Mazgai turintys į juos įeinančią ir išeinančias briaunas yra vadinami vidiniais mazgais (angl. internal node). Vidinio mazgo pavyzdys yra „Guna pajamas?“ klausimas 4 pav. Galiausiai mazgai turintys tik įeinančią šaką vadinami lapais. Lapus atitinka 4 pav. atsakymai į klausimus. Medis „skaitomas“ pradedant nuo šaknies, judant vidiniais mazgais iki kol pasiekiamas lapas – sprendinys. Tokia struktūra leidžia vizualiai parodyti sprendimus, kartu pateikiant pasirinkimo logiką.



4 pav. Sprendimo medžio pavyzdys

Sprendimo medis gali priklausyti vienai iš dviejų rūšių:

- **Klasifikavimo medis.** Kuomet prognozuojama baigtis yra diskretus dydis t.y. sprendžiama klasifikavimo problema. Pavyzdžiui spėjama ar krepšinio komanda laimės sekančias rungtynes.
- **Regresijos medis.** Kai prognozuojamas rezultatas yra tolydus dydis t.y. realusis skaičius. Sprendžiamas regresijos uždavinys. Pavyzdžiui vertinama namo kaina.

Šiame darbe nagrinėjama klasifikavimo problema, todėl naudojami klasifikavimo medis.

Sprendimo medžio mokymas susideda iš dviejų dalių. Pirmiausiai vyksta sudarymo („augimo“) fazė, kuomet remiantis duota duomenų aibe kintamieji rekursyviai suformuoja medžio struktūrą. Procedūros metu siekiama gauti didžiausią prognozavimo tikslumą duotiems duomenims. Sudarius medį vyksta genėjimo fazė (angl. pruning). Ji nėra privaloma, tačiau „auginant“ medį jo

struktūra dažnu atveju tampa per sudėtinga. Tai reiškia, kad sprendimo medis per smarkiai pritaiko prie mokymo imties ir naujų duomenų atveju prognozės būna netikslios. Dėl šios priežasties atliekamas genėjimas. Pašalinama dalis žemiausniųjų lygmenų šakų, taip padidinant prognozavimo tikslumą nematytiems duomenims.

Sprendimo medžio taikymas turi keletą privalumų. Sprendimo medis nereikalauja normalaus duomenų pasiskirstymo ir kintamųjų tarpusavio nepriklausomumo. Įvestys gali būti tiek skaitinės, tiek kategorinės. Šios dvi savybės palengvina duomenų paruošimą ir duoda pranašumą prieš kitus modeliavimo metodus, reikalaujančius normalaus duomenų pasiskirstymo. Galiausiai sprendimo medį (iki kol jis nėra per daug didelis) lengva interpretuoti grafiškai.

Sprendimo medžio taikymo pagrindinis trūkumas yra dažnu atveju per didelis prisitaikymas prie mokymo imties. Taip atsitinka, kuomet medžiui „augimo“ fazėje leidžiama tapti per daug sudėtingu, išmokstant tik tos konkrečios duomenų imties ypatybes. Nuo prisitaikymo prie duomenų gelbsti genėjimas arba medžio dydžio apribojimas sudarymo fazėje.

Dar viena problema pasitaikanti naudojant sprendimo medį yra jo tendencija pasiekti lokaliai geriausią sprendimą, nepriartėjant prie globaliai optimalaus atsakymo. Šią problemą nulemia tai, jog dauguma sprendimo medžių realizuojančių algoritmų yra godūs. Medį apmokant keletą kartų, su skirtingomis imtimis galima šią problemą sumažinti.

Dažniausiai sprendimų medžiai taikomi medicinoje, daugiausiai diagnostikai. SM pritaikymas finansiniam modeliavimui nėra ypač gausus, tačiau sutinkamas literatūroje.

4.2. Sprendimo medžio mokymas

Egzistuoja nemažai SM realizuojančių algoritmų. Gerą apžvalgą pateikia Anyanwu et al. [2]. Darbe skirtiengi SM mokymo algoritmai nagrinėti juos taikant įvairiems duomenų rinkiniams (finansų, transporto, mokslo, sveikatos). išskirti geriausiai pasirodę algoritmai: CART, IDE (ID3), C4.5. Tiesa C4.5 ir ID3 priklauso tai pačiai algoritmų šeimai, todėl iš esmės pagal populiarumą varžosi dvi grupės, pastaroji ir CART. ID3 šeima ir CART kartu yra ir populiariausi literatūroje sutinkami SM mokymo algoritmai.

CART metodas pasiūlytas Breiman [5] knygoje 1984 metais. Jo pagrindinės savybės tai, kad gali būti naudojamas tiek klasifikavimo, tiek regresijos medžių mokymui. Klasifikavimo medžio atveju kiekvienas, mazgas turi po 2 vaikus t.y. formuojamas binarinis medis. Mazgams vertinti dažniausiai naudojamas Gini koeficientas.

Darbe naudoti pasirinktas CART metodas, tai motyvuojama metodo prieinamumu ir populiarumu. Jis dažnai sutinkamas įvairaus spektro tyrimuose.

4.3. Taikymas finansiniuose tyrimuose

Sprendimų medžiai naudojami skirtinguose finansinio prognozavimo uždaviniuose. Wang et al. [32] darbe prognozuoja „Microsoft“, „Intel“ ir „IBM“ akcijų kainų pokyčius ir kuria strategiją kaip pelningai jas prekiauti. Autoriai panaudoja dvigubą sprendimo medį, pirmasis SM nusprendžia ar akcijos kaina kitą dieną kils. Antrasis renkasi ar pirkti akciją, jei nuspręsta, kad jos kaina augs. Lyginant su atsitiktiniais akcijos pirkimais, tokia strategija prognozuoja kainą 10-20% tiksliau. Taip pat anot autorių sprendimo medžių prognozavimo metodo privalumai yra greitas apmokymas, stabilumas ir aiškus suprantamumas.

Wu et al. [34] panaudoja C4.5 algoritmu paremtą sprendimo medį prognozuoti akcijoms kotiruojamoms Taivano ir NASDAQ akcijų biržose.

Literatūroje galima rasti tyrimų, kur finansinio modeliavimo kontekste lyginami ir šiame tyrime naudojami prognozavimo įrankiai: sprendimo medis ir dirbtinis neuroninis tinklas. Vienas pavyzdys Qian et al. [23], autoriai savo darbe DOW Jones indekso kainas prognozuoja naudodami keletą skirtingų instrumentų, tarp kurių yra sprendimo medis ir DNT. Mažiausias klaidas mokymo imtyje pasiekia DNT, o testinėje imtyje geriausius rezultatus rodo SM, palikdami už nugaros k – artimiausių kaimynų metodą.

Nair et al. [20] tyrime taip pat nagrinėjami sprendimų medžiai. Autoriai parodo, kad sprendimo medis prognozavimo tikslumu lenkia dirbtinį neuroninį tinklą. Tai motyvuojama tikslesniu trendų prognozavimu Mumbajaus akcijų rinkoje.

Dar vienas skirtingų prognozavimo metodų palyginimas pateikiamas Tsai et al. [29] darbe. Tyrime Taivano akcijų biržos akcijų kainas prognozuojamos naudojantis daugiasluoksnį perceptroną, CART algoritmu paremtą sprendimo medį ir logistinę regresiją. Geriausią spėjimo tikslumą parodo DNT, o kiti du likę metodai nežymiai atsilieka.

5. Duomenys

5.1. Bendras aprašymas

Šiame tyrime panaudoti vadinamieji 1 lygio rinkos duomenys (angl. Level 1 Market Data). Naudojamai duomenų masyvai neturi sutarto periodo (pvz. vienas įrašas kas sekundę, minutę, valandą, dieną ar bet kokią kitą laiko vienetą). Užuoat pateikus periodinę rinkos situaciją, tyrime naudojamuose duomenyse atspindimi visi rinkoje įvykę pokyčiai. Tokiu atveju neprarandama jokios rinkoje buvusios informacijos ir galima kiek įmanoma tiksliau atlikti simuliacijas. 1 lygio rinkos duomenų masyvai saugo tokią informaciją:

- **Pirkimo kaina:** geriausia rinkoje siūloma kaina, už kurią kas nors nupirktų finansinį turtą.
- **Pirkimo kiekis:** siūlomas kiekis už geriausią pirkimo kainą.
- **Pardavimo kaina:** geriausia rinkoje siūloma kaina, už kurią kas nors parduotų finansinį turtą.
- **Pardavimo kiekis:** siūlomas kiekis už geriausią pardavimo kainą.
- **Paskutinė suprekiauta kaina:** kaina kuria įvyko paskutinis sandoris rinkoje.
- **Paskutinis suprekiautas kiekis:** paskutiniu sandoriu rinkoje suprekiautas kiekis finansinio turto.

Tokie duomenys leidžia pakankamai tiksliai simuliuoti sandorių vykdymą, kadangi bet kuriuo metu žinoma geriausios rinkoje pasiekiamos pirkimo ir pardavimo kainos, be to duomenų masyvuose atvaizduoti visi rinkos knygoje įvykę pasikeitimai. 1 lygio rinkos duomenys leidžia pasiekti didesni tikslumą simuliuojant, negu literatūroje sutinkami OHLC¹ ar tikiniai² duomenys, naudoti [24] tyrime. 1 lygio rinkos duomenys yra viršaišis tiek tikiniams, tiek OHLC duomenų tipams. Tai reiškia, jog pastarieji duomenys gali būti sukonstruoti iš 1 lygio rinkos duomenų³.

2 lentelė. Duomenų pavyzdys. Aukso ateities sandorio, 2016 Balandžio kontrakto duomenų iškarpa.

Data, laikas	Pard. kaina	Pard. kiekis	Pirk. kaina	Pirk. kiekis	Pask. kaina	Pask. kiekis
...						
2016-01-28 10:06:11.920	1118,6	7	1118,5	3	1118,5	0
2016-01-28 10:06:13.493	1118,7	20	1118,5	6	1118,6	1
2016-01-28 10:06:13.501	1118,7	19	1118,5	6	1118,6	0
...						

¹OHLC – angl. open high low close. Duomenų masyvai su tam tikru periodu (minutė, sekundė, valanda ir kiti), kaupiantys tik datą, laiką, atidarymo, uždarymo, aukščiausią ir žemiausią kainas.

²angl. tick data. Duomenys turintys visus rinkoje įvykdytus sandorius. Juose kaupiama informacija: data, laikas, sandorio kaina, sandorio dydis

³1 lygio rinkos duomenų viršaišis būtų 2 lygio rinkos duomenys (angl. Level 2 Market Data). 2 lygio rinkos duomenys dar papildomai turi daugiau negu po vieną geriausią siūlomą/prašomą kainą (paprastai nuo 2 iki 15).

Tyrime panaudota prekybos istorija iš 10 įvairaus spektro likvidžių ateities sandorių, prekiaujamų skirtingose pasaulio biržose: CME, CBOT, CBOEF, NYMEX, COMEX, EUREX. Šių skirtingų instrumentų pasirinkimas grindžiamas siekiu iširti sandorių kaštų vykdymo sumažinimo galimybę vidutiniu rinkos atveju. Pasirinkti instrumentai iš skirtingų turto klasių tikėtina panaikins priklausomybę nuo tam tikros vienos turto klasės. 2 iš pasirinktų ateities sandorių atspindi akcijų indeksų pokyčius, 3 obligacijų. Po 1 energijos, metalų, valiutų kursų, agrokultūrų ir kintamumo indekso pokyčius.

Duomenų istorijos periodas apima laikotarpį nuo 2015 birželio iki 2018 vasario. Tyrime naudoti duomenys plačiau aprašyti 3 lentelėje. Duomenų pavyzdys pateiktas 2 lentelėje. Lentelėje 4 atskirai paaiškinti duomenų masyvo laukeliai.

3 lentelė. Tyrime naudotų istorinių rinkos duomenų aprašymas. Pirmame stulpelyje „Trump.“ reiškia ateities sandorio trumpinį. Šie trumpiniai toliau naudojami darbe.

Trump.	Aprašymas	Birža	Pirma data	Pask. data	Eilučių sk.
ES	S&P 500 akcijų indeksas	CME	2015-12-13	2018-02-28	191,7 mln.
TY	10 metų JAV obligacijos	CBOT	2015-06-15	2018-02-28	120,3 mln.
CL	Nafta	NYMEX	2015-06-15	2018-02-28	236,4 mln.
FV	5 metų JAV obligacijos	CBOT	2015-08-28	2018-02-28	79,3 mln.
GC	Gold	COMEX	2015-06-15	2018-02-28	108,2 mln.
EC	Euras	CME	2015-06-15	2018-02-28	170,7 mln.
C	Kukurūzai	CBOT	2015-06-15	2018-02-28	31,2 mln.
FGBL	Euro obligacijos	EUREX	2015-09-14	2018-02-28	98,2 mln.
VX	CBOE Kintamumo indeksas	CBOEF	2015-09-06	2018-02-28	55,1 mln.
FDAX	DAX akcijų indeksas	EUREX	2015-06-15	2018-02-28	90,7 mln.

4 lentelė. Tyrime naudotų duomenų masyvo laukeliai, jų tipai ir trumpas paaiškinimas.

Stulpelio vardas	Duomenų tipas	Aprašymas
Data ir laikas	Datos ir laiko	Data ir laikas biržoje
Pardavimo kaina	Realiųjų skaičių	Geriausia siūloma pardavimo kaina
Pardavimo kiekis	Natūraliųjų skaičių	Siūlomas pardavimo kiekis
Pirkimo kaina	Realiųjų skaičių	Geriausia siūloma pirkimo kaina
Pirkimo kiekis	Natūraliųjų skaičių	Siūlomas pirkimo kiekis
Paskutinio sandorio kaina	Realiųjų skaičių	Paskutinio įvykusio sandorio kaina
Paskutinio sandorio kiekis	Natūraliųjų skaičių	Paskutinio įvykusio sandorio kiekis

5.2. Prognozavimui naudojamos įvestys

Panašaus pobūdžio darbuose įvesčių pasirinkimas prognozavimo metodams paprastai būna vieno iš dviejų tipų. Viena grupė tyrėjų naudoja pasirinktą skaičių paskutinių rinkoje žinomų kainų ir niekaip jomis nemanipuliuoja t.y. palieka DNT, SM ar kitam prognozavimo metodui pačiam rasti kainų pokyčiuose galimai esančią informaciją. Tokio pasirinkimo atveju dažniausia naudojami gilaus mokymo metodai, mokymo gretis juose gana lėtas.

Kita grupė bando palengvinti darbą prognozavimo metodams apdorodami duomenis prieš prognozavimą. Dažnu atveju naudojami įvairūs indikatoriai (pavyzdžiui slenkantis vidurkis) ar skai-

čiuojamos kitos duomenų charakteristikos. Taip tikintis, kad manipuliacija duomenimis padidins prognozavimo tikslumą bei mokymosi greitį.

Šiame tyrime pasirinktas tam tikras kompromisas tarp dviejų stovyklų. DNT ir SM duodama tiek nepaliestų duomenų, tiek apdorotos informacijos.

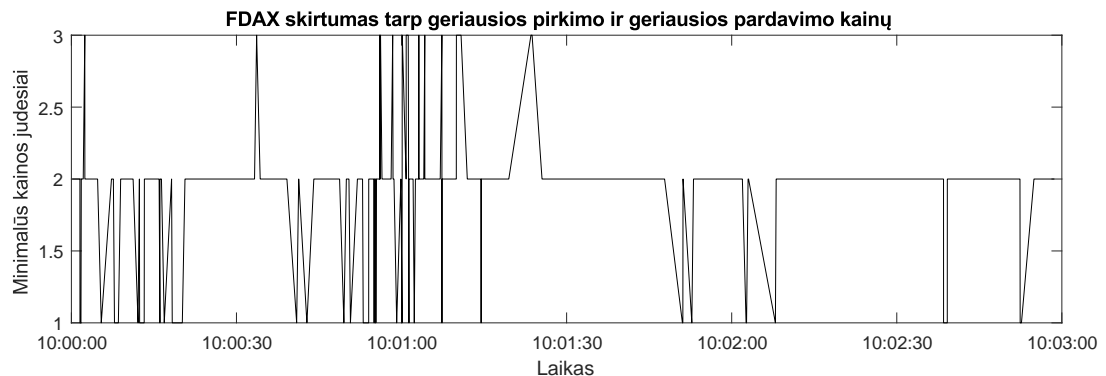
Parinkta 18 įvesčių, kurios naudotos prognozavimui su DNT ir SM:

- Paskutinis žinomas siūlomas kiekis ties geriausia pardavimo kaina.
- Paskutinis žinomas siūlomas kiekis ties geriausia pirkimo kaina.
- Skirtumas tarp paskutinės žinomos pardavimo ir paskutinės žinomos pirkimo kainų („spreadas“). Tai svarbus dydis. Kadangi „spreadas“ linkęs nusistovėti, bet kokie jo nukrypimai į vieną ar kitą pusę indikuoja apie tikėtiną grįžimą prie vidurkio artimu laiku, tai iliustruoja 5 pav.
- Paros laikas, suskirsčius parą minutėmis nuo 0 iki 1439. Akivaizdu, jog kintamumas priklauso nuo paros laiko, jis linkęs būti didesniu rinkos atsidarymo ir uždarymo momentais. Dieną kintamumas taip pat didesnis lyginant su nakties periodu. Esant didesniai kintamumui kaina labiau linkusi judėti į vieną ar kitą pusę.
- Biržoje fiksuota apyvarta per paskutinę minutę. Parodo rinkos aktyvumą.
- Biržoje fiksuota apyvarta per paskutines 5 minutes. Parodo rinkos aktyvumą.
- Skirtumas tarp aukščiausios prašytos ir žemiausios siūlytos kainų per paskutinę minutę t.y. kainos amplitudė (angl. range). Dažnai amplitudė nusakomas rinkos kintamumas.
- Skirtumas tarp aukščiausios prašytos ir žemiausios siūlytos kainų per paskutines 5 minutes. Analogiška kainos amplitudė tik šiek tiek ilgesnio periodo.
- Paskutinė žinoma kaina minus kaina prieš minutę. Parodo kaip pasikeitė kaina per paskutinę minutę t.y. nusako kryptį (teigiamą ar neigiamą) ir pokyčio dydį.
- Paskutinė žinoma kaina minus kaina prieš 5 minutes. kainos kryptis šiek tiek ilgesniu periodu.
- Šarpo santykis⁴ suskaičiuotas paskutinės minutės kainos pokyčiams. Šis rodiklis įtraukia visus paskutinės minutės kainų judesius.
- Šarpo santykis suskaičiuotas paskutinių 5 minučių kainos pokyčiams.
- Vidurkis paskutinių žinomų siūlomų kiekių ties geriausia pardavimo kaina per paskutinę minutę.
- Vidurkis paskutinių žinomų siūlomų kiekių ties geriausia pirkimo kaina per paskutinę minutę.
- Vidurkis paskutinių žinomų siūlomų kiekių ties geriausia pardavimo kaina per paskutines 5 minutes.

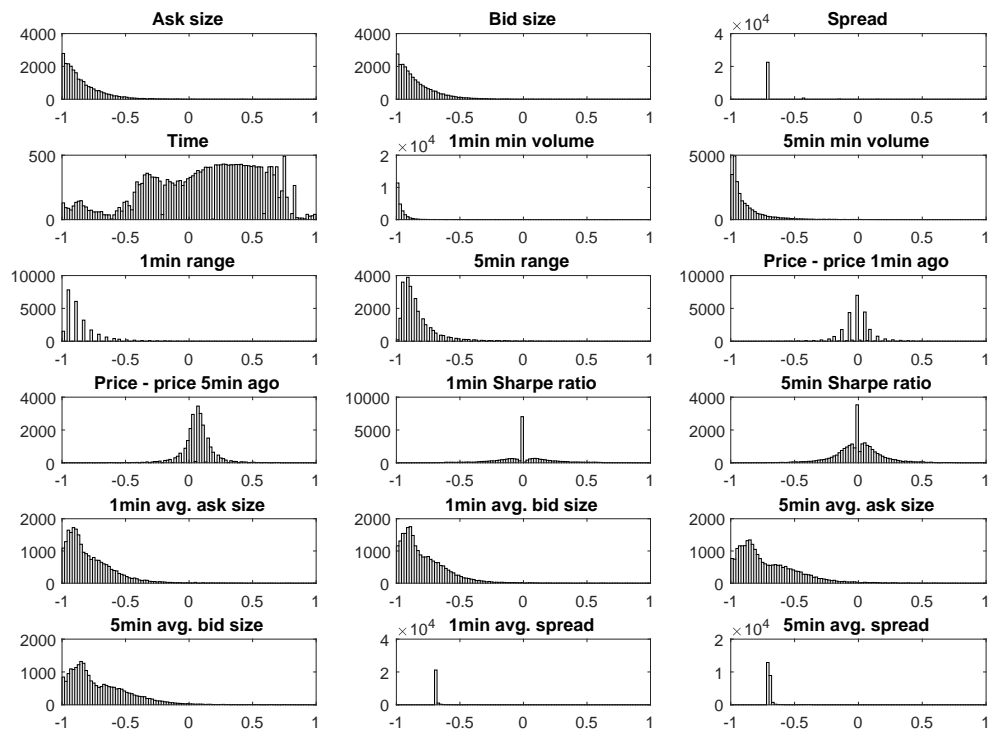
⁴Angl. Sharpe ratio. Pasiūlytas William F. Sharpe skaičiuoja investicijos grąžos santykį atsižvelgiant į riziką. Rodiklis vidutinę grąžą dalina iš standartinio nuokrypio. Jei dviejų investicijų vidutinė grąža sutampa, aukštesnį Šarpo santykį gaus turinti mažesnę variaciją.

- Vidurkis paskutinių žinomų siūlomų kiekių ties geriausia pirkimo kaina per paskutines 5 minutes.
- „Spredo“ vidurkis per paskutinę minutę. Turint tam tikro paskutinio periodo vidurkį prognozuoti „spredo“ kartu ir kainos judesį turėtų būti paprasčiau.
- „Spredo“ vidurkis per paskutines 5 minutes.

Apmokant DNT visos įvestys transformuotos į intervalą $[-1; 1]$ siekiant, kad neuroninio tinklo prognozėms įtakos nedarytų per didelių įvesties reikšmių iškreipiama aktyvacijos funkcija.



5 pav. DAX akcijų indekso ateities sandorio „spredas“ 2016 rugpjūčio 31 dieną nuo 10:00 iki 10:03. Pavyzdyje matoma, jog „spredas“ linkęs nusistovėti ties 2 minimaliais kainos judesiais (MKJ), tačiau trumpais periodais sumažėja iki 1 MKJ arba pakyla iki 3 MKJ.



6 pav. ES įvesčių histograma po transformacijos į intervalą $[-1; 1]$

5.3. Sandorių simuliacijų aprašymas

Šiame poskyryje supažindinama su situacijomis, kurios nagrinėjamos praktiniame tyrime. Situacijos simuliuotos atskirai kiekvieno iš instrumentų atveju. Jų prekybos istorinė informacija suskirstyta į 15 minučių intervalus. Kiekviename iš intervalų generuotas atsitiktinis laiko momentas tapęs simuliuotų sandorių pradžios tašku. Kiekvienu laiko momentu išsaugomos 5.2 apibūdintos įvestys ir simuliuojami du sandoriai: rinkos įsakymas ir atidėtas pavidimas. Išsaugoti praslydimų kaštai abiejų sandorių tipų atveju. Simuliacijos atliktos tiek pirkimo, tiek pardavimo sandoriams.

Duomenų paruošimui ir sandorių simuliacijoms nebuvo naudojamos jokių išorinių programų bibliotekų ar įrankių. Programuota „Matlab“ programavimo aplinkoje.

5 lentelėje pateikta kiek situacijų simuliuota kiekvienam instrumentui. Tai reiškia, jog pvz. ES atveju buvo atlikta 24018 rinkos įsakymų ir toks pats skaičius analogiškais laiko momentais pradėtų vykdyti atidėtųjų pavidimų. Tai pakartota pirkimo ir pardavimo atveju. Simuliuotų situacijų skaičius svyruoja dėl keletos priežasčių. Pirmiausiai skiriasi turima duomenų istorija. Taip pat instrumentų aktyvios prekybos periodo ilgis yra nevienodas.

5 lentelė. Simuliuotų situacijų skaičius kiekvienam nagrinėtam instrumentui.

Instrumentas	Simuliuotų situacijų kiekis
ES	31075
TY	33029
CL	44178
FV	26258
GC	36420
EC	42598
C	10412
FGBL	28855
VX	15127
FDAX	29690

Kadangi tyrimas paremtas istorinių duomenų simuliacija, tenka formuoti tokias prielaidas ir apribojimus:

1. **Sandorio dydis visuomet 1 kontraktas.** Tokia prielaida daroma dėl dviejų priežasčių. Pirmiausiai šiuo darbu nesiekama ištirti įtakos rinkai, todėl pasirinkus mažiausią galimą sandorio dydį minimizuojama ir galima įtaka. Antra priežastis, nemažai smulkesnių prekeivių ar kitų rinkos agentų dėl gana didelių ateities kontraktų dydžių pinigine išraiška⁵, apsiriboja prekiaudami tik po 1 ateities kontraktą per sandorį. Šis tyrimas orientuotas į tokius prekeivius.
2. **Simuliuojant sandorius nedaroma įtaka rinkai.** Dėl mažos prekybos apimties tariama, kad sandorių vykdymo algoritmas niekaip neveikia rinkos.
3. **Sandoriai rinkoje vykdomi be uždelsimo.** Netaikomas uždelsimas (angl. lag), simuliacija pradedama vykdyti jau nuo kito rinkos knygos pasikeitimo, po atėjusio prekybos signalo

⁵Pavyzdžiui 1 gamtinių dujų ateities kontraktas CME biržoje vertas apie 30000\$, o rezervuojamas užstatas prekeivio sąskaitoje (marža) yra 1320\$, 2018 metų Birželio kainomis.

pradžios. Susijusioje literatūroje ši problema nagrinėjama mažai ir yra individuali kiekvieno prekeivio atveju.

4. P_{laukta} (žr. lygtį 2.1) prilyginama prekybos signalo gavimo metu geriausia rinkos knygoje siūloma pirkimo kaina (pirkimo atveju) ir geriausia rinkos knygoje siūloma pardavimo kaina (pardavimo atveju).
5. **Rinkos įsakymas bus įvykdytas sekančiu rinkos knygos pasikeitimu.** Pirkimo rinkos įsakymas bus įvykdytas sekančią rinkos pasikeitimą, geriausia pardavimo kaina. Analogiškai pardavimo rinkos įsakymas bus įvykdytas sekančią rinkos pasikeitimą, geriausia pirkimo kaina.
6. **Atidėtasis pavedimas bus įvykdytas jei jo kainą pasieks priešinga rinkos knygoje siūloma kaina.** Tai reiškia pirkimo atidėtasis pavedimas kaina P bus įvykdytas už kainą P jei geriausia rinkoje siūloma pardavimo kaina pasieks arba nukris žemiau kainos P . Analogiškai pardavimo atidėtasis pavedimas kaina P bus įvykdytas už kainą P jei geriausia rinkoje siūloma pirkimo kaina pasieks arba pakils aukščiau kainos P .
7. **Atidėtasis pavedimas tap pat bus įvykdytas, jei įvykusių sandorių kiekis rinkoje ta pačia kaina pasieks priskirtą sandoriui eilės numerį.** Pavyzdžiui į rinką nusiųstas atidėtas pavedimas pirkti arba parduoti už kainą P . Rinkoje ta kaina jau pateikta k vienetų, todėl nusiųstam sandoriui priskiriamas $k + 1$ eilės numeris. Jei bet kuriuo sandorio vykdymo momentu sukaupta apyvarta ties kaina P pasieks $k + 1$, tuomet laikoma, kad atidėtasis pavedimas įvykdytas už kainą P .

1 algoritmas apibūdina sandorių vykdymo simuliacijų algoritmą pseudokodu. Algoritmas simuliuoja tiek rinkos įsakymus, tiek atidėtuosius pavedimus. Algoritmui įvedama rinkos informacija: laiko bei datos, geriausių siūlomų pirkimo ir pardavimo kainų, kiekių, paskutinio sandorio kainų ir kiekių masyvai. Nurodomas momentas nuo kurio pradėti simuliaciją ir laikas t . Jei laikas t lygus 0 algoritmas vykdo rinkos įsakymą, kitu atveju būna vykdomas atidėtasis pavedimas ir t indikuoja kiek laiko atidėtasis pavedimas gali būti aktyvus. Algoritmo išvestis yra sandorio praslydimas t.y. sandorio vykdymo kaštai.

Pirmiausiai vykdymo eigoje išsaugomas sandorio vykdymo pradžios momentas ir pasižymima P_{laukta} remiantis aukščiau aprašyta 4 prielaida (1-2 eilutės). Rinkos įsakymo atveju taikoma 5 prielaida (5 eilutė). Jei vykdomas atidėtasis pavedimas pasižymima pavedimo eilė biržoje pagal 7 prielaidą (7 eilutė). Kol neviršijamas laikas t tikrinamos 6 ir 7 prielaidos (10-19 eilutės). Jei atidėtasis pavedimas nebuvo įvykdytas per laiką t , jis verčiamas rinkos įsakymu (22 eilutė). Galiausiai algoritmas apskaičiuoja simuliacijos pagrindu gautą praslydimą (28 eilutė).

Pirkimo atveju simuliacijų algoritmas esmingai nesiskiria nuo pardavimo atvejo ir tėra tik priešingas 1 algoritmo variantas, jo pseudokodas pateiktas A priede.

1 algoritmas. Sandorių vykdymo simuliacijos algoritmas (pardavimo atvejis)

Ivestis: $mTime$ # Biržos laikas
 $bestAsk, bestBid$ # Geriausios siūlomų pardavimo, pirkimo kainų istoriniai duomenys
 $askSize, bidSize$ # Kiekiai ties geriausiomis siūlomomis pardavimo, pirkimo kainos
 $lastTrade$ # Paskutinio biržoje įvykusio sandorio kaina
 $lastTradeSize$ # Paskutinio biržoje įvykusio sandorio dydis
 idx # Indeksas, nurodantis momentą kai sandoris išsiųstas į rinką
 t # Laikas kiek atidėtasis pavedimas bus aktyvus

Išvestis: $Praslydimas$

```
1:  $firstIdx \leftarrow idx$ 
2:  $P_{laukta} \leftarrow bestAsk(firstIdx)$ 
3:
4: if  $t = 0$  then # Rinkos įsakymo atvejis
5:    $P_{gauta} \leftarrow bestBid(idx + 1)$ 
6: else # Atidėtojo pavedimo atvejis
7:    $queuePlace \leftarrow askSize(idx) + 1$ 
8:    $idx \leftarrow idx + 1$ 
9:
10:  while  $mTime(idx) - mTime(firstIdx) < t$  and  $queuePlace > 0$  do
11:    if  $P_{laukta} \leq bestBid(idx)$  then
12:       $queuePlace \leftarrow 0$ 
13:      break
14:    end if
15:    if  $P_{laukta} = lastTrade(idx)$  then
16:       $queuePlace \leftarrow queuePlace - lastTradeSize(idx)$ 
17:    end if
18:     $idx \leftarrow idx + 1$ 
19:  end while
20:
21:  if  $queuePlace > 0$  then # Jei nebuvo įvykdyta per laiką  $t$ 
22:     $P_{gauta} \leftarrow bestBid(idx)$ 
23:  else # Pavyko įvykdyti atidėtąjį pavedimą
24:     $P_{gauta} \leftarrow P_{laukta}$ 
25:  end if
26: end if
27:
28:  $Praslydimas \leftarrow P_{laukta} - P_{gauta}$ 
```

5.4. Prognozuojamų įvykių aprašymas

Tiek dirbtiniai neuroniniai tinklai, tiek sprendimų medžiai gavę 5.2 poskyryje aprašytas įvestis prognozuoja vieną iš trijų galimų baigčių:

- Kaina periodo gale t bus pakilusi.
- Kaina periodo gale t liks nepasikeitusi.

- Kaina periodo gale t bus nukritusi.

Norint nepadaryti tyrimo per daug sudėtingu, tai pat dėl ribotų skaičiavimo pajėgumų nuspręsta nagrinėti tik vieną atvejį kai $t = 60s$ t.y. dirbtinis neuroninis tinklas ir sprendimų medis prognozuos kainos kryptį po 1 minutės. Tokio periodo pasirinkimas motyvuotas tuo, jog fiksuojant $t = 60$ galimos baigtys (kaina kils, nepakis, kris) pasiskirsto beveik vienodai. Pavyzdžiui ES instrumentui prognozavimo klasės pasiskirsto beveik idealiai (32,8%; 33%; 34,2%).

Jei periodą trumpintume, didėtų nepakitusių kainos atvejų skaičius, prognozavimas sudėtingėtų, klasės būtų neproporcingos. Jei ilgintume – beveik neliktų nepakitusių kainų. Tiesa ilginimo atveju išsisukti iš nevienodų proporcijų prognozavimo klasių situacijos būtų galima priskyrus mažus kainų pasikeitimus į nepakitusių kainos klasę. Dar vienas ilgesnio periodo trūkumas tikėtina būtų prognozavimo tikslumo mažėjimas. Kadangi praeities informacija su laiku tampa vis nereikšmingesnė ateities prognozavimui.

Sandorių vykdymo algoritmas universalus, todėl tyrimus nesunkiai būtų galima pakartoti ir kitiems laiko periodams t , tačiau tai paliekama ateities darbams.

Prognozavimo tikslumas vertinamas trimis metrikomis:

- **Vidutinė kvadratinė paklaida (MSE).** Šis dydis parodo kaip tiksliai DNT ir SM prognozuoja kainos kryptį.
- **Spėjimo tikslumas.** Procentinė teisingai klasifikuotų kainų pokyčių išraiška.
- **Vidutinis prasydimas.** Simuliuotų sandorių kaštų vidurkis, skaičiuojamas 1 ir 2 algoritmais.

Nagrinėjant pardavimo sandorius ir DNT arba SM prognozuojant kainos kilimą po 60s, sandorių algoritmui bus nurodoma vykdyti sandorį su parametru $t = 60s$ t.y. vykdyti atidėtąjį pavedimą, kuris aktyvus 60s. Motyvuojant tuo, kad ateityje kaina bus aukštesnė ir bus galima parduoti už brangiau. Kitu atveju, jei prognozuojamas kainos kritimas arba nesikeitimas, paduodamas $t = 0s$ t.y. nieko nelaukti ir vykdyti rinkos įsakymą. Kadangi prognozuojamas kainos kritimas arba nesikeitimas ir tikėtina bus galima parduoti už pigiau.

Pirkimo sandorių atveju, kuomet DNT arba SM prognozuojant kainos kritimą po 60s, sandorių algoritmui bus nurodoma vykdyti sandorį su parametru $t = 60s$ t.y. vykdyti atidėtąjį pavedimą, kuris aktyvus 60s. Motyvuojant tuo, kad ateityje kaina bus žemesnė ir bus galima nupirkti už pigiau. Kitu atveju, jei prognozuojamas kainos kilimas arba nesikeitimas, paduodamas $t = 0s$ t.y. nieko nelaukti ir vykdyti rinkos įsakymą. Kadangi prognozuojamas kainos kilimas arba nesikeitimas ir tikėtina bus galima nupirkti už brangiau.

Situacijas geriau padeda suprasti 6 lentelė.

6 lentelė. Kaip elgsis sandorių vykdymo algoritmas skirtingais kainos prognozės atvejais.

Prognozė Sandorio kryptis	Kaina kils	Kaina nesikeis	Kaina kris
Pardavimas	Atidėtasis pavedimas	Rinkos įsakymas	Rinkos įsakymas
Pirkimas	Rinkos įsakymas	Rinkos įsakymas	Atidėtasis pavedimas

6. Neuroninio tinklo taikymo praktinis tyrimas

Šiame skyriuje ieškoma optimalių parametru apmokant dirbtinį neuroninį tinklą. Siekiama kiek galima sumažinti vidutinį praslydimą, tačiau atsižvelgiama ir į vidutinės kvadratinės paklaidos (MSE) bei kainos pokyčio klasifikavimo tikslumo rodiklius. Tyrimuose neuroninio tinklo architektūrai realizuoti pasinaudota „Matlab“ įskiepiu „Neural Network Toolbox“.

6.1. Neuronų skaičius paslėptame sluoksnyje

Pradžioje pasirinkta nagrinėti optimalų neuronų skaičių paslėptame sluoksnyje. Pasirinktas vienas paslėptas sluoksnis. Mokoma gradientų nusileidimo metodu. Klaidos funkcija – MSE. Mokymosi greitis fiksuotas 0,1. Inercijos konstanta (angl. momentum constant) parinkta 0,9. Mokymosi ciklų skaičius – 5000. Testinei imčiai palikta 20% duomenų imties (imčių dydžiai kiekvienam instrumentui pateikti 5 lentelėje). DNT mokyti kiekvienam instrumentui, pirkimo ir pardavimo atveju atskirai. Norint geriau palyginti prognozavimo vidutinę paklaidą, neuroninis tinklas buvo mokomas 5 kartus. DNT apmokymui, kaskart sumaišant, buvo atsitiktinai atrenkama 80% likusios mokymo imties (64% pradinės imties dydžio), o testinė imtis visada likdavo ta pati.

Testuoti neuronų skaičiai paslėptame sluoksnyje: 5, 10, 25, 50, 75, 100, 150, 200. Todėl tyrime iš viso apmokyti $10 * 2 * 5 * 8 = 800$ dirbtinių neuroninių tinklų.

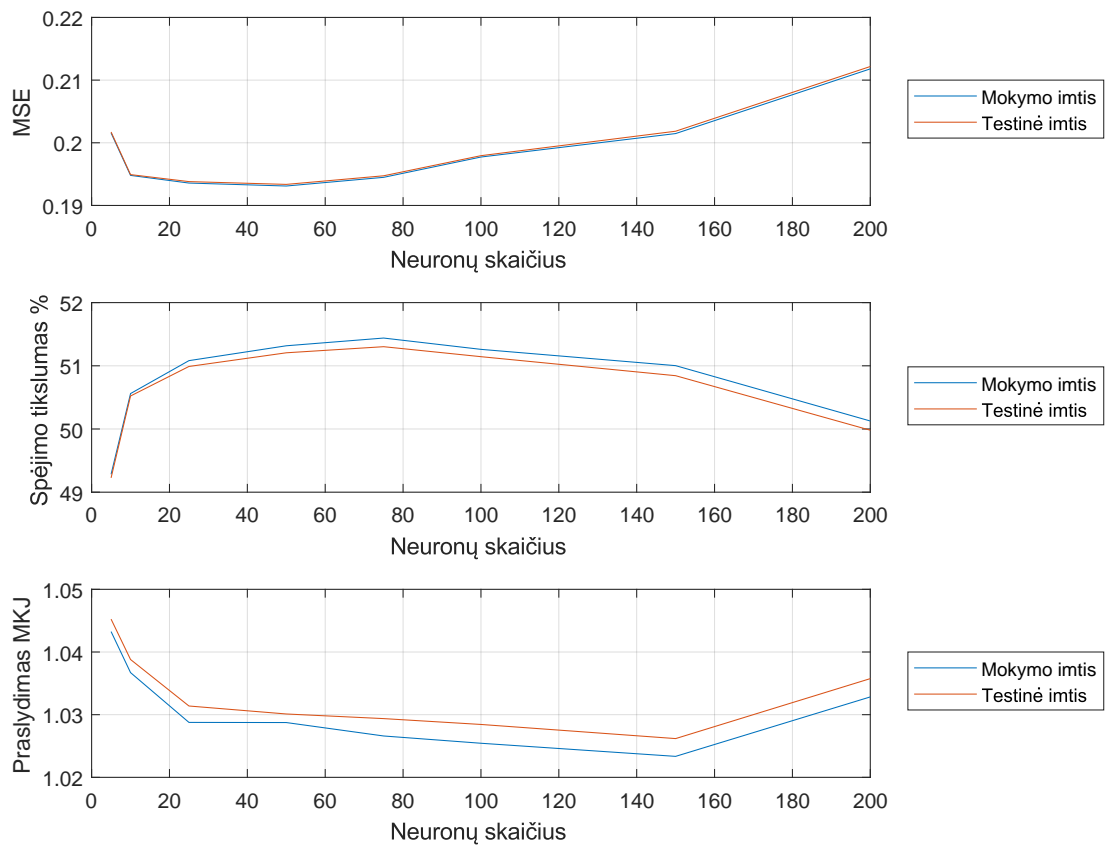
Testinės imties rezultatai per daug nesiskiria nuo mokymo imties rezultatų ir tarp jų matoma aiški koreliacija. Visų instrumentų rezultatų vidurkis pateiktas 7 paveiksle. Mažiausia vidutinė paklaida MSE pasiekama su 50 neuronų paslėptame sluoksnyje. MSE yra mažiausia ir nusistovi [10; 75] neuronų režyje. Akivaizdus MSE padidėjimas matomas mažinant neuronų skaičių paslėptame sluoksnyje (5 neuronai) ir didinant (daugiau kaip 75).

Apskritai didžiausias kainos krypties spėjimo tikslumas pasiekiamas ties 75 neuronais paslėptame sluoksnyje. Prognozės tiksluma stabiliausias ir įgyja didžiausias reikšmes neuronų skaičiui paslėptame sluoksnyje esant tarp 20 ir 100. Prognozavimo tikslumas smarkiai krenta mažinant neuronų skaičių iki 25 ir mažiau. Lėčiau, tačiau tikslumas mažėja ir didinant neuronų skaičių paslėptame sluoksnyje virš 100. Tikslumas skaitine išraiška siekia šiek tiek virš 51% ir artėja link 51,5%. Kadangi DNT klasifikuoja kainos pokytį trimis atvejais (pakilo, nepakito, nukrito), todėl spėjimo tikslumas procentine išraiška nukenčia. Lyginant su literatūroje nagrinėjamais darbais, kur prognozuojami tik du galimi variantai (kils arba kris).

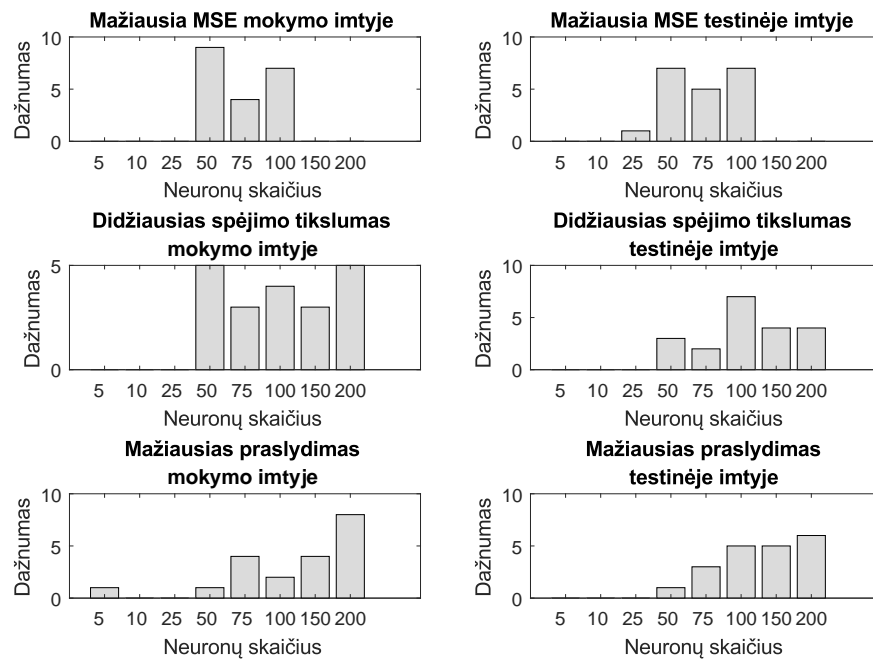
Kol neuronų skaičius paslėptame sluoksnyje neviršija 100, praslydimas koreliuoja neigiamai su spėjimo tikslumu ir teigiamai su klaidos funkcija MSE. Didėjant kainos prognozių tikslumui, mažėja praslydimas. Krentant MSE praslydimas irgi mažėja. Tačiau neuronų skaičiui pasiekus 150 koreliacija dingsta. Kitos metrikos prastėja, kai tuo tarpu praslydimas mažiausias. Ties 200 neuronų paslėptame sluoksnyje visos trys prognozavimo charakteristikos vėl elgiasi vienodai.

Didžiausias praslydimo pagerėjimas pastebimas didinant neuronų skaičių intervale [5; 25]. Po to praslydimas vis dar mažėja ir pasiekia piką ties 150 neuronų, prekybos kaštai vėl kyla didinant neuronų skaičių iki 200.

Nagrinėjant ne vidutinį atvejį, o geriausių rezultatų dažnumą, kurie pateikti 8 paveiksle, galima padaryti daugiau išvalgų. Mokymo klaida MSE visada būna mažiausia [50; 75] intervale. Nežymiai skiriasi MSE mažiausių pasikartojimų dažnumas testinėje imtyje. Tai iš esmės sutampa su vidutiniu atveju nagrinėtu 7 pav.



7 pav. Vidutinė MSE, vidutinis prognozavimo tikslumas ir vidutinis praslydimas visiems instrumentams, naudojant skirtingą neuronų skaičių paslėptame sluoksnyje.



8 pav. Histograma pagal geriausius rezultatus, žiūrint į neuronų skaičius paslėptame sluoksnyje. Viena histograma atspindi 20 atvejų, kiekvienam tiriamam instrumentui pardavimo ir pirkimo atvejais.

Žvelgiant į didžiausią kainos krypties spėjimo tikslumą iliustruojančią histogramą matoma, jog tikslumas vidutiniu atveju nesutampa su tikslumu paskiriems instrumentams. Nors mokymo imčiai vidutiniu atveju spėjimai teisingiausi kai neuronų skaičius intervale $[20; 100]$, tačiau pagal dažnį geriausias rezultatas svyruoja nuo 50 iki pat maksimalaus tirto kiekio – 200 neuronų. Testinei imčiai didžiausias spėjimo tikslumas šiek tiek panašesnis į vidurkį, tačiau vis tiek atsiranda atvejų, kuomet didesnę spėjimo tikslumą duoda daugiau (150 - 200) neuronų paslėptame sluoksnyje.

Pagal mažiausią praslydimą atskiriems instrumentams optimalus neuronų skaičius paslėptame sluoksnyje svyruoja nuo 50 iki 200. Daugiausiai reikšmių koncentruojasi intervale nuo 100 iki 200.

Apibendrinant šį tyrimą, galima daryti išvadą, jog apmokant DNT ir siekiant mažiausio praslydimo reikėtų naudoti apie 150 neuronų paslėptame sluoksnyje. Tai galioja tiek vidutiniu atveju, tiek nagrinėjant atskirų instrumentų atvejus. Tiesa, ši prielaida padaryta kai mokymo gretis fiksuotas ties 0,1, inercijos konstanta lygi 0,9 ir mokoma bent 5000 mokymo ciklų.

6.2. Mokymosi ciklų skaičius

Šiuo tyrimu siekiama rasti optimalų mokymosi ciklų skaičių. Dauguma pašalinių mokymo parametrų pasirinkti tokie patys kaip ir praėjusiame poskyryje aprašytame tyrime. Pasirinktas vienas paslėptas sluoksnis. Mokoma gradientų nusileidimo metodu. Klaidos funkcija – MSE. Mokymosi greitis fiksuotas 0,1. Inercijos konstanta (angl. momentum constant) parinkta 0,9. Neuronų skaičius paslėptame sluoksnyje – 150, motyvuojant 6.1 poskyryje atliktu tyrimu. Testinei imčiai vėl palikta 20% duomenų imties. DNT mokyti kiekvienam instrumentui, pirkimo ir pardavimo atveju atskirai. Norint geriau palyginti prognozavimo vidutinę paklaidą, neuroninis tinklas buvo apmokomas 5 kartus. DNT apmokymui, kaskart sumaišant, buvo atsitiktinai atrenkama 80% likusios mokymo imties (64% pradinės imties dydžio), o testinė imtis visada likdavo ta pati.

Testuotas mokymo ciklų skaičius: 50, 100, 200, 500, 1000, 2000, 5000, 10000, 15000, 20000, 25000.

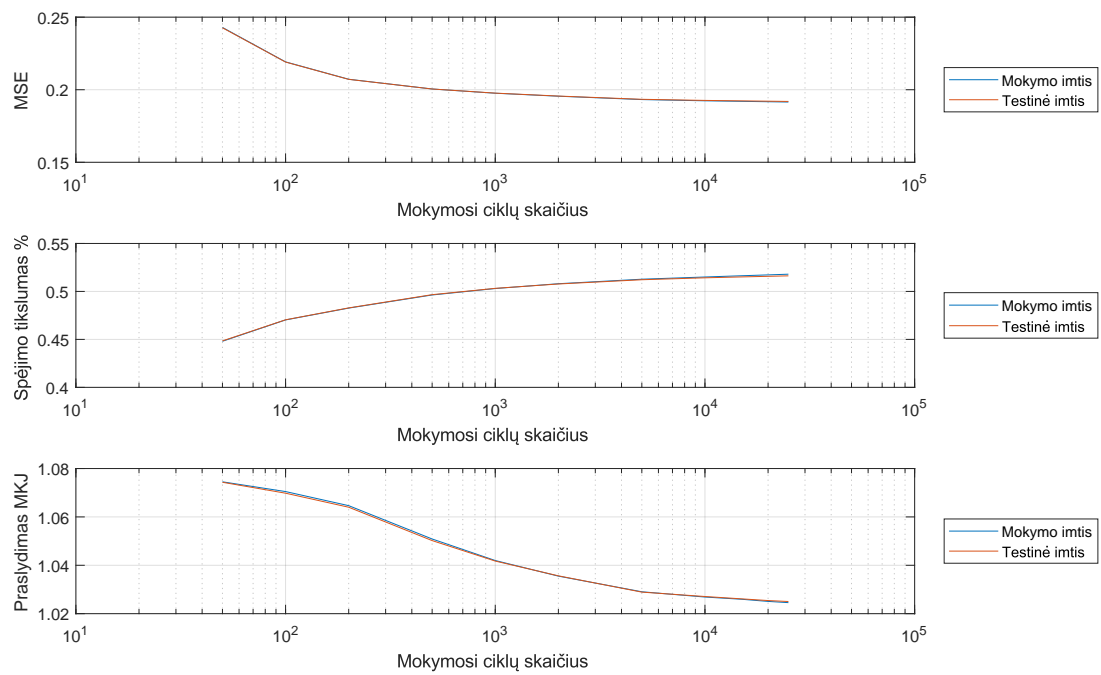
Rezultatai pateikti 9 pav., kur matoma MSE, spėjimų tikslumo ir praslydimo vidurkiai visiems instrumentams. 10 pav. rodo tų pačių charakteristikų pokytį didinant mokymosi ciklų skaičių.

Didžiausias MSE sumažėjimas vyksta tol, kol mokymosi ciklų skaičius nesiekia 1000. Po 10000 mokymosi ciklų MSE visiškai priartėja prie nulio.

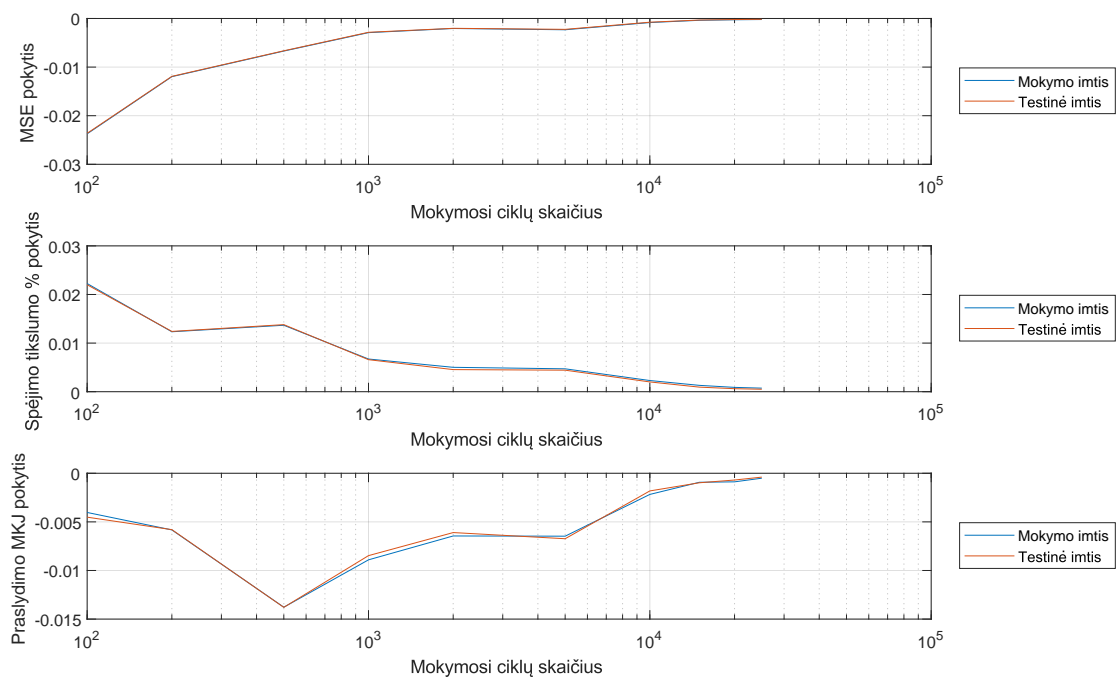
Spėjimų tikslumo padidėjimas gan chaotiškas ir nepastovus esant mažam mokymo ciklų skaičiui (iki 200). Mokymo ciklų skaičiui esant 200 ir didesniai spėjimų procentinis tikslumas lėtėjančiai, tačiau stabiliai auga.

Praslydimo kitimas irgi gana nepastovus esant mažam mokymo ciklų skaičiui (500 ir mažiau). Mokymo ciklų skaičiui pasiekus 500 kaštai stabiliai lėtėjančiai mažėja.

Apibendrinant šį tyrimą, galima daryti išvadą, jog DNT mokymosi epochų skaičius turi siekti bent 500, kad išnyktų chaotiškumas ir būtų pasiektas bent lokalus sprendinys. Didinant mokymosi skaičių stabiliai, tačiau palaipsniui lėtėjančiai gerėja ir prognozavimo rezultatai visoms charakteristikoms.



9 pav. Vidutinė MSE, vidutinis prognozavimo tikslumas ir vidutinis praslydimas visiems instrumentams, pritaikant skirtingą mokymosi ciklų skaičių.

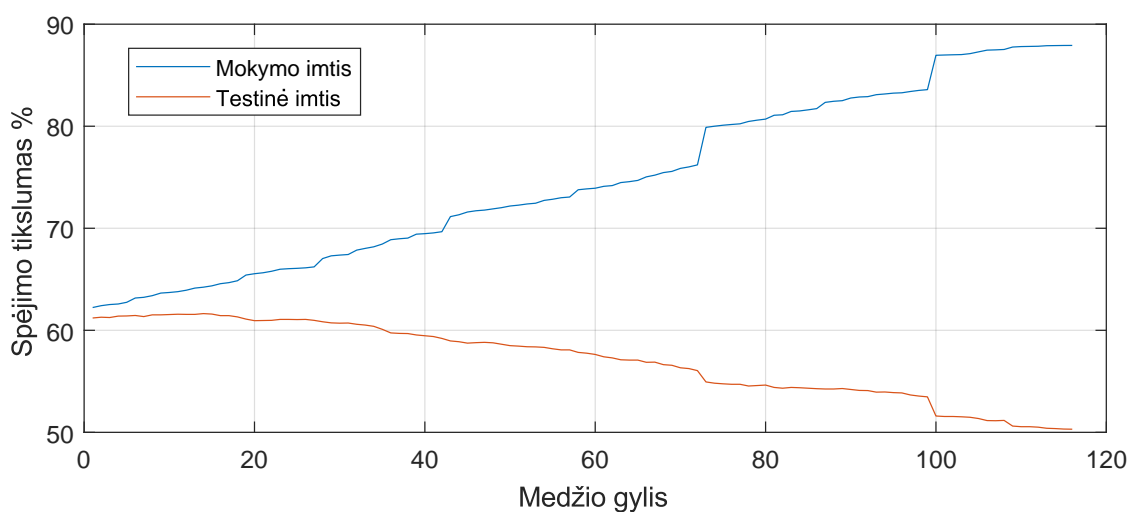


10 pav. MSE, prognozavimo tikslumo ir praslydimo pokyčiai didinant mokymosi ciklų skaičių.

7. Sprendimo medžio taikymo praktinis tyrimas

Šiame darbe mokyti sprendimo medį naudojami CART metodas. Siekiama sumažinti sandomų vykdymo kaštus vidutiniu atveju. Tyrime taip pat atsižvelgiama į vidutinę kvadratinę paklaidą (MSE) ir kainos pokyčio klasifikavimo tikslumą. Įvestys skirtos SM nebuvo niekaip normalizuotos. Sprendimo medis realizuotas pasinaudojant „Matlab“ įskiepiu „Statistics and Machine Learning Toolbox“.

Naudojant sprendimo medį klasifikavimui, pagrindinė užduotis yra rasti ribą nuo kurios jis pradeda prisitakyti prie mokymosi imties. Kitu atveju toks sprendimo medis bus bevertis prognozavimui. Paveiksle 11 pavaizduota kas vyksta leidus medžiui augti iki labai didelio gylio TY atveju. Galima, pastebėti, kad mokymo imties rezultatas niekada neprastėja augant medžiui, kai testinės imties – priešingai. Nuo tam tikro mokymo etapo testinės imties rezultatas pradeda nesustojamai prastėti dėl SM prisitaikymo prie testinės imties duomenų.



11 pav. Sprendimo medžio prisitakymo prie mokymo imties duomenų pavyzdys TY instrumentui.

7.1. Optimalių parametrų sprendimo medžiui tyrimas

Šiame tyrime bandoma optimizuoti 2 parametrus, kurių pagalba kontroliuojamas sprendimo medžio gylis ir kartu prisitaikymas prie mokymo imties. Pirmasis yra skaičius kiek daugiausiai kartų medis gali dalytis (angl. max number of splits). Patogumo dėlei jis pažymimas: *MaxNumSplits*. Šis kriterijus tiesiogiai fiksuoja maksimalų mazgų skaičių, o tai kartu kontroliuoja medžio gylį.

Antrasis – minimalus stebėjimų skaičius mazge (angl. min parents size). Sutrumpintai pavadintas: *MinParentSize*. Šis parametras mokant medį neleis sukurti mazgo, kurio šakoje bus mažiau negu tam tikras, nurodytas stebėjimų. Minimalus stebėjimų skaičius mazge netiesiogiai daro įtaką medžio dydžiui. Nustačius *MinParentSize* pakankamai didelį gali įvykti taip, kad mokymo algoritmas nebesugebės sukurti tokio naujo mazgo, kad jo šakose būtų daugiau negu *MinParentSize* stebėjimų ir turės baigti mokymo procesą. Priešingas ekstremumas būtų *MinParentSize* nustatymas santykinai mažu. Tuomet mazgo šakose galėtų būti ir labai nedidelis stebėjimų skaičius, o tai iš dalies veda prie medžio prisitaikymo mokymo imčiai.

Medžiams apmokyti naudotas CART metodas. Po apmokymo genėjimas nebuvo atliekamas.

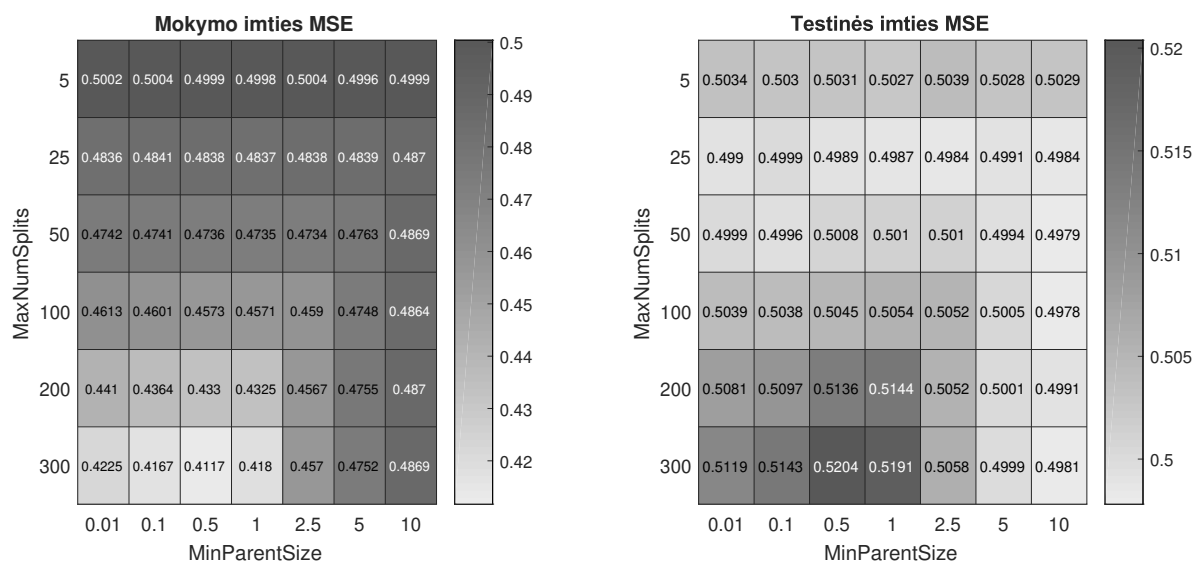
Mokymo ir testinės imtys padalintos 80%-20% santykiu. Kiekvienam instrumentui, pirkimo ir pardavimo atvejais apmokyta po 10 SM. Kiekvienu iš 10–ies atvejų iš mokymo imties atsitiktinai atrinkta 80% duomenų. Taip siekiama išlyginti atsitiktinumą ir ištirti vidutinį atvejį. Testinė imtis kiekvienu iš 10–ies mokymo atvejų likdavo ta pati.

MaxNumSplits parametras tirtas su tokiais reikšmėmis: 5, 25, 50, 100, 200, 300. Tai reiškia, kad sprendimų medis galėjo turėti nuo 5 iki 300 mazgų.

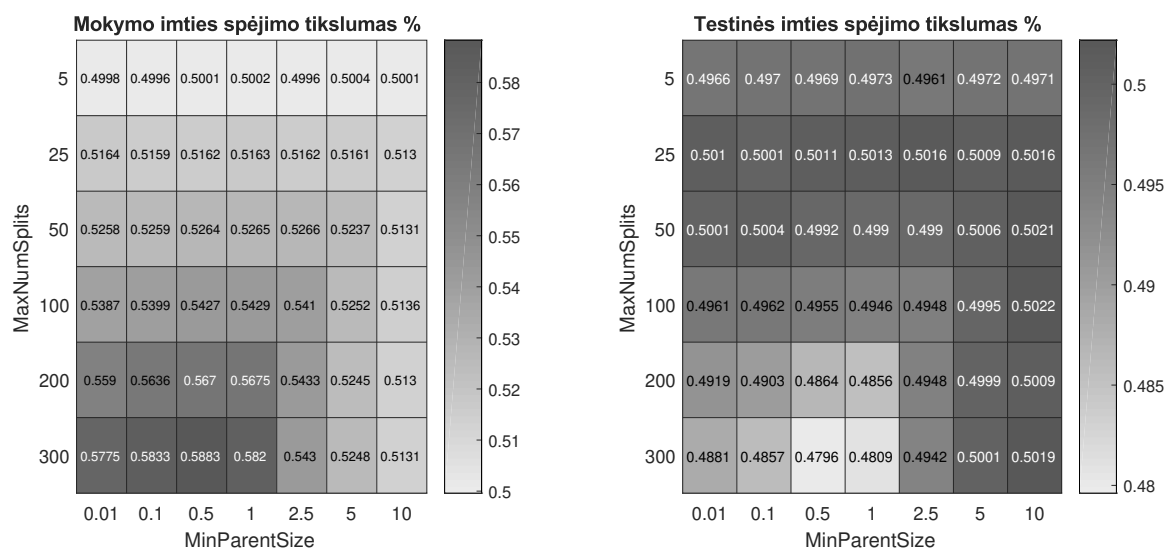
MinParentSize kriterijus turi būti nurodomas kaip konkretus skaičius. Stebėjimų skaičius kiekviename mazge privalės viršyti *MinParentSize*. Kadangi šiame tyrime imčių skaičius skirtingiems instrumentams nesutampa, tenka *MinParentSize* susieti su konkrečiu tyrimo duomenų kiekiu. Tai atlikta *MinParentSize* pavertus į procentinę išraišką nuo mokymo imties dydžio. Pavyzdžiui turint mokymo imtį lygią 1000, o *MinParentSize* parinkus 2, mažiausias galimas stebėjimų skaičius bet kuriame mazge bus 2% nuo 1000 t.y. 20. *MinParentSize* tyrime įgavo šias vertes: 0,01, 0,1, 0,5, 1, 2,5, 5, 10.

Tyrimo rezultatai pateikti 10-13 iliustracijose. 15 pav. pateiktas vidutinis sprendimo medžio gylis, jis svyruoja nuo maždaug 2 iki 50. 12 paveiksle matoma, jog MSE testinei imčiai mažiausia kai *MaxNumSplits* nedidelis ir yra apie 25-50 arba kai *MinParentSize* didelis 5-10. 13 pav. rodo, kad spėjimo tikslumas irgi didžiausias su tokiais pačiais parametrais.

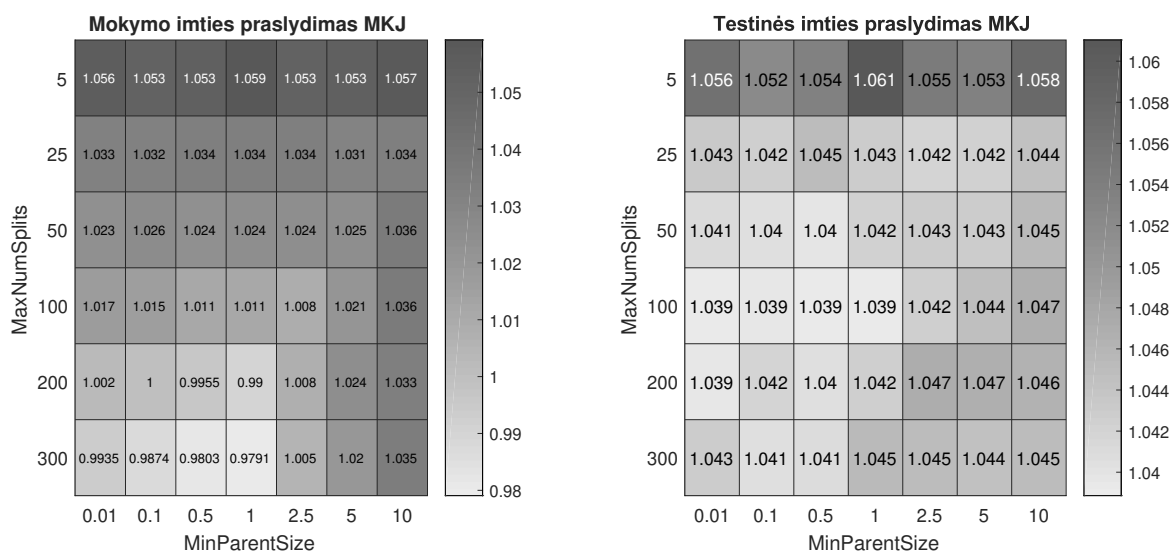
Kiek kitokia situacija nagrinėjant sandorių kaštų priklausomybę nuo tirtų parametrų (14 paveikslas). Nors nežymiai, bet rezultatai testinei imčiai geresni *MaxNumSplits* esant tarp 50 ir 200, o *MinParentSize* būnant nuo 0,01 iki 1. Kiek smulkiau šis intervalas panagrinėtas 16 pav. Matoma, jog optimalūs parametrai *MaxNumSplits* ir *MinParentSize* turėtų būti atitinkamai 125 ir 0,01.



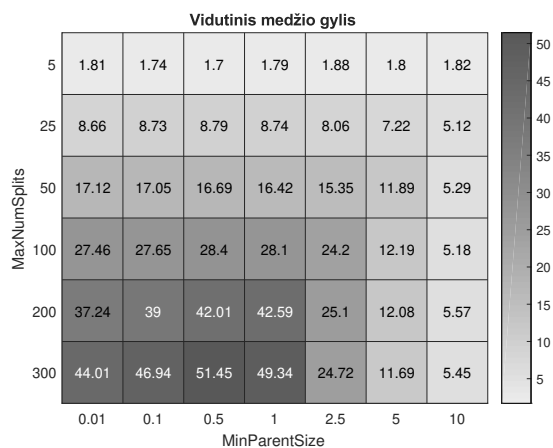
12 pav. Sprendimo medžio vidutinės kvadratinės paklaidos (MSE) priklausomybė nuo *MaxNumSplits* ir *MinParentSize*. Vidutinis atvejis visiems 10 finansinių instrumentų pirkimo ir pardavimo sandoriams. *MaxNumSplits* kinta nuo 5 iki 300, o *MinParentSize* nagrinėtas tarp 0,01 ir 10.



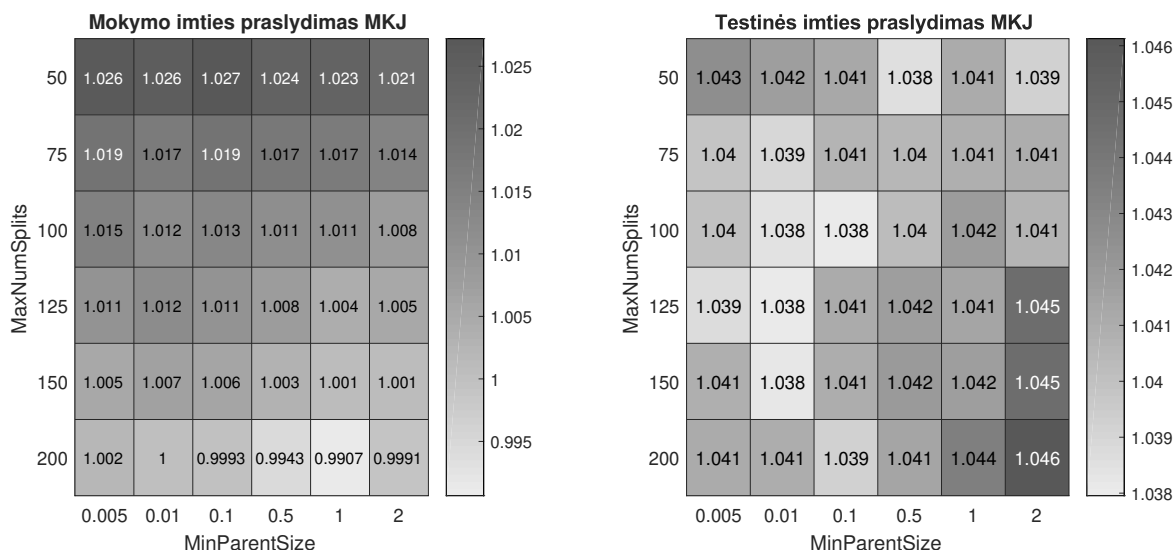
13 pav. Sprendimo medžio prognozavimo tikslumo priklausomybė nuo *MaxNumSplits* ir *MinParentSize*. Vidutinis atvejis visiems 10 finansinių instrumentų pirkimo ir pardavimo sandoriams. *MaxNumSplits* kinta nuo 5 iki 300, o *MinParentSize* nagrinėtas tarp 0,01 ir 10.



14 pav. Sandorių kaštų priklausomybė nuo *MaxNumSplits* ir *MinParentSize*, prognozavimui naudojant sprendimo medį. Vidutinis atvejis visiems 10 finansinių instrumentų pirkimo ir pardavimo sandoriams. *MaxNumSplits* kinta nuo 5 iki 300, o *MinParentSize* nagrinėtas tarp 0,01 ir 10.



15 pav. Sprendimo medžio gylis priklausomybė nuo *MaxNumSplits* ir *MinParentSize*. Vidutinis atvejis visiems 10 finansinių instrumentų pirkimo ir pardavimo sandoriams. *MaxNumSplits* kinta nuo 5 iki 300, o *MinParentSize* nagrinėtas tarp 0,01 ir 10.



16 pav. Sandorių kaštų priklausomybė nuo *MaxNumSplits* ir *MinParentSize*, prognozavimui naudojant sprendimo medį. Vidutinis atvejis visiems 10 finansinių instrumentų pirkimo ir pardavimo sandoriams. *MaxNumSplits* kinta nuo 50 iki 200, o *MinParentSize* nagrinėtas tarp 0,005 ir 2.

8. Dirbtinio neuroninio tinklo ir sprendimo medžio rezultatų palyginimas

Šiam galutiniam apibendrinimui ir rezultatų tarp skirtingų metodų palyginimui panaudoti naujai sugeneruoti imčių duomenys. Naujos imtys prieš tai nebuvo naudotos atliekant 6 ir 7 skyriuose aprašytus eksperimentus.

Visiems 10 finansinių instrumentų apmokyta po 2 sprendimų medžius ir po 2 dirbtinius neuroninius tinklus. „SM geriausias“ sprendimo medis apmokytas konkrečiam instrumentui su geriausias tam instrumentui gautais parametrais 7.1 tyrime, testinėje imtyje. „SM optimalus“ visiems instrumentams apmokytas sprendimo medis su tokiais pačiais parametrais: $MaxNumSplits = 125$ ir $MinParentSize = 0,01$.

Abu dirbtiniai neuroniniai tinklai mokyti su tais pačiais parametrais kaip ir tyrime 6.1. „DNT geriausias“ atveju dirbtinio neuroninio tinklo paslėptų neuronų skaičius sluoksnyje buvo parinktas konkrečiam instrumentui geriausią rezultatą rodžiusį rezultatą testinėje imtyje 6.1 tyrime. „DNT optimalus“ apmokytas su 150 neuronų paslėptame sluoksnyje.

Po apmokymo SM ir DNT testinių imčių praslydimo ir kainos krypties spėjimo tikslumų rezultatai pateikti 7 ir 8 lentelėse. 7 lentelėje apskaičiuotas procentinis kaštų sumažėjimas, jeigu būtų taikytas atitinkamas metodas vykdyti sandorį, užuot jį vykdžius rinkos įsakymu.

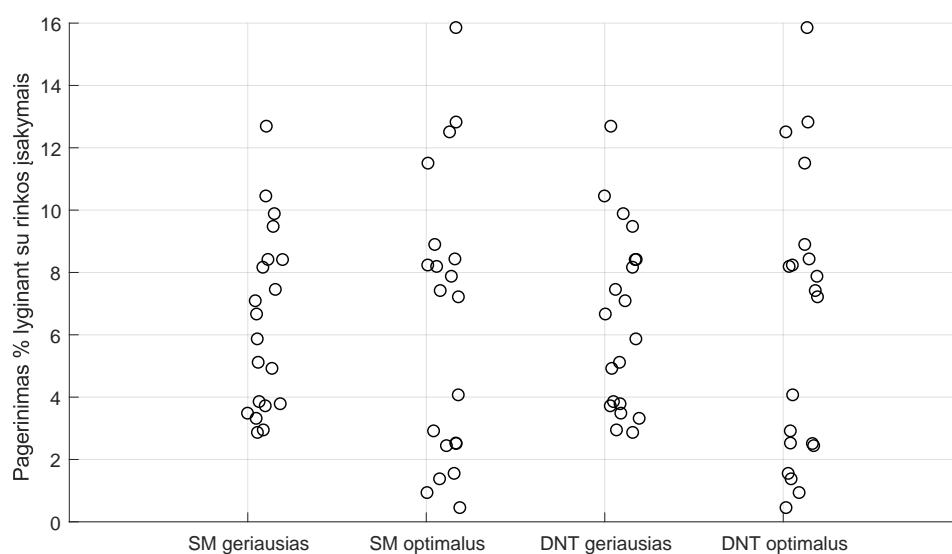
Iš rezultatų galima matyti, jog spredimų medis atsilieka nuo dirbtinio neuroninio tinklo tiek praslydimo sumažinimu, tiek teisingai klasifikuotų kainos pokyčių tikslume. Geriausiai tarp 4 metodų pasirodė „DNT optimalus“ apmokytas su 150 neuronų paslėptame sluoksnyje.

7 lentelė. Simuliuotų situacijų skaičius kiekvienam nagrinėtam instrumentui.

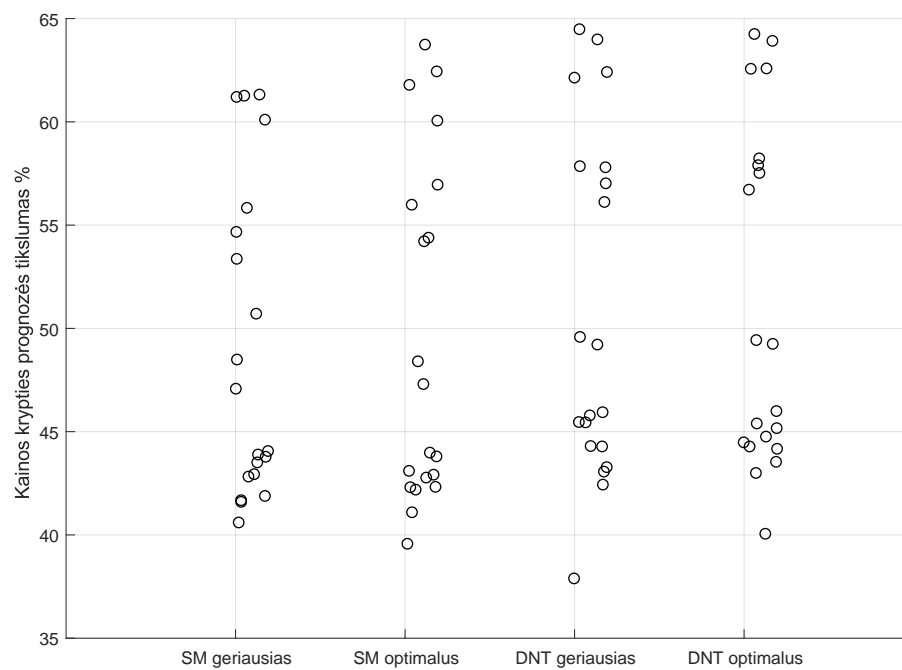
Trump.	Kryptis	SM geriausias	SM optimalus	DNT geriausias	DNT optimalus
ES	Pirkimas	6,64%	4,05%	7,79%	7,02%
ES	Pardavimas	8,14%	8,88%	8,07%	8,18%
TY	Pirkimas	3,46%	1,53%	1,42%	1,67%
TY	Pardavimas	3,70%	2,50%	1,01%	1,36%
CL	Pirkimas	7,43%	11,48%	14,32%	15,20%
CL	Pardavimas	4,90%	7,40%	10,23%	10,72%
FV	Pirkimas	5,09%	1,36%	2,21%	2,45%
FV	Pardavimas	2,85%	2,89%	1,27%	1,94%
GC	Pirkimas	5,85%	8,22%	13,37%	12,46%
GC	Pardavimas	9,45%	7,20%	9,09%	9,57%
EC	Pirkimas	7,07%	15,83%	17,03%	16,95%
EC	Pardavimas	10,43%	8,41%	3,43%	13,05%
C	Pirkimas	2,93%	0,92%	0,52%	1,02%
C	Pardavimas	3,83%	0,43%	1,58%	1,00%
FGBL	Pirkimas	8,39%	12,80%	13,36%	12,55%
FGBL	Pardavimas	12,67%	8,17%	13,02%	13,28%
VX	Pirkimas	3,30%	2,42%	2,09%	1,86%
VX	Pardavimas	3,77%	2,49%	2,46%	3,64%
FDAX	Pirkimas	8,40%	7,86%	7,77%	8,43%
FDAX	Pardavimas	9,86%	12,49%	11,64%	12,65%
Vidurkis	Pirkimas	5,86%	6,65%	7,99%	7,96%
Vidurkis	Pardavimas	6,96%	6,09%	6,18%	7,54%
Vidurkis	Bendrai	6,41%	6,37%	7,08%	7,75%

8 lentelė. Simuliuotų situacijų skaičius kiekvienam nagrinėtam instrumentui.

Trump.	Kryptis	SM geriausias	SM optimalus	DNT geriausias	DNT optimalus
ES	Pirkimas	43,86%	42,16%	45,76%	45,37%
ES	Pardavimas	43,75%	43,77%	45,91%	45,97%
TY	Pirkimas	60,07%	60,03%	62,38%	62,56%
TY	Pardavimas	61,18%	61,76%	62,11%	62,54%
CL	Pirkimas	43,48%	43,95%	45,42%	45,14%
CL	Pardavimas	42,91%	43,07%	45,44%	43,51%
FV	Pirkimas	54,64%	55,96%	57,77%	58,20%
FV	Pardavimas	55,80%	56,93%	57,82%	57,87%
GC	Pirkimas	44,03%	42,75%	42,41%	44,45%
GC	Pardavimas	42,80%	42,28%	44,25%	44,73%
EC	Pirkimas	41,65%	42,30%	44,27%	44,25%
EC	Pardavimas	41,86%	42,89%	37,86%	44,14%
C	Pirkimas	50,68%	54,19%	56,99%	57,49%
C	Pardavimas	53,34%	54,36%	56,09%	56,69%
FGBL	Pirkimas	40,57%	41,07%	43,03%	42,97%
FGBL	Pardavimas	41,57%	39,54%	43,25%	40,03%
VX	Pirkimas	61,29%	63,71%	64,45%	64,22%
VX	Pardavimas	61,24%	62,41%	63,96%	63,89%
FDAX	Pirkimas	48,46%	48,37%	49,19%	49,40%
FDAX	Pardavimas	47,05%	47,27%	49,55%	49,22%
Vidurkis	Pirkimas	48,87%	49,45%	51,17%	51,41%
Vidurkis	Pardavimas	49,15%	49,43%	50,62%	50,86%
Vidurkis	Bendrai	49,01%	49,44%	50,90%	51,13%



17 pav. 7 lentelės duomenų pasiskirstymas.



18 pav. 8 lentelės duomenų pasiskirstymas.

Išvados ir rekomendacijos

Šiame darbe apibrėžtas vertybinių popierių biržos veikimo modelis. Suformuota sandorių vykdymo problema. Atlikta susijusios literatūros analizė, kuri parodo, jog darbe nagrinėjamą problemą galima spręsti naudojant tiek tiesinius, tiek netiesinius prognozavimo metodus.

Kaštų sumažinimo uždavinys suvestas į kainos krypties prognozę trumpame periode. Bandyta prognozuoti tris galimus kainos pokyčius 1min bėgyje: kils, kris, nesikeis.

Tiriant dirbtinio neuroninio tinklo parametrus nustatyta, jog vidutiniu atveju geriausias prognozavimo tikslumas pasiekiamas su maždaug 150 neuronų paslėptame sluoksnyje.

Tyrinėjant dirbtinio neuroninio tinklo mokymosi ciklą skaičiaus įtaką prognozėms rezultatams parodyta, jog esant mažam mokymosi ciklą skaičiui prognozės nėra stabilios. Ryškesnis tikslumas pasiekiamas nuo maždaug 200-500 mokymo ciklą skaičiaus. Ciklą skaičius ištirtas iki 25000. Analizė parodė, jog pasiekus 10000 klaidos mažėjimo ir prognozavimo tikslumo greičiai gerokai sumažėja. Greičiausiai klaida krinta ir prognozavimas tikslėja kai mokymosi ciklą skaičius yra maždaug nuo 500 iki 10000.

Sprendimo medžio tyrimas parodė, jog su pasirinktais įvesties duomenimis įvyksta labai staigus prisitakymas prie mokymo imties. Testinės imties rezultatai visiškai nekoreliuoja su mokymo imties spėjimų tikslumu ir klaidos dydžiu.

Lyginant dirbtinį neuroninį tinklą ir sprendimo medį prieita išvada, jog dirbtinis neuroninis tinklas su šiame darbe su nagrinėjamu uždaviniu susitvako geriau.

DNT pasiekia apie 51% tikslumą prognozuojant kainos kryptį, kai sprendimo medžio rezultatas svyruoja ties 49%. Dirbtinis neuroninis tinklas geba sumažinti sandorių kaštus lyginant su rinkos įsakymu apie 7-8%. Sprendimų medis pasiekia maždaug 6-6,5% kaštų sumažėjimą.

Ateities tyrimų planas

Ateityje galima būtų atlikti platesnį tyrimą įvairiems prognozavimo periodams. Patikrinti kaip dirbtinio neuroninio tinklo ir sprendimo medžio prognozavimo tikslumas koreliuoja su skirtingu prognozės periodu. Patvirtinti ar paneigti hipotezę, jog kainos prognozės tikslumas mažėja didinant prognozės periodo ilgį.

Pritaikyti gilaus mokymo neurono tinklą neapdorojus įvesčių t.y. leisti jam pačiam rasti įvairias duomenų priklausomybes.

Kita kryptis galėtų būti optimalių įvesčių paieška. Tam galima pasitelkti genetinį algoritmą arba koreliacijų analizę.

Literatūros šaltiniai

- [1] Bahram Adrangi, Arjun Chatrath, Kanwalroop Kathy Dhanda, and Kambiz Raiee. Chaos in oil prices? evidence from futures markets. *Energy Economics*, 23(4):405–425, 2001.
- [2] Matthew N Anyanwu and Sajjan G Shiva. Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3(3):230–240, 2009.
- [3] Ahmed Bensaïda. Noisy chaos in intraday financial data: Evidence from the american index. *Applied Mathematics and Computation*, 226:258–265, 2014.
- [4] Milan Borkovec, Ian Domowitz, Brian Kiernan, and Vitaly Serbin. Portfolio optimization and the cost of trading. *The Journal of Investing*, 19(2):63–76, 2010.
- [5] Leo Breiman, Jerome Friedman, Charles J Stone, and RA Olshen. *Classification and Regression Trees*. CRC Press, 1984.
- [6] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.
- [7] Morton Glantz and Robert Kissell. *Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era*. Academic Press, 2013.
- [8] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [9] John C Hull and Sankarshan Basu. *Options, futures, and other derivatives*. Pearson Education India, 2016.
- [10] Monica Isfan, Rui Menezes, and Diana A Mendes. Forecasting the portuguese stock market time series by using artificial neural networks. In *Journal of Physics: Conference Series*, volume 221, page 012017. IOP Publishing, 2010.
- [11] Bekir Karlik and A Vehbi Olgac. Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122, 2011.
- [12] Alec N Kercheval and Yuan Zhang. Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance*, 15(8):1315–1329, 2015.
- [13] Ramon Lawrence. Using neural networks to forecast stock market prices. *University of Manitoba*, 333, 1997.
- [14] Blake LeBaron. Chaos and nonlinear forecastability in economics and finance. *Phil. Trans. R. Soc. Lond. A*, 348(1688):397–404, 1994.
- [15] Charles-Albert Lehalle, Othmane Mounjid, and Mathieu Rosenbaum. Optimal liquidity-based trading tactics. *arXiv preprint arXiv:1803.05690*, 2018.
- [16] Shiyu Liang and R Srikant. Why deep neural networks for function approximation? *arXiv preprint arXiv:1610.04161*, 2016.

- [17] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [18] David G McMillan. Non-linear predictability of uk stock market returns. *Oxford Bulletin of Economics and Statistics*, 65(5):557–573, 2003.
- [19] Joachim Nagel. Markets committee. 2016.
- [20] Binoy B Nair, VP Mohandas, and NR Sakthivel. A decision tree—rough set hybrid system for stock market trend prediction. *International Journal of Computer Applications*, 6(9):1–6, 2010.
- [21] YIJIN Pei. Neural network: A powerful tool for classification. In *Prepared for*, pages 703–708. Citeseer, 1995.
- [22] Edgar E Peters. *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*, volume 1. John Wiley & Sons, 1996.
- [23] Bo Qian and Khaled Rasheed. Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1):25–33, 2007.
- [24] Aistis Raudys, Saulius Blažiūnas, Linas Žvirblis, and Darius Plikynas. Atomic order execution tactics in futures markets. *Information Technology And Control*, 46(1):70–85, 2017.
- [25] Thilo A Schmitt, Desislava Chetalova, Rudi Schäfer, and Thomas Guhr. Non-stationarity in financial time series: Generic features and tail behavior. *EPL (Europhysics Letters)*, 103(5):58003, 2013.
- [26] Everton Silva, Douglas Castilho, Adriano Pereira, and Humberto Brandao. A neural network based approach to support the market making strategies in high-frequency trading. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 845–852. IEEE, 2014.
- [27] Justin Sirignano. Deep learning for limit order books. 2016.
- [28] Michael Small and Chi Kong Tse. Evidence for deterministic nonlinear dynamics in financial time series data. In *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pages 339–346. IEEE, 2003.
- [29] Chih-Fong Tsai, Yuah-Chiao Lin, David C Yen, and Yan-Min Chen. Predicting stock returns by classifier ensembles. *Applied Soft Computing*, 11(2):2452–2459, 2011.
- [30] Alfredo Vellido, Paulo JG Lisboa, and J Vaughan. Neural networks in business: a survey of applications (1992–1998). *Expert Systems with applications*, 17(1):51–70, 1999.
- [31] Linas Žvirblis. Ateities sandorių vykdymo analizė. *Mokslo tiriamojo darbo projektas*. 2018.
- [32] Jar-Long Wang and Shu-Hui Chan. Stock market trading rule discovery using two-layer bias decision tree. *Expert Systems with Applications*, 30(4):605–611, 2006.
- [33] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

- [34] Muh-Cherng Wu, Sheng-Yu Lin, and Chia-Hsin Lin. An effective application of decision tree to stock trading. *Expert Systems with Applications*, 31(2):270–274, 2006.
- [35] Seung-Ryong Yang and B Wade Brorsen. Nonlinear dynamics of daily futures prices: conditional heteroskedasticity or chaos? *Journal of Futures Markets*, 13(2):175–191, 1993.
- [36] Ban Zheng, Eric Moulines, and Frédéric Abergel. Price jump prediction in limit order book. *arXiv preprint arXiv:1204.1381*, 2012.

Priedai

A. Sandorių vykdymo simuliacijų algoritmas pirkimo atvejui

2 algoritmas. Sandorių vykdymo simuliacijos algoritmas (pirkimo atvejis)

Išvestis: $mTime$ # Biržos laikas
 $bestAsk, bestBid$ # Geriausios siūlomų pardavimo, pirkimo kainų istoriniai duomenys
 $askSize, bidSize$ # Kiekiai ties geriausiomis siūlomomis pardavimo, pirkimo kainos
 $lastTrade$ # Paskutinio biržoje įvykusio sandorio kaina
 $lastTradeSize$ # Paskutinio biržoje įvykusio sandorio dydis
 idx # Indeksas, nurodantis momentą kai sandoris išsiųstas į rinką
 t # Laikas kiek atidėtasis pavedimas bus aktyvus

Išvestis: $Praslydimas$

```
1:  $firstIdx \leftarrow idx$ 
2:  $P_{laukta} \leftarrow bestBid(firstIdx)$ 
3:
4: if  $t = 0$  then # Rinkos įsakymo atvejis
5:    $P_{gauta} \leftarrow bestAsk(idx + 1)$ 
6: else # Atidėtojo pavedimo atvejis
7:    $queuePlace \leftarrow bidSize(idx) + 1$ 
8:    $idx \leftarrow idx + 1$ 
9:
10:  while  $mTime(idx) - mTime(firstIdx) < t$  and  $queuePlace > 0$  do
11:    if  $P_{laukta} \geq bestAsk(idx)$  then
12:       $queuePlace \leftarrow 0$ 
13:      break
14:    end if
15:    if  $P_{laukta} = lastTrade(idx)$  then
16:       $queuePlace \leftarrow queuePlace - lastTradeSize(idx)$ 
17:    end if
18:     $idx \leftarrow idx + 1$ 
19:  end while
20:
21:  if  $queuePlace > 0$  then # Jei nebuvo įvykdyta per laiką  $t$ 
22:     $P_{gauta} \leftarrow bestAsk(idx)$ 
23:  else # Pavyko įvykdyti atidėtąjį pavedimą
24:     $P_{gauta} \leftarrow P_{laukta}$ 
25:  end if
26: end if
27:
28:  $Praslydimas \leftarrow P_{gauta} - P_{laukta}$ 
```
