



**KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

Erika Mekšunaitė

**EUROPOS ŠALIŲ AKCIJŲ KAINŲ GRAŽŲ ANALIZĖ IR
PROGNOZĖ**

Baigiamasis magistro projektas

Vadovai

Doc. dr. Kristina Lukoševičiūtė

Doc. dr. Alina Stundžienė

KAUNAS, 2017

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

**EUROPOS ŠALIŲ AKCIJŲ KAINŲ GRAŽŲ ANALIZĖ IR
PROGNOZĖ**

Baigiamasis magistro projektas
Didžiųjų verslo duomenų analitika (kodas 621G12002)

Vadovai

Doc. dr. Kristina Lukoševičiūtė
Doc. dr. Alina Stundžienė

Recenzantai

Dr. Mindaugas Kavaliauskas
Doc. dr. Aura Drakšaitė

Projektą atliko

Erika Mekšunaitė

KAUNAS, 2017



KAUNO TECHNOLOGIJOS UNIVERSITETAS

(Fakultetas)

(Studento vardas, pavardė)

(Studijų programos pavadinimas, kodas)

„Europos šalių akcijų kainų gražų analizė ir prognozė“ AKADEMINIO SAŽININGUMO DEKLARACIJA

20 ____ m. _____ d.
Kaunas

Patvirtinu, kad mano, **Erikos Mekšunaitės**, baigiamasis projektas tema „Europos šalių akcijų kainų gražų analizė ir prognozė“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

TURINYS

Ižanga.....	9
1. Literatūros apžvalga.....	10
1.1. Nagrinėjama problema	10
1.1.1. Akcijų kainų tyrimo problemos ir aktualumas	10
1.1.2. Investavimo strategijos ir akcijų grąža	11
1.1.3. Akcijų kainą sąlygojantys veiksniai	12
1.1.4. Akcijų tyrimų apžvalga	13
1.2. Akcijų kainų tyrime analizuojami metodai	15
1.2.1. Analizuojami klasterizavimo metodai	16
1.2.2. Analizuojami prognozavimo metodai	18
1.3. Uždavinio formuluotė, įgyvendinimas, taikymas.....	21
2. Medžiagos ir tyrimų metodai	22
2.1. Duomenų šaltiniai ir jų apdorojimas	22
2.1.1. Europos šalių biržų akcijos.....	22
2.1.2. Makroekonominiai rodikliai	23
2.1.3. Paieškos sistemos Google tendencijos	24
2.1.4. Duomenų gavybos procesas	25
2.2. Naudojama programinė priemonė	28
2.3. Prognozuojamas kintamasis	29
2.4. Klasterizavimo metodų taikymas	30
2.4.1. Atstumo mato parinkimas.....	30
2.4.2. Padalijimo pagrįstu klasterizavimo metodo jungimo algoritmai.....	30
2.4.3. Hierarchinio klasterizavimo jungimo metodai	31
2.4.4. Klasterizavimo kokybės indeksai	32
2.5. Ypatingųjų reikšmių dekompozicija.....	33
2.6. Prognozavimo metodų taikymas	33
2.6.1. Apibendrintas tiesinis modelis – regresija.....	33
2.6.2. Autoregresinis neuroninis tinklas	34
2.6.3. Prognozės įvertinimo metrikos	35
3. Tyrimų rezultatai ir jų aptarimas.....	36
3.1. Pirminė duomenų analizė	36
3.1.1. Ekstremalių reikšmių tyrimas	36

3.1.2. Vidurkio ir sklaidos apie jį tyrimas	37
3.1.3. Autokoreliacijos tyrimas.....	39
3.1.4. Stacionarumo tyrimas	39
3.1.5. Multikolinearumo tyrimas	39
3.1.6. Koreliacijos tyrimas tarp gražos ir nepriklausomų kintamųjų.....	41
3.2. Klasterizavimo metodų tyrimas.....	42
3.2.1. Padalijimu pagrįstų metodų tyrimas	42
3.2.2. Hierarchinių metodų tyrimas	43
3.2.3. Klasterizavimo metodų palyginimas	45
3.2.4. Sudarytų klasterių aprašymas	46
3.4. Akcijų gražos prognozių modelių sudarymas ir palyginimas	48
3.4.1. 1-ojo klasterio akcijų prognozės modeliai.....	49
3.4.2. Kitų klasterių akcijų prognozės modeliai	53
Išvados	62
Literatūros sąrašas	63
1 priedas. Analizuojamų biržų sąrašas.....	65
2 priedas. Makroekonominių rodiklių kitimas analizuojamose šalyse.....	66
3 priedas. Finansinių terminų paieškos apimčių google sistemoje kitimas.....	68
4 priedas. Klasterių aprašomoji statistika ir požymiai.....	70
5 priedas. Klasterių reiškiančių akcijų nustatymo rezultatai.....	72
6 priedas. 2-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	74
7 priedas. 3-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	76
8 priedas. 4-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	78
9 priedas. 5-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	80
10 priedas. 6-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	82
11 priedas. 7-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	84
12 priedas. 8-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	86
13 priedas. 9-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	88
14 priedas. 10-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	90
15 priedas. 11-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	92
16 priedas. 12-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	94
17 priedas. 13-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	96
18 priedas. 14-ojo klasterio akcijų prognozės modelių tyrimų rezultatai.....	98

Mekšunaitė, Erika. Europos šalių akcijų gražų analizė ir prognozė. Magistro baigiamasis projektas / vadovai doc. dr. Kristina Lukoševičiūtė, doc. dr. Alina Stundžienė; Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas.

Mokslo kryptis ir sritis: duomenų analitika

Reikšminiai žodžiai: *akcijos, rinka, Google tendencijos, makroekonominiai rodikliai, klasterizavimas, prognozė, regresija, neuroniniai tinklai.*

Kaunas, 2017. 99 p.

SANTRAUKA

Akcijų rinka, tai sudėtinga ir dinamiška sistema, kuriai įtaką daro begalė išorinių veiksnių. Todėl akcijų rinkos prognozė yra vienas iš sudėtingiausių laiko eilučių analizės uždavinių. Modeliai, kurie atlieka patikimą šių laiko eilučių prognozę, gali suteikti didelę naudą investuotojams, panaudojant modelį kaip atramą priimant investavimo sprendimus.

Šiame darbe sudarėme analizuojamų Europos šalių akcijų vidutinės mėnesinės gražos prognozės modelį. Tuo tikslu buvo atlikta akcijų padalijimu pagrįstu ir hierarchinio klasterizavimo metodų lyginamąją analizę, tiesinės regresijos ir netiesinio autoregresinio neuroninio tinklo prognozių analizę. Tokiu būdu prognozė buvo gauta naudojant ne vien akcijų gražos ankstesnes vertes, bet taip pat naudojant išorinę informaciją – *Google* paieškos sistemos tendencijas ir makroekonominius rodiklius.

Pirma, buvo atliktas duomenų gavybos procesas, kurį sudaro duomenų išgavimas iš trijų duomenų bazių, jų išvalymas ir apdorojimas. Modeliavimo etape buvo atlikta aprašomoji analizė, kurios metu parinkti išorinių veiksnių rodikliai. Tuomet akcijos pagal panašumą buvo suskirstytos į grupes taikant keletą skirtingų klasterizavimo metodų, atstumo matų, klasterių jungimo algoritmų, o optimalus klasterių skaičius parinktas atsižvelgiant į *Silhouette*, *Davies-Bouldin* ir *Dunn* klasterizavimo kokybės vertinimo metrikas. Galutiniame darbo etape, kiekvienai akcijų grupei atskirai sudaryti du prognostiniai modeliai pagal dominuojančią grupės akcijos gražos laiko eilutę ir atlikta visų analizuojamų Europos šalių akcijų gražos prognozė. Siekiant įvertinti sudarytų modelių kokybę buvo įvertinta vidutinė kvadratinė paklaida ir naujai pristatytos prognozės krypties ir prognozės gražos teigiamumo/ neigiamumo (pelno ar nuotolio) metrikos.

Atliktų tyrimų rezultatai leidžia daryti išvadą, kad tiek autoregresinis neuroninio tinklo modelis tiek tiesinė regresija akcijų gražai prognozuoti yra tinkami modeliai. Tačiau neuroninio tinklo modelis šiek tiek tikslesnis kokybės metrikų atžvilgiu ir gali būti naudojamas kaip finansinių laiko eilučių prognozavimo standartinių statistinių metodų alternatyva. Taip pat analizuojant kasdienes gražų duomenis galimai būtų sudarytas tikslesnis modelis.

Mekšunaitė, Erika. *European Stocks Returns Analysis And Forecast*: Master's thesis / supervisors assoc. prof. Kristina Lukoševičiūtė, assoc. prof. Alina Stundžienė. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: data analysis

Key words: *stock, exchange, Google trends, macroeconomic indicators, clustering, forecast, regression, neural network.*

Kaunas, 2017. 99 p.

SUMMARY

Considering the fact that markets are generally influenced by different external factors, the stock market prediction is one of the most difficult tasks of time series analysis. The use of models that provide a reliable prediction in financial time series may bring valuable profits for the investors, because they can use the prediction model as a valuable decision support.

The European stocks monthly average return prediction model, based on nonlinear autoregressive neural networks and linear regression comparison with partitional and hierarchical clustering comparison for stocks returns time series forecasting is presented in this work. In this way, the prediction was obtained by not just using the previous values of the series but also by using information external to the main series – search engine *Google* trends and macroeconomic indicators.

Firstly, data preprocessing was done to download data from three databases, clean up and process it. In the modeling step, technical analysis has been conducted to select indicators. Then clustering approach was applied to group the stocks by similarity. Applied different clustering methods, distance measures, linkage technics and optimal number of clusters were evaluated depending on *Silhouette*, *Davies-Bouldin* and *Dunn* indexes. Finally, the two forecasting models were used to train the data and perform analyzed European stock return time series forecast in each group separately. The mean root squared error and newly created prediction on change direction and prediction on return sign measures were used to evaluate the performances of proposed model.

The experimental results lead to the conclusion that the proposed autoregressive neural networks model was more successfully than linear regression and it can be used as an alternative method to standard statistical techniques for financial time series forecasting.

SANTRUMPOS

- CPI – vartotojų kainų indeksas;
D – *Dunn* klasterizavimo kokybės indeksas;
DB – *Davies-Bouldin* klasterizavimo kokybės indeksas;
ED – Euklido atstumas;
GLM – apibendrintas tiesinis modelis;
Hits – *Google* sistemoje užklauso paieškos dydis;
INFPP – infliacijos rodiklis;
IPG_PP – pramonės produkcijos augimas;
IRLT – ilgalaikė palūkanų norma;
IR3TIB – trumpalaikė palūkanų norma;
MSE – vidutinė kvadratinė paklaida;
NARX – netiesinis autoregresinis su papildomu įėjimu neuronų tinklas;
POCID – prognozės krypties pokyčiu paremta metrika;
POS – prognozės reikšmių ženklų sutapimo metrika;
RMSE – kvadratinė šaknis iš vidutinės kvadratinės paklaidos;
SBD – laiko eilutės forma paremtas matas;
SHARE_PRICES – akcijų kainų indeksas;
Sil – *Silhouette* klasterizavimo kokybės indeksas;
SVD – ypatingųjų reikšmių dekompozicija.
TERM_SP – palūkanų normų apimtis/ paplitimas;
UNEMP_R – nedarbo lygis;

IŽANGA

Tiriama problema. Europos šalių akcijų gražos prognozė, panaudojant išorinius veiksnius tokius kaip paieškos sistemos *Google* tendencijos ir makroekonominiai rodikliai.

Tyrimo aktualumas. Akcijų kainos prognozių modeliai tiriami jau dešimtmečius, tačiau sudaryti tikslų modelį yra labai sudėtingas uždavinys, kadangi akcijų rinka yra sudėtinga ir dinamiška sistema. Šį dinamiškumą įtakoja mažiausias ekonominis vienetas – žmogus, jo elgsena akcijų rinkoje. Tokią informaciją puikiai atspindi šių dienų orakulas – internetinės paieškos sistemos. Todėl norint įvertinti akcijų rinkos dinamiką reikia įvertinti ne vien šalių ekonomiką, bet ir su investavimu susijusių paieškų internete tendencijas. Toks sėkmingas būsimų akcijų kainų žinojimas duotų investuotojams didelį pelną.

Temos naujumas. Darbe akcijų gražai prognozuoti palyginami netiesinis autoregresinis neuroninis tinklas su papildomais įėjimais ir tiesinės regresijos metodai. Tiriama 2015 metais pristatyta akcijų kainos priklausomybė nuo paieškos sistemos duomenų. Akcijų grupavimui naudojama nauja laiko eilučių klasterizavimo metodika paremta laiko eilučių forma. Apjungus šiuos metodus ir veiksnius tikėtina sudaryti sėkmingą akcijų gražos prognozės modelį.

Tyrimo tikslas. Išanalizuoti akcijų ir išorinių veiksnių duomenis, palyginti keletą klasterizavimo algoritmų akcijoms sugrupuoti į panašias grupes, palyginti regresijos ir neuroninių tinklų metodus, taip sudarant Europos šalių akcijų gražos prognozės modelį, panaudojant išorinius veiksnius tokius kaip paieškos sistemos *Google* tendencijos ir makroekonominiai rodikliai.

Sprendžiami uždaviniai:

- 1) Duomenų paruošimas, išgavimas ir apdorojimas;
- 2) Pirminė duomenų analizė – kitimo tendencijų apžvalga, priklausomybių tyrimas;
- 3) Akcijų klasterizavimas, atliekant keleto metodų ir įvairių jungimo algoritmų, bei atstumo matų lyginamąją analizę;
- 4) Reikšmingiausių akcijų gražų laiko eilučių klasteriuose išskyrimas;
- 5) Kiekvieno klasterio akcijų gražos prognozės modelių sudarymas pagal jame dominuojančią akciją, palyginant regresinę analizę ir autoregresinius neuroninius tinklus.

1. LITERATŪROS APŽVALGA

Šiame skyriuje suformuluosime sprendžiamą uždavinį bei pasirinksimė sprendimo metodus. Visa tai atliekama atsižvelgiant į jau atliktus tyrimus nagrinėtoje mokslinėje literatūroje ir kituose mokslinės informacijos šaltiniuose.

1.1. Nagrinėjama problema

Akcija – tai vertybinis popierius pagrindžiantis indėlį į akcinį kapitalą, nuosavybės dalį įmonėje ir duodantis jo savininkui teisę gauti dividendą. Akcijos išleidžiamos ir jomis prekiaujama akcijų rinkoje. Listinguojamų įmonių vertybiniais popieriais prekiaujama akcijų biržoje, kurioje užtikrinama teisinga ir efektyvi prekyba.

1.1.1. Akcijų kainų tyrimo problemos ir aktualumas

Akcijų rinka tai sudėtinga sistema, kadangi akcijų kainos kinta kiekvieną dieną. Todėl akcijų pardavėjams ir pirkėjams labai sunku numatyti būsimą jų vertę rinkoje. Dėl šios priežasties jau dešimtmečius akcijų duomenų tyrimas yra labai populiarus investuotojų ir mokslininkų tarpe. Sėkmingas būsimų akcijų kainų prognozavimas duotų investuotojams didelį pelną. Tačiau teoriniai modeliai parodo įtakų ryšius, bet sunkiai paaiškina didelių kainų svyravimus, kadangi investuotojai dažniausiai prieš investuojant įvertina riziką. Jau nuo XX amžiaus pradžios pradėta plėtoti efektyvios rinkos teorija, kuri teigia, kad jei investuotojai naują informaciją rinkoje priima efektyviai ir akcijų kainą nustato racionaliai, taip ateities pelningumas yra nuspėjamas, t. y. vertybinių popierių rinkos kainos pilnai atspindi esamą ir greitai keičiasi pagal naują viešą ir neviešą informaciją. Iš esmės tokia situacija yra todėl, kad jei būtų nuspėjamas aukštas pelningumas ateityje, tai investuotojai pirktų akcijas dabar ir kaina pakiltų tiek, kad pelningumas ateityje dėl to sumažėtų. Tačiau yra daug mokslininkų kritikuojančių šią teoriją ir teigiančių, kad egzistuoja begalė metodų ir technologijų, kurios leidžia nustatyti būsimų kainų informaciją [1].

Tokios prognozavimo metodikos skirstomos į tris plačias kategorijas, kurios gali persidengti. Pirmoji tai fundamentali analizė – metodų visuma, kuriais siekiama nustatyti „tikrąją“ akcijų vertę ir prognozuoti jos kitimą. Šiam tikslui naudojami ne biržų duomenys, o įmonės finansiniai ir veiklos duomenys ir įvairūs makroekonominiai rodikliai. Atliekama rinkos (ekonomikos), pramonės šakos ir įmonės analizė. Antroji tai techninė analizė – metodų visuma, kuriais siekiama nuspėti akcijų kainas, remiantis jų kitimo istorija. Šios analizės pagrindas yra prielaida, kad visa reikalinga informacija apie reikšmingus veiksnius atsispindi akcijų kreivėje ir įvykiai rinkoje gali pasikartoti. Naudojami įvairūs paprasti statistiniai laiko eilučių analizės metodai. Trečioji ir šiuo metu plačiausiai taikoma kategorija – technologiniai metodai (duomenų tyryba), atsiradę kartu su skaitmeninių kompiuterių atsiradimu. Šių metodų tikslas atliekant duomenų apdorojimo ir skaičiavimų procesus rasti tam tikrą struktūrą dideliuose duomenų rinkiniuose. Tai dirbtinio

intelekto (angl. *artificial intelligence*), „besimokinančių“ sistemų (angl. *machine learning*), statistikos ir duomenų bazių sistemų sankirta [1]. Būtent šios kategorijos metodus tirsime šiame darbe.

1.1.2. Investavimo strategijos ir akcijų grąža

Bet kuriam investuotojui yra labai svarbu žinoti dvi akcijų kainas: turimos investicijos ar planuojamos įsigyti esamą kainą ir jos būsimą pardavimo kainą. Nepaisant to, investuotojai nuolat peržiūri kainų kitimo istoriją ir pagal jas priima investicinius sprendimus. Kai kurie investuotojai pirks akcijas, kurių kaina labai smarkiai nukrito, manydami kad tai įvyko dėl kainos korekcijos. Tuo tarpu kiti investuotojai vengia pirkti akcijas, kurių kaina krenta, bijodama kad kainos nuosmukis tęsis ir toliau. Patyrę investuotojai, kurie matę daug rinkos pakilimų ir nuosmukių, dažnai laikosi nuomonės, kad laikui bėgant rinka išsilygina, t. y. akcijų kainos tendencija laiku bėgant konverguoja prie vidutinės reikšmės (vidurkio reversija). Kita populiari investavimo strategija yra kai neatsižvelgiama nei į buvusias akcijos kainas, nei į kainos prognozes, o investuojama „martingale“ principu. Laikoma, kad geriausia rytojaus kainos prognozė yra šiandienos akcijos kaina plus nedidelis jos pakilimas, taip investuotojai sutelkia dėmesį tik į rizikos valdymą, būdingą jo nepastovioms investicijoms. Taip pat yra investuotojų, kurie perka atpigusias akcijas tikėdamiesi, kad akcijos yra nuvertintos ir ilgainiui jų kaina pasikoreguos, tuomet bus gaunamas pelnas [2].

Nors yra daug įvairių investavimo strategijų ir kiekvienas investuotojas ją pasirenka savo nuožiūra, tačiau visų finansinių investicijų tikslas yra pelnas. Rodiklis, parodantis iš investicijos į akciją gautą naudą vadinamas akcijų grąža. Šią grąžą laiko momentu t galime apibrėžti tokia formule:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}; \quad (1)$$

čia p_t – akcijų kaina laiko momentu t (pardavimo kaina), p_{t-1} – akcijų kaina prieš tai buvusiu laiko momentu (pirkimo kaina). [3] straipsnyje taip pat siūloma tiriant ir lyginant įvairias akcijas analizuoti jų grąžas vietoj kainos, kadangi tuomet akcijas išmatuotas ta pačia metrika galima prasmingai palyginti.

Šiame darbe sudarysime modelį akcijų grąžai prognozuoti, nevertinant rizikos dydžių ir nesiekiant sudaryti investicinį akcijų portfelį.

1.1.3. Akcijų kainą sąlygojantys veiksniai

Mokslinėje literatūroje atlikta daugybę tyrimų, kurių metu nustatytas akcijų gražų nuspėjamumas naudojant makroekonominius rodiklius. Šie rezultatai nestebina, kadangi makroekonomika daro svarbią įtaką įmonių finansinei veiklai. Tačiau rasti konkrečius rodiklius darančius įtaką konkrečių akcijų gražos kitimui yra gana sudėtingas uždavinys.

Siūloma akcijų gražos prognozavimui naudoti tokius rodiklius kaip santykinę pinigų rinkos palūkanų normą, santykinę 3 mėnesių išdo sąskaitos normą (trumpalaikių palūkanų norma), santykinę ilgalaikių vyriausybės obligacijų pajamingumą (ilgalaikių palūkanų norma), skirtumą tarp ilgalaikių ir trumpalaikių palūkanų normų, infliacijos dydį, pramonės produkcijos augimą, pinigų rinkos rodiklius, nedarbo lygio kitimą [4]. [5] straipsnyje papildomai be jau išvardytų rodiklių siūloma ištirti akcijų indeksus ir kredito maržos dydį. [6] publikacijoje be jau išvardintų analizuojami ir valiutos kurso bei naftos kainos rodikliai.

Taip pat literatūroje siūloma akcijų kainų dinamikai analizuoti ir prognozuoti naudoti įmonių kiekybinius ir kokybinius finansinius rodiklius. Kokybiniai rodikliai panaudojant teksto tyrybą gali būti išskiriami iš finansinių ataskaitų. Pavyzdžiui [7] straipsnyje naudojami dvireikšmiai kintamieji nurodantys ar ataskaitose buvo paminėti tokie terminai kaip efektyvumas, augimas, pranašumas, pagerėjimas, nepakankamumas, reorganizacija, sudėtingumas ir nusiskundimai. Tame pačiame straipsnyje išskiriami tokie kiekybiniai finansiniai rodikliai kaip veiklos pelningumas (angl. *operating margin*), nuosavo kapitalo pelningumo rodiklis (angl. *Return On Equity*, ROE), veiklos pelnas tenkantis turtui (angl. *return on total assests*, ROTA), nuosavo kapitalo santykis (angl. *Equity to capital*) ir gautinų sumų apyvartos santykis (angl. *Receivables Turnover*).

Manoma, kad rinkos pakilimai ir nuosmukiai priklauso nuo didelės grupės žmonių pasirinkimų, toks „bandos elgesys“ akcijų rinką padaro labai chaotiška ir sunkiai prognozuojama, kadangi nežinome žmonių mąstymo prieš nusprendžiant pirkti ar parduoti akcijas. Todėl norint prognozuoti rinką reikia duomenų apie žmonių elgseną prieš priimant finansinius sprendimus. Vienas tokių duomenų šaltinių, tai paieškos užklausų internete apimtys. Tokia informacija laisvai prieinama per *Google* tendencijas (angl. *trends*) – tai vienas didžiausių realaus laiko paieškos internete duomenų rinkinių visame pasaulyje. Paieškos užklausų duomenys atspindi ekonominio gyvenimo mažiausio galimo vieneto – asmens elgseną. Jau 2010 metais mokslininkai [8] straipsnyje analizavo kaip priklauso akcijų kaina ir apyvarta nuo to kiek kartų įmonių vardai buvo naudojami *Google* paieškos užklausoje. [9] darbe autoriai siūlo prognozuojant akcijų kainų kitimą naudoti terminų, susijusių su finansais ir ekonomika, paieška *Google* sistemoje.

1.1.4. Akcijų tyrimų apžvalga

Kaip anksčiau minėta yra atlikta begalės tyrimų norint nustatyti akcijų kainų ar gražų kitimų tendencijas šalių, biržų ar tik įmonių lygyje. Šiame skyrelyje apžvelgsime gana sėkmingai atliktus tyrimus mokslinėje literatūroje ir kituose mokslinės informacijos šaltiniuose. Šiuose darbuose siekiant sudaryti įvairius akcijų prognozės modelius su įvairiais nepriklausomais kintamaisiais atliekami įvairūs akcijų grupavimai, pagrindinių požymių išskyrimai.

2005 metais [4] straipsnyje ištirta 12-os išsivysčiusių šalių (Belgijos, Kanados, Danijos, Prancūzijos Vokietijos, Italijos, Japonijos, Nyderlandų, Norvegijos, Švedijos, Jungtinės Karalystės Jungtinės Amerikos Valstijos) atskirai akcijų mėnesinės gražos nuspėjamumas naudojant buvusio periodo makroekonominius kintamuosius. Grangerio priežastingumo ryšio testu gauta, kad palūkanų norma ir infliacija yra patikimiausi rodikliai prognozuojant akcijų gražą įvairiose šalyse. Taip pat šių rodiklių koeficientai gauti esantys reikšmingi regresijos modelyje, nors nustatyta, kad akcijų gražą sudaro labai nedidelė nuspėjama komponentė.

Straipsnyje [5] autoriai sudarė 10-ies išsivysčiusių šalių tikimybinis akcijų klasifikavimo modelius, į kuriuos įtraukė po vieną ir po keletą makro rodiklių. Čia gauta, kad prognozuojant akcijų kilimą ar nuosmukį, reikšmingiausi rodikliai yra infliacija (ilgojo periodo prognozei) ir ilgalaikių palūkanų norma (trumpojo periodo).

Stokholmo biržos 30-ies akcijų topologinis klasterizavimas pagal akcijų 5-erių metų kainų kitimą tarp atskirų įmonių atliktas [10] šaltinyje. Hierarchinė struktūra išgauta iš koreliacijos koeficientų tarp visų porų akcijų logaritminių kasdienių gražų matricos. Taikyti vidutinės ir vienetinės jungties klasterizavimo algoritmai, naudotas Euklido atstumas tarp koreliacijos koeficientų. Nustatyta, kad geriau tinkantis metodas klasterizavimui yra vidutinės jungties metodas, kuriuo gauti penki klasteriai ekonomine prasme interpretuojami pagal visuotinį ekonominės veiklos klasifikavimo standartą (angl. *Global Industry Classification Standard*, GICS).

Siekiant ištirti tinkamiausią klasterizavimo metodą sugrupuojant akcijas pagal priklausomumą tam tikriems sektoriams, [11] straipsnyje mokslininkai atliko nacionalinės vertybinių popierių biržos akcijų kasdienų duomenų klasterizavimą tankiu paremtu (DBSCAN algoritmas), hierarchiniu (vienetinės jungties algoritmas) ir padalijimo (k -vidurkių algoritmas) metodais. Naudojant Euklido atstumą kaip klasterių nepanašumo metriką nustatyta, kad geriausiai duomenų klasterizavimui į sektorius tinka k -vidurkių metodas, čia gautas optimalus grupių skaičius lygus 4 (kvadratinių paklaidų klasterių viduje suma lygi 127.02).

Taip pat akcijų klasterizavimas į sektorius aprašomas [12] šaltinyje, čia įrodoma, kad akcijų uždarymo kainos kinta kartu visame sektoriuje. Šiame darbe naudojami laiko eilučių hierarchinio klasterizavimo algoritmas bei hierarchinio klasifikavimo schema ir neuroniniu tinklu paremtas

TreeGNG (angl. *Tree Growing Neural Gas*) algoritmas, naudojamas Euklido atstumas. Ištyrus Jungtinės Karalystės įmonių akcijų kainų indeksą FTSE 100 sudarančių 73 įmonių akcijų kasdienių logaritminių gražų 10-ies metų laikotarpio duomenis gauti sėkmingi klasterizavimo rezultatai atitinkantys FTSE globalaus klasifikavimo schemą.

Mokslininkai [7] publikacijoje sudarė akcijų kainų kitimo prognozės modelį panaudojant įvairius įmonių finansinius kiekybinius ir kokybinius rodiklius. Hierarchinio ir k -vidurkio metodų pagalba atrinktos akcijų kainas reprezentuojantys požymių vektoriai, kurie panaudojami akcijų kainų klasifikavimui (kils, smuks ar nesikeis kaina). Lyginami klasifikavimo metodai tokie kaip hierarchinio surinkimo ir pasikartojančio k -vidurkių klasterizavimo hibridinis metodas (angl. *Hierarchical agglomerative and Recursive K means clustering*, HRK) ir atraminių vektorių metodas (angl. *Support Vector Machines*, SVM). Čia modeliai taikomi atsitiktinai parinktomis 35-erių akcijų (priklausančių įvairiems sektoriams) 13-os metų laikotarpio kainų laiko eilutėms.

Kitas akcijų klasterizavimas atliktas siekiant sudaryti akcijų kainos prognozės modelį [13]. Čia autorius siekia panaudojant duomenų tyrybos algoritmus, tokius kaip klasterizavimą ir regresiją, sudaryti analizės sistemą, kuri padeda investuotojams identifikuoti pelningesnes įmones rinkoje. Darbe analizuojami nacionalinės vertybinių popierių biržos 10-ies įmonių akcijos 6-erių mėnesių laikotarpio vidutinė akcijų kaina. Mokslininkai pritaikę ir palyginę pagal kokybės vertinimo indeksus hierarchinio, padalijimo ir tankiu pagrįstus klasterizavimo metodus nustatė, kad padalijimo metodu (k -vidurkių jungimo algoritmas) gaunami geriausi rezultatai. Šiame straipsnyje tyrėjai atlikę akcijų klasterizavimą daugialypės regresijos metodu sudarė uždarymo kainos 21-erių dienų prognozę, kurioje naudojami nepriklausomi kintamieji – atidarymo, uždarymo, aukščiausia, žemiausia ir dienos prieš uždarymą kainos.

Mėnesinių akcijų gražos ženklo (krypties) prognozės modelis sudarytas [6] šaltinyje. Tam tikslui autorius sudarė *probit* modelius Australijos, Kanados, Prancūzijos, Vokietijos, Italijos, Japonijos, Nyderlandų, Švedijos, Šveicarijos, Jungtinės Karalystės ir Jungtinių Valstijų biržų akcijoms, kuriame nepriklausomi kintamieji buvo vėlintos gražos reikšmės, vėlintos gražos krypties dvireikšmis kintamasis ir kiti makroekonominiai rodikliai.

Tyrėjai [14] straipsnyje palygino du neuroninių tinklų modelius prognozuojant dviejų pasirinktų Brazilijos įmonių kasdienių akcijų kitimą tik nuo akcijų istorinių kainų ir įtraukus išorinius veiksnius (makroekonominių rodiklių, akcijų indeksus). Čia palyginus daugiasluoksniu perceptrono (angl. *Multi-layer Perceptron*, MLP) ir radialinės bazės funkcijos (angl. *Radial Basis Function*, RBF) neuroninius tinklus, gauta, kad į modelius įtraukus išorinius veiksnius įvairių metrikų atžvilgiu tinkamesnis metodas yra MLP, o be išorinių veiksnių RBF.

2015 metais mokslininkas atlikęs principinių komponentių analizę (angl. *Principal Component Analysis*, PCA) – sudaręs iš 35 kintamųjų požymių vektorius, sudarė netiesinį

autoregresinį su papildomu įėjimu neuronų tinklą (angl. *Nonlinear Autoregressive Network with Exogenous input*, NARX) Nasdaq OMX Baltic biržos vienos pasirinktos įmonės akcijų kasdienių uždarymo kainos prognozei [15]. Gauta išvada, kad toks modelis gali sėkmingai prognozuoti akcijų kainą.

Panašus tyrimas atliktas taip pat 2015 metais, čia tyrėjai palygino autoregresinį slenkamųjų vidurkių (angl. *Autoregressive Integrated Moving Average*, ARIMA) ir NARX modelius [16]. Naudojami duomenys Bukarešto akcijų biržos pasirinktos įmonės savaitiniai duomenys ir 7 išoriniai indikatoriai. Darbo rezultate gauta, kad NARX modelis tinkamesnis akcijų uždarymo kainai prognozuoti.

2010 metais Tobias Preis pristatė *Google* tendencijų duomenų panaudojimą akcijų kainos analizei. Jis palygino du savaitinių duomenų rinkinius 6 metų laikotarpyje: kiek kartų S&P 500 indekso įmonės vardas buvo naudojamas *Google* paieškos užklausoje ir tos įmonės akcijų uždarymo kainas ir prekybos apimtis [8]. Šio darbo rezultate gauta, kad *Google* duomenys neįtakoja savaitinių akcijų kainų svyravimų, tačiau buvo nustatyta stipri koreliacija tarp įmonės pavadinimo paieškų internete ir jos akcijų prekybos apimtys.

To paties mokslininko ir jo kolegų 2013 metais publikuotame straipsnyje teigiama, kad *Google* tendencijų duomenys gali būti naudingi prognozuojant savaitinį akcijų kainų kitimą [9]. Mokslininkai lygino 98 terminų, susijusių su finansais ir ekonomika (tokie kaip „skolos“, „krizė“), paieška *Google* sistemoje su DJIA (*The Dow Jones Industrial Average*) rinkos indekso akcijų paskutine kaina. Ištyrus investavimo strategijas pagal *Google* tendencijų duomenis (jei prieš tai buvusį periodą paieškos apimtys sumažėjo tuomet akcijos parduodamos, jei išsaugo tuomet perkamos), išsiaiškinta, kad prieš akcijų kainų nuosmukį išauga finansinių rinkų raktažodžių paieškos apimtys. Taip pat nustatyta, kad terminas „skolos“ reikšmingiausias prognozuojant rinkos pakilimus ir nuosmukius.

1.2. Akcijų kainų tyrime analizuojami metodai

„Besimokinančios“ sistemos – dirbtinio intelekto sritis, plačiai naudojama technika norint iš didelių duomenų rinkinių apibūdinančių akcijų kainų dinamiką nustatyti jų tendencijas. „Besimokinančios“ sistemos technikos duomenų tyrybos metodai klasifikuojami į aprašomuosius („be mokytojo“) ir prognostinius („su mokytoju“). Aprašomieji metodai nustato ryšius egzistuojančius tarp duomenų, prognostiniai skirti reikšmių prognozių radimui. Šiame darbe taikysime laiko eilutėms ir analizuosime aprašomuosius klasterizavimo metodus ir prognostinius regresijos ir neuroninių tinklų metodus.

1.2.1. Analizuojami klasterizavimo metodai

Duomenų klasterizavimo tikslas – suskirstyti duomenis į grupes pagal jų panašumą, optimizuojant jų panašumą toje pačioje grupėje ir nepanašumą tarp skirtingų grupių. Dažniausiai klasterizavimo metodai taikomi tuomet, kai iš anksto duomenų grupės yra nežinomos. Klasterizavimo rezultatai gali būti naudingi aprašant ir modeliuojant ekonominius faktorius, kurie nulemia grupės akcijų kainą.

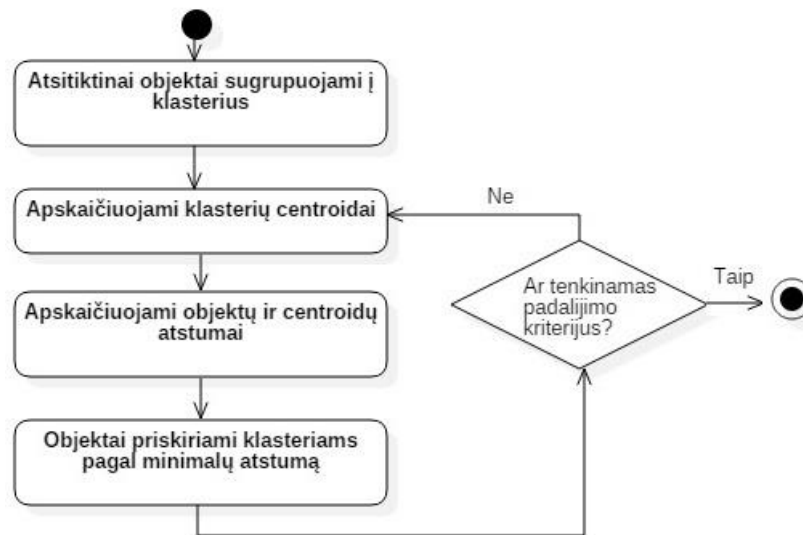
Siekiant surasti atitinkamus panašumus ir skirtumus tarp dviejų laiko eilučių yra svarbu parinkti tinkamą atstumo matą (matas nustatantis kiekybinį ryšį tarp dviejų laiko eilučių). Tam tikslui reikia atlikti keletą skirtingų matų palyginimą. Populiariausi intervalų skalėje išmatuotiems kintamiesiems klasterizuoti panašumo matai yra šie: Euklido, Euklido atstumo kvadratas, koreliacijos koeficientas, Čebyšovo ir ilgio skirtumo. Labiausiai paplitę atstumo tarp objektų grupių (skirtingumo) matai: artimiausio kaimyno, tolimiausio kaimyno, ryšių tarp grupių, *Vordo* ir kiti.

Įvairūs algoritmai ir panašumo matai laiko eilutėms yra aprašyti [17, 18, 19] šaltiniuose. Šiame darbe išstirsime ankstesniame skyrelyje aprašytoje literatūroje naudojamus matus akcijų gražoms grupuoti ir pamėginsime keletą naujų, sėkmingai taikytų laiko eilučių klasterizavimui.

Populiariausi laiko eilučių klasterizavimui taikomi metodai yra hierarchinis (angl. *hierarchical*), perdengiantis, tikimybiniai ir dalijimo (angl. *partitional clustering*) [19]. Šiame darbe taikysime hierarchinius ir padalijimo metodus, kurių aprašymai pateikti žemiau.

Padalijimu pagrįstas metodas. Duomenų rinkinį iš n objektų dalijame į k klasterių aibę, taip kad kvadratinių atstumų suma klasteryje būti minimali (optimizuojamas pasirinktas padalijimo kriterijus). Pagrindinės klasterizavimo technikos (jungimo metodai) yra tokios kaip, k -vidurkių, „ k -medoidų“, dalijimas aplink „medoidus“ (PAM). [19] straipsnio autoriaus 2015 metais pristatytas kaip aukšto laiko eilučių klasterizavimo tikslumo k -formos (angl. *k-shape*) metodas.

Algoritmas pradedamas atsitiktinai priskiriant skirtingiems objektams atsitiktinai sugeneruotus centrus. Laiko eilučių klasterizavime centroidai taip pat yra laiko eilutės. Kiekvienos iteracijos metu perskaičiuojamas taškų priskyrimas klasteriams pagal centrus ir atnaujinami klasterių centrai. Procesas kartojamas kol nebekinta padalijimo kriterijus arba nei vienas taškas nepakeičia klasterio (žr. 1.1. pav.).



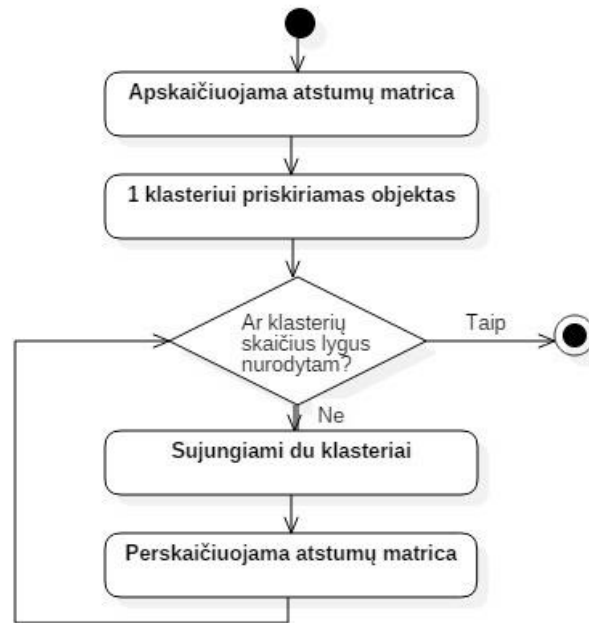
1.1. pav. Klasterizavimo padalijimo metodu algoritmo schema

Pagrindinis šio metodo trūkumas – iš anksto reikia žinoti klasterių skaičių.

Hierarchinis metodas. Hierarchiniu klasterizavimo metodu duomenys yra ne dalijami į atskirus klasterius, o gaunami sudarant klasterių medį. Hierarchiniai algoritmai skirstomi į jungimo ir skaidymo. Jungimo metodais skaičiuojant atstumus tarp klasterių prie klasterio iš vieno elemento prijungiami kiti elementai kol sudaromas vienas stambus klasteris, o skaidymo metodai veikia priešingai. Žinomiausi hierarchinio jungimo metodai:

- Vienetinės jungties arba artimiausio kaimyno (angl. *single linkage*);
- Pilnosios jungties arba tolimiausio kaimyno (angl. *complete linkage*);
- Vidutinės jungties (angl. *average linkage*);
- Vordo jungties (angl. *Wards's linkage*);
- Svertinis vidutinės jungties (angl. *weighted average linkage*);
- Centroidų jungties (angl. *centroid linkage*);
- Medianų jungties (angl. *median linkage*).

Klasterizuojant jungimo metodu atliekami tokie žingsniai: apskaičiuojama atstumo tarp visų objektų matrica, pagal šį atstumą panašiausi objektai (tarp jų atstumas mažiausias) jungiami į klasterius, taip šis procesas kartojamas kol suformuojamas hierarchinis medis, nustatomas lygis kuriame medis „nukertamas“ taip priskiriant objektus galutiniams klasteriams (žr. 1.2. pav.).



1.2. pav. Hierarchinio klasterizavimo algoritmo schema

Klasterinės analizės prielaidos. Nors taikant klasterizavimo metodą duomenų normalizuoti nebūtina, tačiau ši procedūra dažnai rekomenduojama, kad visi kintamieji turėtų vienodą įtaką atstumo skaičiavime.

Optimalaus klasterių skaičiaus parinkimas ir klasterizavimo kokybės įvertinimas. Šis uždavinys nėra sprendžiamas analitiškai, kadangi „teisingas“ klasterių skaičius priklauso nuo interpretacijos. Literatūroje [20] išskiriami trys indeksų tipai: išoriniai – klasterio atskiriamumo; vidiniai – klasterio kompaktiškumo; santykiniai – apima išorinių ir vidinių indeksų kategorijas. [13] straipsnyje autorius gautų akcijų klasterių įvairiais metodais tikslumą vertina pagal tokius indeksus: *C*, *Jaccard*, *Random* ir *Silhouette*. [18, 20] šaltiniuose klasterių skaičiaus parinkimui siūloma naudoti *Silhouette*, *Dunn*, *Davies- Bouldin* koeficientą. [20] straipsnyje autorius išskiria tokius patikros indeksus, kaip standartinio nuokrypio vidutinė kvadratinė šaknis (angl. *Root Mean Square Standart Devation*, *RMSSTD*), pusiau dalinis R kvadratinis (angl. *Semi Partial R Squared*, *SPR*), R kvadratinis (angl. *R Squared*, *RS*). Taip pat šis autorius siūlo vertinant hierarchinį klasterizavimą vertinti išorinius indeksus: atstumą tarp dviejų klasterių (angl. *Distance Between Two Clusters*, *CD*), neraiškiajam klasterizavimui siūlo tirti santykinius indeksus: sudalinimo koeficientą (angl. *partition coefficient*, *PC*) ir entropinio sudalijimo koeficientą (angl. *partition entropy coefficient*, *PE*).

1.2.2. Analizuojami prognozavimo metodai

Mokslinėje literatūroje laiko eilutėms prognozuoti taikomi įvairūs metodai pagal turimų duomenų tipą ir norimą prognozės modelį [21]. 1.1.4. skyrelyje nagrinėtoje literatūroje siekiant sudaryti akcijų kainų prognostinį modelį iš akcijų kainų praeities duomenų ir įvairių išorinių veiksnių plačiausiai ir gana sėkmingai taikomi regresijos ir neuroninių tinklų modeliai. Todėl

šiam darbe tirsime būtent regresijos ir neuroninių tinklų metodus. Šių metodų bendrasis aprašymas pateikiamas žemiau.

Apibendrintas tiesinis modelis (angl. *Generalized Linear Model, GLM*). Tai plati regresinių modelių klasė, kur modeliai charakterizuojami trimis komponentėmis (čia \mathbf{y} – nepriklausomas atsako kintamasis, \mathbf{X} – priklausomų kintamųjų rinkinys):

- atsitiktine dedamąja, kuri apibrėžiama atsako \mathbf{y} skirstiniu. Jei atsitiktinis dydis \mathbf{y} tolydus – tai ši dedamoji \mathbf{y} tikėtinumo funkcija, jei diskretus – tikimybė.
- sisteminė dedamoji, kuri nusakoma neatsitiktiniu dydžiu $\eta_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i}$
- ryšio funkcija g , tokia kad $E(\mathbf{y}) = \mu = g^{-1}(\eta)$, čia μ nepriklausomo kintamojo \mathbf{y} vidurkis. Ši funkcija nusakoma ryšį tarp sisteminės dedamosios ir atsitiktinės dedamosios vidurkio.

Žemiau pateikta lentelė dažniausiai naudojamų skirstinių ir ryšio funkcijų skirtingiems duomenims [22]:

1.1 lentelė. Dažniausiai naudojamos apibendrinto tiesinio modelio komponentės

y ryšys su x_i	y skirstinys	y skalė	Ryšio funkcija
Tiesinis	Normalusis	Realioji $(-\infty ; +\infty)$	$\mathbf{X}\boldsymbol{\beta} = \mu$
Eksponentinis	Eksponentinis	Realioji $(0 ; +\infty)$	$\mathbf{X}\boldsymbol{\beta} = \mu^{-1}$
	Gama		
Įvykių skaičius fiksuotame laiko tarpe	Puasono	Sveikoji $(0,1,2,..)$	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$
„Taip“ / „Ne“ skaičius	Bernulio	Dvireikšmė $\{0,1\}$	$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$
„Taip“ įvykių skaičius ir visų įvykių	Binominis	Sveikoji $(0,1,2,..)$	
Priklausymas kategorijai	Kategorinis	Kategorijų	

Modelio parametrai įvertinami didžiausio tikėtinumo metodu naudojant iteracinį svartinį mažiausių kvadratų algoritimą.

Šiame darbe intervalų skalėje išmatuotam priklausomam kintamajam taikysime vieną paprasčiausių GLM atvejų (kai skirstinys normalusis) – daugialypę tiesinę regresiją [22].

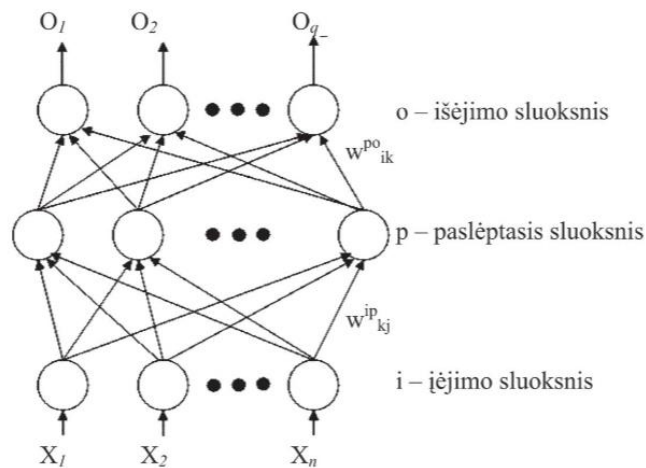
GLM modelio prielaidos. Taikant šį modelį turi būti tenkinamos tam tikros sąlygos. Vieną jų, priklausomo kintamojo \mathbf{y} reikšmės turi būti nepriklausomos, \mathbf{y} nebūtinai turi būti pasiskirstęs pagal normalųjį skirstinį, tačiau paprastai daroma prielaida, kad jo skirstinys yra iš eksponentinių šeimos skirstinių. Nebūtinai tiesinis ryšys tarp priklausomo ir nepriklausomų kintamųjų, tačiau daroma prielaida, kad egzistuoja tiesinė priklausomybė tarp ryšio funkcija transformuoto priklausomo kintamojo ir sisteminės dedamosios. Taip pat neturi būti netenkinama homogeniškumo (dispersijų pastovumo) sąlyga. Gauta modelio paklaidos turi būti

nepriklausomos, tačiau neprivalo būti normaliai pasiskirsčiusios (nebent y parenkamas skirstinys normalusis, tai paklaidos turi būti normaliosios) [22].

Dirbtis neuroninis tinklas. Tai nuosekliam informacijos apdorojimui skirta struktūra, kuri buvo sukurta remiantis biologinės nervų sistemos pagrindu. Šis tinklas pasižymi puikiomis galimybėmis sudarant sudėtingus modelius ir susideda iš daugelio tarpusavyje susijusių neuronų – komponentų, kurie atlieka skaičiavimus, o tinklo apmokymo metu keičiami tarp jų esančių jungčių svoriai. Dauguma neuroninių tinklų neuronus suskirsto į poabius (sluoksnius). Yra išskiriamos tokios neuroninių tinklų architektūros: tiesioginio perdavimo (angl. *feed-forward*), atgalinio perdavimo (angl. *feed-back*), sluoksnių (angl. *layers*) ir perceptroniniai (angl. *perceptron*) tinklai. Šiame darbe taikysime tik tiesioginio perdavimo tinklus, todėl kitų architektūrų neaprašinsime. Tiesioginio perdavimo tinklas dar vadinamas informaciją skleidžiančiu pirmyn neuroniniu tinklu, jo struktūra vieno paslėptąjo sluoksnio atveju ir ją atitinkanti formulė pateiktos žemiau 1.3. schemoje ir (2) formulėje [20]:

$$O_i = \left\{ \sum_{k=1}^h w_{ik}^{po} f \left\{ \sum_{j=1}^n w_{kj}^{ip} x_j \right\} \right\}; \quad (2)$$

čia w_{kj}^{ip} svoris tarp j -tojo neurono įėjimo sluoksnyje ir k -tojo neurono paslėptajame sluoksnyje, w_{ik}^{po} – svoris tarp k -tojo neurono paslėptajame sluoksnyje ir i -tojo neurono išėjimo sluoksnyje, $f(\cdot)$ – neurono perdavimo (aktyvavimo) funkcija.



1.3. pav. Neuroninio tinklo struktūra

Neuroninio tinklo mokymas atliekamas minimizuojant pasirinkto kriterijaus funkciją. Prieš formuojant neuroninį tinklą reikia nustatyti paslėptų sluoksnių skaičių, šiuose sluoksniuose esančių neuronų skaičių, neurono perdavimo funkciją. [20] šaltinyje pateikiamos tokios rekomendacijos šių parametų parinkimui: paslėptųjų sluoksnių skaičius daugiausiai lygus dviem, kadangi teoriškai neuroninis tinklas su vienu sluoksniu ir turinti pakankamai neuronų gali

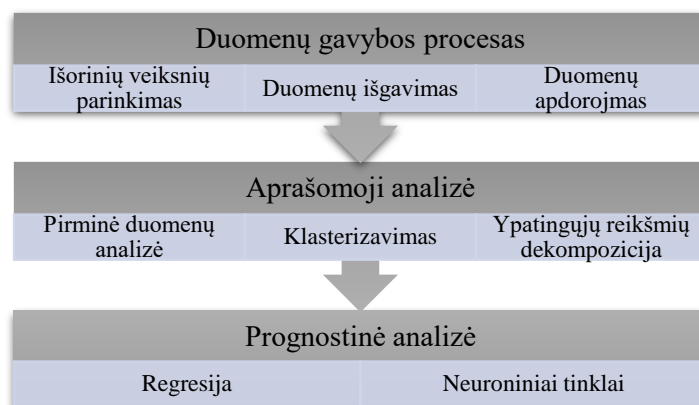
aprosimuoti bet kokią funkciją, neuronų skaičiau parinkimui reikia atlikti eksperimentinį vertinimą.

Pagrindinės neuronų tinklo savybės: adaptyvus mokymasis, savi-organizacija, lygiagretūs skaičiavimai, klaidų tolerancija.

Prognozavimo tikslumo įvertinimas. Vieni autoriai prognozavimo tikslumą vertino remiantis vidutine kvadratine paklaida (angl. *Mean Squared Error*, MSE), vidutine procentine absoliutine paklaida (angl. *Mean Absolute Percent Error*, MAPE), normalizuota vidutine kvadratine paklaida (angl. *Normalized Mean Squared Error*, THEIL), krypties pokyčio prognozės metrika (angl. *Prediction On Change In Direction*, POCID), vidutinė santykinė dispersija (angl. *Average Relative Variance*, ARV), praradimų ir išlošimų suma (angl. *Sum of Losses and Gains*, SLG) [14]. Kituose darbuose modeliai lyginti ir jų kokybę vertinta naudojant tik MSE metriką [15, 16]. Dvireikšmio kintamojo prognozės kokybei nustatyti [6] autorius vertino *pseudo* determinacijos koeficientą, ROC (angl. *Receiver Operating Characteristic*) kreivės kitimą ir AUC dydį (angl. *The Area Under the ROC Curve*), AIC kriterijų, kvadratinį tikėtimumo įvertį (QPS), klasifikavimo tikslumą (angl. *accuracy*), *Pesaran* ir *Timmermann* testą.

1.3. Uždavinio formuluotė, įgyvendinimas, taikymas

Šio darbo tikslas sudaryti Europos šalių akcijų gražų prognozių modelius panaudojant duomenų tyrybos metodus. Kadangi akcijos kainos kinta chaotiškai ir yra sunkiai nuspėjamos vien tik iš jų praeities kitimų, šiame darbe naudosime išorinius veiksnius tokius kaip *Google* tendencijos ir makroekonominiai rodikliai. Darbo tikslui įgyvendinti išgausime ir apdorosime šių rodiklių ir Europos šalių rinkų duomenis, toliau atliksime jų pirminę duomenų analizę (kitimo tendencijų tyrimas, aprašomoji statistika, priklausomybių tyrimas), sugrupuosime analizuojamas akcijas į klasterius naudojant padalijimu pagrįstą ir hierarchinius metodus – iš šių gautų grupių dekompozicijos metodu atrinksime dominuojančias akcijų gražos laiko eilutes ir joms sudarysime regresinį bei neuroninių tinklų modelius, kuriuos pritaikysime visų grupės analizuojamų akcijų gražai prognozuoti. Ši darbo eigos schema pateikta žemiau 1.4 pav. schemeje.



1.4. pav. Darbo eigos schema

2. MEDŽIAGOS IR TYRIMŲ METODAI

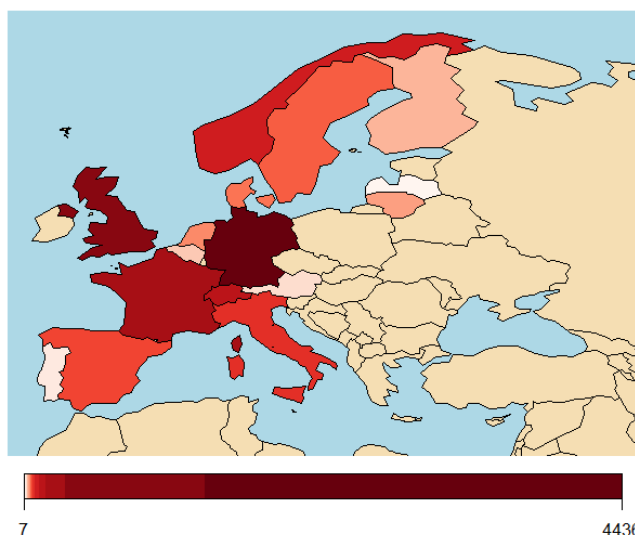
2.1. Duomenų šaltiniai ir jų apdorojimas

Šiame darbe analizuosime duomenis apjungtus iš trijų laisvai interneto vartotojams prieinamų duomenų šaltinių. Todėl viena iš svarbiausių tyrimo dalių, tai duomenų gavybos procesas. Šis procesas apima korektiškų duomenų parsisiuntimą, jų supratimą ir tikslų apdorojimą (išgryninimą). Pažymime, kad tiriamų duomenų periodas lygus 63 mėnesiams (nuo 2012 01 01 dienos iki 2017 03 31 dienos).

2.1.1. Europos šalių biržų akcijos

Šiais technologijų laikais internete gausu svetainių, kuriuose laisvai vartotojams prieinami vieši finansiniai duomenys. Viena populiariausių – tai interneto paslaugų bendrovės *Yahoo* finansinių duomenų svetainė *Yahoo Finance* (žiniatinklio prieiga per <https://finance.yahoo.com>). Dėl gana patogaus duomenų gavimo būdo – tiesiogiai iš duomenų bazės per interneto paslaugų žiniatinklio aplikaciją, šiame darbe naudojame *Yahoo* svetainėje pateiktus akcijų duomenis. Verta paminėti ir kitą populiarią akcijų rinkų duomenų svetainę – *Google Finance*. Tačiau lyginant su *Yahoo Finance* duomenų parsisiuntimo prasme galime išskirti, kad iš *Yahoo Finance* duomenų bazės parsisiuntimui prieinama daugiau akcijų duomenų nei *Google* duomenų bazėje (čia mažų rinkų duomenys prieinami tik peržiūrai). Joje prieinamos 46-ias šalis apimančios 81-os biržų rinkos duomenys. Šiame darbe analizuosime 27-ių Europos šalių akcijų biržas, kurios išvardintos 1 priedo lentelėje. Vienas didžiausių *Yahoo Finance* trūkumų lyginat su *Google Finance*, tai sudėtingas visų prieinamų Europos įmonių akcijų sąrašo išgavimas. Todėl šiame etape buvo sudaromas įmonių sąrašas iš *Google Finance* įmonių simbolių ir *Yahoo Finance* biržų simbolių.

Žemiau 2.1 paveiksle pavaizduota pradiniam akcijų duomenų faile įmonių skaičiaus pasiskirstymas analizuojamose Europos šalyse spalvų skalėje. Kaip matome daugiausia skirtingų akcijų yra Vokietijos ir Londono biržose.



2.1. pav. Analizuojamų įmonių skaičiaus pasiskirstymas šalyse spalvų skalėje

2.1.2. Makroekonominiai rodikliai

Kaip anksčiau minėta [4, 5] straipsniuose nustatyti sėkmingi ryšiai tarp akcijų gražos kitimo ir analizuotų makroekonominių rodiklių. Todėl mūsų tyrime sudarant akcijų gražų prognozės modelį pasirenkame šiuose straipsniuose analizuotus rodiklius. Taip pat pasirinkdami atsižvelgiame į tai, kad visose 16 analizuojamų šalių šie rodikliai būtų matuojami bent jau mėnesiniu periodu ir tokioje pat skalėje. Taigi, pasirenkame tokius rodiklius (visi rodikliai matuojami pagal bazinius 2010 metus, jų kitimas analizuojamose šalyse kas mėnesį pateiktas 2 priede):

- Infliacija (angl. *inflation*, INFPP) – rodiklis, apibrėžtas kaip tam tikrų namų ūkių grupių įprastai perkamų prekių ir paslaugų krepšelio kainų pokytis (dėl šio pokyčio kinta piniginio vieneto perkamoji galia). Matuojama atsižvelgiant į metinį kainų augimo procentinį dydį.
- Vartotojų kainų indeksas (angl. *consumer price index*, CPI) – rodiklis, rodantis vartojimo prekių ir paslaugų, kurias įsigyja, už kurias sumoka ir kurias namų ūkiai panaudoja tiesiogiai patenkinti vartojimo poreikius, vidutinį kainų lygio pokytį per tam tikrą laikotarpį. Šis indeksas naudojamas nustatyti infliacijos lygį. Darbe naudojamas indeksas apimantis visų prekių ir paslaugų kategorijas.
- Ilgalaikė palūkanų norma (angl. *long-term interest rates*, IRLT) – ilgalaikių vyriausybės obligacijų pajamingumas 10 metų periode (procentinis kasdienių normų vidurkis). Šią normą iš esmės nulemia mokesčių dydis skolintojui, skolininko rizika, kapitalo vertės nuosmukis. Ilgalaikė palūkanų norma yra žinoma kaip vienas iš verslo investicijas nulemiančių veiksnių. Žemas palūkanų normos dydis skatina investicijas, aukštas veikia priešingai.
- Trumpalaikė palūkanų norma (angl. *short-term interest rates*, IR3TIB) – norma, pagal kurią trumpalaikės skolos yra vykdomos tarp finansinių institucijų arba norma, pagal kurią trumpalaikiai valstybės vertybiniai popieriai išleidžiami arba prekiaujami rinkoje (procentinis kasdienių normų vidurkis). Trumpalaikių palūkanų normų pagrindas yra trijų mėnesių pinigų rinkos palūkanų normos, kurios vadinamos išdo vekselių norma (angl. *treasury bill rate*).
- Palūkanų normų apimtis/ paplitimas (angl. *term spread*, TERM_SP) – tai skirtumas tarp ilgalaikio vyriausybės obligacijų pajamingumo (ilgalaikių palūkanų normų) ir 3-ų mėnesių išdo vekselių normos (trumpalaikės palūkanų normos).
- Pramonės produkcijos augimas (angl. *industrial production growth*, IPG_PP) - nusako šalies pramoninių įmonių gamybos masto pokyčius. Apima tokius sektorius kaip kasyklų, gamyklų ir komunalinių paslaugų teikėjus (elektra, dujos, vanduo). Šis rodiklis yra matuojamas remiantis ankstesnio laikotarpio gamybos apimtimis.

- Nedarbo lygis (angl. *unemployment rate*, UNEMP_R) – tai bedarbių žmonių procentinis santykinis darbo jėgos dydis. Bedarbiai – tai galintys dirbti žmonės užsiregistravę kaip darbo neturintys asmenys ir ėmęsi aktyvių veiksmų susirasti darbą per pastarąsias keturias savaites. Į šį rodiklį įtraukti visų grupių žmonės.
- Akcijų kainų indeksas (angl. *share prices index*, SHARE_PRICES) – įmonių listinguojamų nacionalinėje arba užsienio vertybinių popierių biržose akcijų kainos. Šiuos indeksus nustato vertybinių popierių birža, išreiškiant iš kasdienių uždarymo kainų mėnesinio aritmetinio vidurkio. Šis dydis matuoja indekso akcijų vertės pokyčius, indekso akcijų krepšelių rinkos kapitalizacijos pokyčius.

Šiuos finansinius rodiklius parsisiunčiame iš viešai prieinamos OECD duomenų bazės (žiniatinklio prieiga per <http://stats.oecd.org>). Tai ekonominio bendradarbiavimo ir plėtros organizacijos (angl. *Organisation of Economic Cooperation and Development*, OECD) duomenų bazė, kurioje laisvai prieinama šios organizacijos pateikta statistinė informacija.

2.1.3. Paieškos sistemos *Google* tendencijos

Google tai tarptautinė bendrovė teikianti interneto paslaugas tokių produktų kaip paieškos sistema, debesų kompiuterija, internetinio reklamavimo technologijos ir programinė įranga. Populiarioji *Google* paieškos sistema dažnai vadinama šiuolaikiniu orakulu, į kurią kreipiamasi ieškant tam tikros informacijos. Tokiu būdu sukaupiama daugiau informacijos apie vartotojus nei, kad apie juos supantį pasaulį.

Google tendencijos – vienas didžiausių realaus laiko duomenų rinkinių visame pasaulyje (trilijonai užklausų generuojama kasmet), sukurtas 2004 metais, o nuo 2015 metų prienami ir realaus laiko duomenys (prieiga per <https://trends.google.com/trends/?hl=lt>). Šių paieškos internete duomenų tyrimas suteikia unikalią perspektyvą sužinoti kuo žmonės domisi šiuo bei ankstesniais laiko momentais. Tendencijų duomenys – tai visų paieškų užklausų *Google* sistemoje anoniminė (nesaugoma asmens identifikacija), kategorinė (suskirstyta pagal paieškos užklausos temas) ir suagreguota informacija. Šie duomenys matuojami kaip normalizuotas santykinis pasirinktos temos visų paieškų *Google* sistemoje pasirinktame periode ir regione indeksas [23]. Formulė apibrėžianti šį dydį laiko momentu t ir regione r :

$$pop_rat_{t,r} = \frac{n_{t,r}}{\sum_i N_{t,r,i}}; \quad (2)$$

$$hit_{t,r} = \frac{pop_rat_{t,r}}{\max_{t_0 \leq j \leq t_n} (pop_rat_{j,r})} \cdot 100; \quad (3)$$

Čia $pop_rat_{t,r}$ – termino populiarumo koeficientas laiko momentu t regione r ; $hit_{t,r}$ termino populiarumo normalizuotas indeksas; n_t – tam tikro termino paieškų skaičius regione r laiko

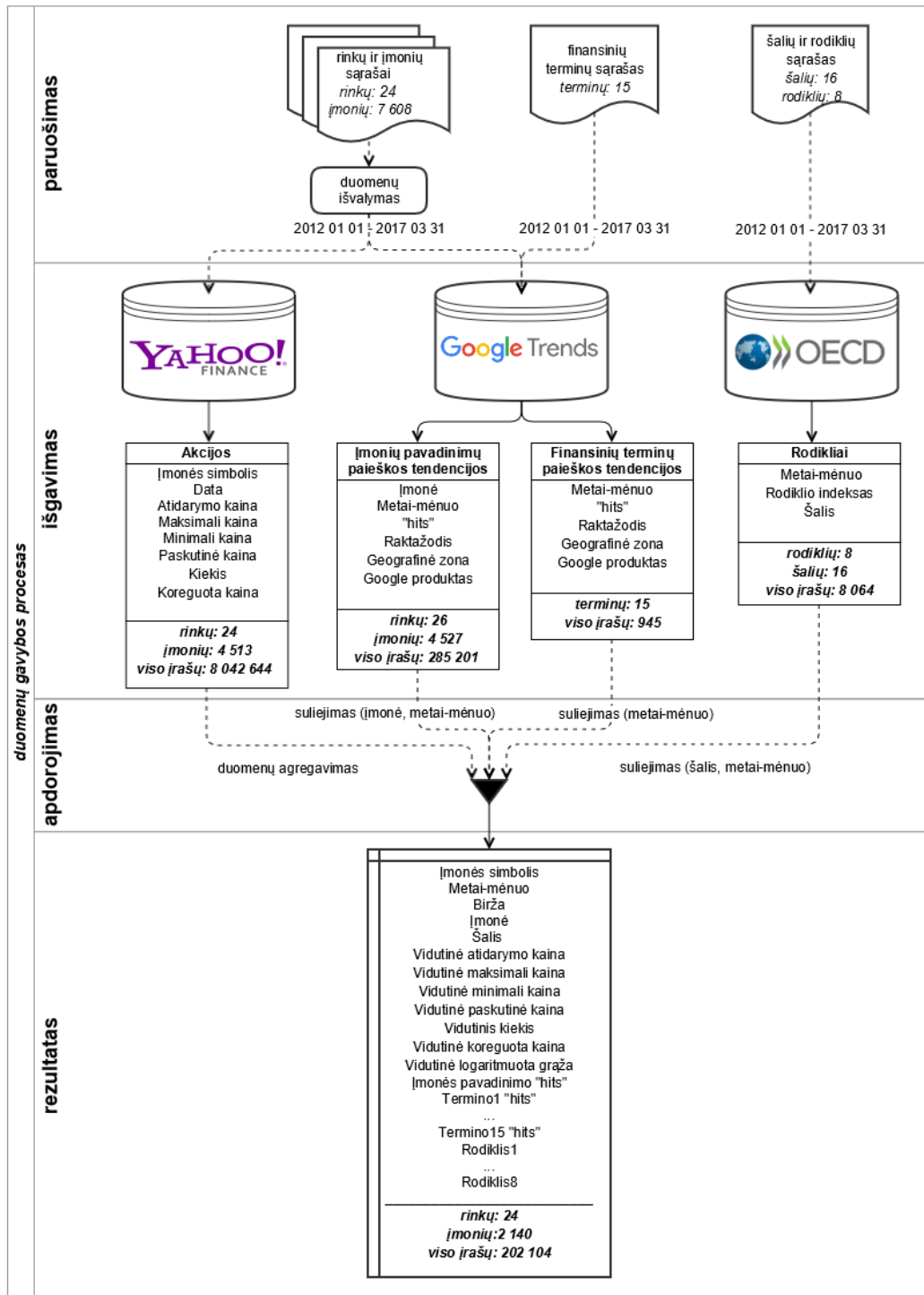
momentu t ; $N_{t,r,i}$ – i -tojo termino paieškų skaičius regione r laiko momentu t ; t_0 ir t_n – pradinis ir galutinis pasirinkto periodo laiko momentai.

Susidomėjimą tema nusakančios reikšmės kinta nuo 1 iki 100, kur 100 nusako maksimalų visų paieškų dydį pasirinktame laike ir vietoje. Tokiu būdu vartotojai gali išmatuoti susidomėjimą tam tikra tema visame pasaulyje arba šalių lygyje.

Šiame darbe sudarant modelį naudojame anksčiau aptartame [9] straipsnyje (kai terminų paieškos vartotojų apimtys yra pasauliniu mastu) gautų reikšmingiausių terminų prognozuojant rinkos akcijų kainas paieškos *Google* sistemoje dydžius. Naudojami šie terminai anglų kalba: ekonomika (angl. *economics*), skola (angl. *debt*), infliacija (angl. *inflation*), metalai (angl. *metals*), parduoti (angl. *sell*), obligacijos (angl. *bonds*), rizika (angl. *risk*), svertas (angl. *leverage*), galimybė (angl. *chance*), *Nasdaq*, pinigai (angl. *money*), akcijos (angl. *stocks*), konfliktas (angl. *conflict*), pelnas (angl. *profit*), pajamos (angl. *earnings*). Taip pat tirsime ir [8] šaltinyje analizuotą kintamąjį – įmonės pavadinimo paieškos tendencijas. Šių duomenų pilnam išgavimui svarbus žingsnis įmonės pavadinimo išvalymas *Google* tendencijos užklausos pateikimo etape. Tai aprašyta detaliau sekančiame skyrelyje. Šiame darbe bus naudojama šių kintamųjų santrumpas, tokia struktūra $\langle terminas \rangle_hits$.

2.1.4. Duomenų gavybos procesas

Iš ankstesniuose skyreliuose minėtų duomenų šaltinių parsisiunčiame duomenis ir apjungiamo į vieną bendrą duomenų rinkinį. Sėkmingam tolimesniam tyrimui svarbu korektiškai apdoroti duomenis. Tam tikslui duomenų gavybos procesą išskaidome į 3 žingsnius: paruošimas, išgavimas, apdorojimas. Šis procesas pavaizduotas žemiau pateiktoje 2.2 schemeje.



2.2. pav. Duomenų gavybos proceso schema

Pirmuoju žingsniu paruošiamie reikiamus failus duomenų filtravimui iš duomenų bazių. Surenkamos *Yahoo Finance* prieinamų 24-ių rinkų ir jose prekiaujamų 7 608-ių Europos įmonių akcijų simbolių ir įmonių pavadinimų sąrašas reikiamu formatu, kuris perduodamas kaip įmonių filtras siunčiant akcijų duomenis. Sudarome 15-os finansinių terminų sąrašą, kuris bus naudojamas terminų paieškos tendencijoms parsisiųsti. Įmonių pavadinimų paieškos tendencijoms parsisiųsti naudojame 7 608-ių įmonių pavadinimų sąrašą. Tam, kad iš *Google* tendencijų duomenų bazės būtų grąžinama korektiška informacija įmonių pavadinimai turi būti „išvalyti“ – pašalinti

pertekliniai įmonės pavadinimo žodžiai/ simboliai, kurių vartotojas vesdamas užklausą į paieškos sistemą dažniausiai nenaudoja (_GI, plc, Imoniu grupe, Corp., &, Group, NCCP, IND, Inc., Co., Plc, p.l.c., PLC, Ltd, Ord, Vz), akcinių bendrovių simboliai naudojami analizuojamose šalyse (AB, S.A., SE, A/S, As, SA, Sa, AG, Oyj, S.A.E., N, ASA), papildoma akcijų biržos informacija (N.V., NV, (Inactive), S.p.A., S.P.A., SpA, Spa, -A-, /NV, SCA, -H-, -B-, -N-, -NPV-, NPV). Ekonominių rodiklių parsisiuntimui iš OECD duomenų bazės sudarome 8-ių ekonominių rodiklių simbolių pagal OECD formatą ir 16-os šalių sąrašą. Sudarome filtrus kiekvieno rodiklio atsisiuntimo užklausiai (rodiklio dažnumas, šalys, tematika, matavimo vienetas). Taip pat visų užklausų atveju perduodamas laikotarpio filtras lygus 2012 01 01 – 2017 03 31.

Antrajame duomenų išgavimo etape per žiniatinklio tarnybas pateikiant užklausas parsisiunčiami duomenys tiesiogiai iš duomenų bazių. Gauti keturi duomenų rinkiniai. Pirmasis – akcijų duomenys, jį sudaro įmonės simbolis, data, atidarymo, uždarymo, minimali, maksimali, koreguota kainos ir apyvartos kiekis. Iš pateiktų 7 608 akcijų užklausų duomenų bazėje rasti tik 4 513 akcijų duomenys (viso 8 042 644 kasdieniai įrašai). Antrasis rinkinys, tai įmonių pavadinimų paieškos tendencijų duomenys, kurių struktūra – metų ir mėnesio simbolis, „hits“ (paieškos santykinis dydis), raktažodis (paieškos terminas – įmonės pavadinimas), geografinė zona (mūsų atveju visur lygi reikšmei „world“ – apima viso pasaulio paieškas) ir produktas (mūsų atveju visur lygi reikšmei „web“ – žiniatinklis). Taip pat suliejimui su kitais duomenų failais prie šio rinkinio pridedame įmonės simbolį. Iš pateiktų 7 608 įmonių pavadinimų užklausų duomenų bazėje rasti tik 4 527 įmonių duomenys (viso 285 201 mėnesiniai įrašai). Tokiu pat principu ir struktūra parsisiunčiamas trečias duomenų failas – finansinių terminų paieškos tendencijų duomenys (viso 16 065 mėnesiniai įrašai). Ketvirtasis duomenų rinkinys, tai makroekonominių rodiklių kitimas analizuojamuose 16 šalių. Duomenis sudaro metų ir mėnesio simbolis, rodiklio simbolis, rodiklio indeksas ir šalis (viso 8 064 įrašų). Šie duomenų rinkiniai tolimesniam darbui išsaugomi lokaliai kompiuterio atmintyje.

Trečiajame žingsnyje apdorojame duomenis, juos apjungiamo į vieną duomenų rinkinį ir išvalome. Tam, kad galėtume sulieti duomenis suagreguojame kasdienes akcijų duomenis į mėnesinius – suskaičiuojame kainų ir apyvartos vidurkius, sukuriame naują metų ir mėnesio kintamąjį. Toliau prie mėnesinių akcijų duomenų prijungiame kitus duomenis pagal raktinius kintamuosius: įmonė, metų ir mėnesio simbolis, šalis. Pagal žemiau aprašytą (4) formulę sukuriame tolimesnei analizei reikalingą kintamąjį – akcijų gražos logaritminę reikšmę ir suskaičiuojame mėnesinį jos vidurkį. Apjungus duomenų rinkinius į vieną bendrą, dėl ne visų įmonių duomenų prieinamumo *Google* ir *Yahoo* duomenų bazėse, gauname daug trūkstamų reikšmių. Kad išlaikyti duomenų korektiškumą pašaliname visus įrašus tų įmonių, kurių grupėse

yra trūkstamų reikšmių. Tokiu būdu galutiniame duomenų faile liko 24 rinkų, 2 140 įmonių, 202 104 mėnesiniai 36 kintamųjų įrašai.

2.2. Naudojama programinė priemonė

Kadangi šiame darbe siekiama išanalizuoti ir sudaryti modelį konkrečiam uždaviniui, pasinaudojant jau sukurtais algoritmais ir nesiekiame sudaryti vartotojo sąsajos, todėl pasirenkame nemokamą ir atviro kodo integruotą programavimo aplinką *RStudio* v. 1.0.143 skirtą programavimo kalbai *R*. Taip pat ši programinė aplinka pasirenkama dėl orientacijos į interaktyvią duomenų analizę, statistinės analizės procedūras, kokybišką grafinį duomenų atvaizdavimą, galimybės naudoti specializuotus paketus. *R* paketai – tai tam tikra funkcijų, procedūrų, duomenų ir dokumentacijos biblioteka, specializuota ir pritaikyta tam tikros srities uždavinių sprendimui. Šiame darbe naudojami šie pagrindiniai paketai:

- *doParallel* – lygiagrečių skaičiavimų adapteris;
- *quantmod* – kiekybinė finansinė modeliavimo struktūra;
- *gtrendsR* – *Google* tendencijų užklausų vykdymas ir informacijos gavimas;
- *OECD* – duomenų paieška ir išgavimas iš OECD (ekonominio bendradarbiavimo ir plėtros organizacija) duomenų bazės;
- *ISOweek* – datų konvertavimas ISO 8601 formatu;
- *ggplot2* – aiški ir kokybiška duomenų vizualizacija;
- *rworldmap* – duomenų vizualizacija pasaulio žemėlapyje šalių ir tinklelių lygmenyje;
- *xts* – įvairių laike pagrįstų duomenų klasių valdymas;
- *qrmdata* – kiekybinių rizikos valdymo idėjų analizė;
- *tidyquant* – detali kiekybinė finansinė analizė;
- *proxy* – išplėstinės atstumų ir panašumo metrikos;
- *TSclust* – laiko eilučių klasterizavimas (laiko eilučių nepanašumo metrikos, specifiniai klasterizavimo metodai, klasterizavimo kokybės vertinimo metrikos) ;
- *dtwclust* – laiko eilučių klasterizavimas paremtas optimizuota technika naudojančia dinamine laiko išskyrimo atstumo metriką (*DTW*);
- *reshape2* – lankstus duomenų restruktūrizavimas ir agregavimas;
- *grid* – grafikų maketavimo įrankis;
- *dplyr* – greito ir nuoseklaus darbo su duomenų rinkiniais (naudojant kompiuterio atmintį ir ne) ;
- *Hmisc* – įvairios duomenų analizės ir jų manipuliavimo Harrell'o funkcijos;
- *corrplot* – koreliacinių matricių vizualizacija;

- *data.table* – greitas didelių duomenų rinkinių agregavimas, grupavimas, sujungimas, paralelinis duomenų nuskaitymas ir įrašymas;
- *Gmisc* – aprašomosios statistikos, duomenų skirtumų, kintamųjų parinkimo įrankis;
- *forecast* – laiko eilučių prognozės metodai ir tiesiniai modeliai;
- *car* – regresijos modelių rezultatų analizės funkcijos.

2.3. Prognozuojamas kintamasis

Kaip anksčiau minėta 1.1.2. skyrelyje šiame darbe tirsime akcijų pelningumo dydį – akcijų grąžą r_t , kurios dydis laiko momentu t apibrėžtas (1) formulėje. Kai kuriuose tyrimuose aprašytuose 1.1.4. segmente analizuojamas ir prognozuojamas akcijų grąžos natūralusis logaritmas [21]:

$$R_t = \ln(p_t) - \ln(p_{t-1}) = \ln\left(\frac{p_t}{p_{t-1}}\right); \quad (4)$$

čia p_t – akcijų kaina laiko momentu t , p_{t-1} – akcijų kaina prieš tai buvusiu laiko momentu.

Šiame tyrime taip pat tirsime grąžos logaritmą, kadangi aritmetinė grąža R_t turi teigiamą tendencingumą. Tarkime laiko momentu t_1 investavome į akciją, kurios kaina 100 piniginių vienetų, laiko momentu t_2 akcijos kaina pakyla iki 200 piniginių vienetų (pagal (4) formulę gaunama 100 % grąža), laiko momentu t_3 akcijos kaina nusmuko iki 100 piniginių vienetų (gaunama neigiama 50 % grąža – nuostolis). Trijų laiko momentų akcijos grąžos aritmetinis vidurkis lygus 25 % – gaunamas 25 % pelnas, nors iš tikrųjų nėra uždirbta nei pelno nei nuostolio. Šį realų nulinį pelną gautume radus logaritminių grąžų vidurkį (žr. 2.1. lentelę).

2.1 lentelė. Akcijų grąžų apskaičiavimas

Laiko momentas	Akcijos kaina	Grąža	Grąžos logaritmas
t_1	100	-	-
t_2	200	100 %	69 %
t_3	100	-50 %	-69 %
<i>Vidurkis</i>		25%	0%

Taip pat [3, 21] šaltiniuose išskiriami teoriniai privalumai analizuojant ir modeliuojant akcijų grąžos logaritmą. Čia teigiama, kad grąžos logaritmas įprastai yra pasiskirstęs pagal normalųjį skirstinį (dauguma matematinių metodų reikalauja šios prielaidos), kad sudėtinė grąža per tam tikrą skaičių periodų yra pradinio laiko momento ir paskutinio laiko momento grąžų logaritmų skirtumas (jei grąža nėra normaliojo skirstinio, tuomet pagal centrinę ribinę teoremą grąžų sumos skirstinys artėja į normalųjį), kad grąžos logaritmas yra tolydaus laiko stacionarus procesas.

Kadangi šiame darbe bus sudarinėjamas modelis akcijų grąžos priklausomybės nuo išorinių veiksnių, tokių kaip makroekonominiai rodikliai ir *Google* tendencijų, kurių duomenys yra kas

mėnesį, priklausomas kintamasis bus akcijų gražų logaritmų mėnesinis vidurkis lygus $\bar{R}_i = \frac{1}{m} \sum_{j=1}^m R_{i,j}$, kur $i = 1, 2, \dots, n$; $n = 63$ mėnesiai; m – dienų skaičius mėnesyje i .

2.4. Klasterizavimo metodų taikymas

2.4.1. Atstumo mato parinkimas

1.1.4. skyrelyje apžvelgus literatūroje naudojamus atstumo matus nustatant kiekybinį ryšį tarp dviejų laiko eilučių, pasirenkame vieną populiariausiai taikomą Euklido atstumą ir vieną naujesnių laiko eilutės forma paremta matą. Darbe palyginsime klasterizavimo kokybę naudojant šiuos abu matus.

Toliau aprašant matus ir algoritmus tarkime, kad \mathbf{x} ir \mathbf{y} tai dvi skirtingos laiko eilutės, tai yra reikšmių seka laikotarpyje n : $\mathbf{x} = (x_1, x_2, \dots, x_n)$, klasterizavimo rezultatas – k klasterių, kurie sudaro laiko eilučių aibes S_i , t. y. gaunamos laiko eilučių grupės $\mathbf{S} = (S_1, S_2, \dots, S_k)$, bendruoju atveju atstumo matas žymimas $d(\mathbf{x}, \mathbf{y})$.

Euklido atstumas (angl. *Euclidean distance, ED*), tai plačiausiai naudojama atstumų metrika. Šis atstumas nusako atstumą tarp dviejų laiko eilučių. Matas apibrėžiamas taip:

$$ED(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (5)$$

Laiko eilutės forma paremtas matas (angl. *shape-based distance, SBD*) pristatytas [19] straipsnyje. Šis matas 2015 metais buvo pasiūlytas, kaip dalis k -formos (angl. *k-shape*) klasterizavimo algoritmo. *SBD* principas – normuotų kryžminių koreliacijų (CC_c) seka tarp dviejų laiko eilučių x ir y . Mato reikšmės yra intervale $[0 ; 2]$, kuo mato reikšmė artimesnė nuliui, tuo laiko eilutės panašesnės. Matas apibrėžiamas taip:

$$SBD(\mathbf{x}, \mathbf{y}) = 1 - \frac{\max\{CC_c(\mathbf{x}, \mathbf{y})\}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}; \quad (6)$$

čia $\|\cdot\|_2$ – laiko eilutės l_2 norma (Euklidinė norma) lygi $\sqrt{\sum_{i=1}^n (x_i)^2}$, $CC_c(\mathbf{x}, \mathbf{y})$ – kryžminių koreliacijų seka, apskaičiuojama kaip Furje transformacijos ir jos atvirkštinės transformacijos sandaugos sąsūka. $CC_c(\mathbf{x}, \mathbf{y}) = \mathcal{F}^{-1}\{\mathcal{F}(\mathbf{x}) * \mathcal{F}^{-1}(\mathbf{y})\}$, kur $\mathcal{F}(\mathbf{x}_k) = \sum_{r=0}^{|\mathbf{x}|-1} x_r e^{-\frac{2jrk\pi}{|\mathbf{x}|}}$, $k = 0, \dots, |\mathbf{x}| - 1$, $\mathcal{F}^{-1}(\mathbf{x}_r) = \sum_{k=0}^{|\mathbf{x}|-1} \mathcal{F}(\mathbf{x}_k) e^{\frac{2jrk\pi}{|\mathbf{x}|}}$, $r = 0, \dots, |\mathbf{x}| - 1$.

2.4.2. Padalijimo pagrįstu klasterizavimo metodo jungimo algoritmai

Šiame darbe dalijimu pagrįstu klasterizavimo metodu, laiko eilutes jungsime į klasterius panaudojant k -vidurkio, dalijimo apie „medoidus“ (PAM) ir k -formos technikas. Pirmame

skyrelyje pateiktame 1.1. klasterizavimo padalijimo metodu algoritmo schemeje, objektų priskyrimas klasteriams vykdomas šiais algoritmais:

- k -vidurkio – minimizuojama klasterių objektų atstumų nuo centrų kvadratų suma:

$$\sum_{i=1}^k \sum_{j=1}^n d(x_j^{(i)}, c_i); \quad (7)$$

čia $d(x_j^{(i)}, c_i)$ – atstumas tarp elementų $x_j^{(i)}$ ir klasterio centro c_i . Metodas dažniausiai naudojamas su Euklido atstumu [20].

- dalijimo apie „medoidus“ (PAM) – skirstome į klasterius pagal artimiausius medoidus (vidurinius taškus), minimizuojama klasterių objektų atstumų nuo vidurinių taškų kvadratų suma, t. y. minimizuojama formulė (7), kurioje vietoje klasterio centro c_i naudojamas klasterio medoidas.

- k -formos (angl. *k-shape*) – klasterių centrai skaičiuojami minimizuojant atstumų kvadratų sumą tarp laiko eilučių. Centrai apskaičiuojami atlikus laiko eilučių transformaciją, maksimizuojant Reilėjaus koeficientus randamas tikrinis vektorius, kuris nusako transformacijos didžiausią tikrinę reikšmę, šis tikrinis vektorius laikomas klasterio centru. Metodas naudojamas kartu su SBD atstumu. Platesnis jo aprašymas pateiktas [19] straipsnyje.

2.4.3. Hierarchinio klasterizavimo jungimo metodai

Šiame darbe taikysime hierarchinės klasterinės analizės metodus – laiko eilutes jungsime į klasterius panaudojant vidutinės, pilnosios ir vienetinės jungties technikas. Pirmame skyrelyje pateiktame 1.2 klasterizavimo hierarchiniu metodu algoritmo schemeje, objektų sujungimas į klasterius vykdomas šiais algoritmais [20]:

- Vidutinės jungties – metodas skaičiuoja visų galimų klasterių porų vidutinį atstumą ir jungiamos tos poros į klasterį, kur šis atstumas minimalus:

$$\frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_{r_i}, x_{s_j}). \quad (8)$$

- Tolimiausio kaimyno arba pilnosios jungties – skaičiuojamas atstumas tarp labiausiai nutolusių taškų:

$$\max_{1 \leq r_i \leq n_r, 1 \leq s_j \leq n_s} d(x_{r_i}, x_{s_j}). \quad (9)$$

- Artimiausio kaimyno arba vienetinės jungties – tarp arčiausių taškų:

$$\min_{1 \leq r_i \leq n_r, 1 \leq s_j \leq n_s} d(x_{r_i}, x_{s_j}). \quad (10)$$

2.4.4. Klasterizavimo kokybės indeksai

Norint įvertinti klasterizavimo rezultatus – įvairių metodų tikslumą, nustatyti optimalų klasterių skaičių imtyje, reikia atlikti kokybinį vertinimą remiantis tam tikromis funkcijomis.

Šiame darbe naudosime vidinius kokybės vertinimo indeksus, kadangi neturime nei faktinės nei ekspertinės informacijos apie galimas duomenų grupes, o klasterizavimo tikslas yra sumažinti tiriamų duomenų kiekį, nagrinėjant gautus klasterius atskirai. Šiame darbe analizuosime tokius klasterizavimo rezultatų patikros indeksus:

- **Silhouette indeksas** palygina vidutinius atstumus tarp laiko eilučių klasteryje su vidutiniais atstumais tarp kitų klasterių [13]. Koeficiento reikšmių intervalas lygus $[-1;1]$, klasterizavimas laikomas sėkmingu, jeigu koeficiento reikšmė artima 1. Indeksas apibrėžiamas taip:

$$Sil_{index} = \frac{1}{n} \sum_{j=1}^n Sil(\mathbf{x}_j), \quad (11)$$

$$Sil(\mathbf{x}) = \frac{q(\mathbf{x}) - p(\mathbf{x})}{\max((p(\mathbf{x}), q(\mathbf{x})))}; \quad (12)$$

čia $p(\mathbf{x})$ – vidutinis laiko eilutės x nepanašumas (atstumas) į kitas to paties klasterio laiko eilutes (kuo mažesnė ši reikšmė, tuo laiko eilutė x geriau priskirta klasteriui), $q(\mathbf{x})$ – mažiausias vidutinis laiko eilutės x nepanašumas (atstumas) į kitus klasterius, kuriems nepriklauso x -oji eilutė. Pagrindinis indekso naudojimo trūkumas – negalimas apskaičiuoti, kai klasterį sudaro vienas elementas.

- **Dunn indeksas** nustato gerai atskirtas klasterių grupes (maksimalus atstumas tarp klasterių, minimalus jų viduje). Kuo didesnė indekso reikšmė, tuo kokybiškesnis klasterizavimas. Indeksas apibrėžiamas taip:

$$D_{index} = \frac{\min_{1 \leq \mathbf{k}_1 < \mathbf{k}_2 \leq k} d(\mathbf{k}_1, \mathbf{k}_2)}{\min_{1 \leq \mathbf{k}_3 \leq k} d'(\mathbf{k}_3)}; \quad (13)$$

čia $d(\mathbf{k}_1, \mathbf{k}_2)$ – atstumas tarp klasterių \mathbf{k}_1 ir \mathbf{k}_2 , $d'(\mathbf{k}_3)$ – yra klasterio \mathbf{k}_3 vidinis atstumas (skersmuo), k – klasterių skaičius. Pagrindinis indekso naudojimo trūkumas toks pat kaip ir Silhouette indekso.

- **Davies-Bouldin indeksas**, kaip ir *Dunn* indeksas nustato gerai atskirtas klasterių grupes. Mažos koeficiento reikšmės rodo klasterių didelį atstumą vienas nuo kito. Indeksas apibrėžiamas taip:

$$DB_{index} = \frac{1}{k} \sum_{\mathbf{k}_1=1}^k \max_{\mathbf{k}_1 \neq \mathbf{k}_2} \left\{ \frac{d'(\mathbf{k}_1) + d'(\mathbf{k}_2)}{d(\mathbf{k}_1, \mathbf{k}_2)} \right\}; \quad (14)$$

čia $d'(\cdot)$ ir $d(\cdot, \cdot)$ funkcijos apibrėžiamos taip pat kaip ir *Dunn* indekse.

2.5. Ypatingųjų reikšmių dekompozicija

Ypatingųjų reikšmių dekompozicija (angl. *singular value decomposition*, SVD) tai algoritmas, kurio pagalba duomenų matrica \mathbf{X} turinti $m \times n$ ($m \geq n$) elementų yra išskaidoma tokiu būdu: $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, čia \mathbf{U} – $m \times n$ dydžio matrica (matricos stulpeliai vadinami kairiaisiais ypatingaisiais vektoriais), \mathbf{S} – $n \times n$ dydžio diagonalioji matrica su teigiamais elementais pagrindinėje įstrižainėje ir nuliniiais kitais elementais (elementai vadinami ypatingosiomis reikšmėmis), \mathbf{V}^T taip pat $n \times n$ dydžio matrica (matricos eilutės vadinamos dešiniaisiais ypatingaisiais vektoriais) [24].

SVD metodu apskaičiuojamos matricos \mathbf{V}^T ir \mathbf{S} diagonalizuojant $\mathbf{X}^T\mathbf{X}$ taip kad, $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{S}^2\mathbf{V}^T$, tuomet randama matrica \mathbf{U} pagal $\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ [24].

Šis metodas yra tiesiogiai susietas su principinių komponentių analize (angl. *Principal Component Analysis*, PCA) tikrinių reikšmių skaičiavimu iš kovariacinės matricos. Todėl matrica $\mathbf{U}\mathbf{S}$ sudaro principinių komponentių įverčiai (mūsų analizuojamu atveju, tai laiko eilučių koordinatės principinių komponentių erdvėje), dešinieji ypatingieji vektoriai – tai principinės komponentės, diagonalioji matrica – principinių komponentių dispersijos proporcijos. Šis metodo taikymas taip pat turi sąsają su Furje analize, pradinių duomenų transformacija ortogonalioje bazėje. Plačiau šis metodas aprašytas [24] straipsnyje.

Remiantis profesoriaus Jeff Leek [25] šaltinyje pateiktas SVD metodo taikymas, norint surasti labiausiai duomenų rinkinyje dominuojantį kintamąjį – „adatą šieno kupetoje“. Iš SVD metodu gautos \mathbf{V} matricos stulpelių atrenkamos maksimalios reikšmės, turinčios didžiausią įtaką tam tikroje komponentėje (įprastai, pirmajame stulpelyje – reikšmingiausioje komponentėje).

2.6. Prognozavimo metodų taikymas

Supaprastinant formulių užrašymą pažymime Y_t – analizuojamos akcijos vidutinė mėnesinė uždarymo kainos logaritminė grąža laiko momentu t ($Y_t = \bar{R}_t$), analizuojami makroekonominiai ir *Google* tendencijų kintamieji laiko momentu t yra $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(n)})$, \hat{Y}_t – grąžos prognozė laiko momentu t .

2.6.1. Apibendrintas tiesinis modelis – regresija

Šiame darbe priklausomam kintamajam – akcijų grąžos mėnesiniam vidurkiui taikysime vieną paprasčiausių GLM atvejų (tarę, kad priklausomo kintamojo skirstinys normalusis) – daugialypę tiesinę regresiją. Toks modelis nustato kiekybinio kintamojo Y_t tiesinę priklausomybę nuo keleto nepriklausomų kintamųjų \mathbf{X}_t :

$$Y_t = \beta_0 + \beta_1 X_t^{(1)} + \beta_2 X_t^{(2)} + \dots + \beta_n X_t^{(n)} + \varepsilon_t; \quad (15)$$

čia $\beta_0, \beta_1, \dots, \beta_n$ – modelio koeficientai, ε_t – atsitiktinė paklaida.

Modelio parametrai įvertinami didžiausio tikėtimumo metodu naudojant iteracinį svertinį mažiausių kvadratų algoritimą.

Pagrindinės tiesinės regresinės analizės prielaidos (reikalavimai, kuriuos turi tenkinti atsitiktinės paklaidos ε_t):

- atsitiktinės paklaidos ε_t yra normaliai pasiskirstę atsitiktiniai dydžiai;
- ε_t vidurkis lygus nuliui – $\mathbf{E}\varepsilon_t = 0$;
- ε_t dispersijos lygios (homoskediško prielaida) – $\mathbf{D}\varepsilon_t = \sigma^2$;
- ε_t yra nepriklausomi dydžiai.

2.6.2. Autoregresinis neuroninis tinklas

Norint prognozuoti laiko eilučių ateities vertes priklausomas nuo kelto kintamųjų laiko srities praeities verčių, taikomas modifikuotas neuroninis tinklas, kuris vadinamas netiesiniu autoregresiniu tinklu su išoriniais įėjimais (NARX). Šis neuroninių tinklų metodas taikytas 1.1.4. skyriuje aprašytuose [16, 15] šaltiniuose atliktuose tyrimuose akcijų kainų kitimui prognozuoti.

Tai tiesioginio perdavimo architektūros vieno sluoksnio neuroninis tinklas, kuris apmokomas su vėlintomis regresorių reikšmėmis. Šis modelis nagrinėjamu atveju matematiškai aprašomas tokia lygtimi:

$$Y_t = \varphi(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p_Y}, \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_{t-p_X}); \quad (16)$$

čia p_Y – nepriklausomo kintamojo vėlinimų dydis, p_X – priklausomų kintamųjų vėlinimų dydis, φ – neuroninio tinklo svorių ir aktyvacijos funkcijos architektūra.

1.1.4 analizuotoje literatūroje teigiama, kad dažniausiai taikomas vieno paslėptojo sluoksnio tinklas, kuris su tinkamai parinktu neuronų skaičiumi gali gauti geras prognozes. Čia taip pat patvirtintas akcijų rinkų dinamiškumas, t. y. akcijų kainą veikia realaus laiko išoriniai faktoriai. Dėl šių priežasčių šiame darbe pasirenkame vieno sluoksnio tinklą, o laiko eilučių vėlinimą lygų vienam, kadangi prognozuojamos bus mėnesinės akcijų grąžos. Išėjimo sluoksnio neuronų skaičius lygus vienam, kadangi prognozuojama tolydi reikšmė. Paslėptojo sluoksnio neuronų skaičių nustatysime eksperimentiniu būdu.

Šiame darbe paslėptojo ir išėjimo sluoksnių neuronų aktyvacijos funkcija naudojama logistinė sigmoidinė funkcija:

$$\text{logsig}(x) = (1 + e^x)^{-1} \quad (17)$$

Neuroninio tinklo parametrai optimizuojami kvazi-Niutono BFGS metodu (angl. *Broyden, Fletcher, Goldfarb and Shanno*).

Pagal anksčiau apibrėžtą neuroninio tinklo bendrojo atvejo (2) formulę mūsų atveju šioje formulėje išėjimo sluoksnio neuronų skaičius lygus 1, o aktyvacijos funkcija $f = \text{logsig}$.

Šiame darbe bus sudaroma 50 neuroninių tinklų su skirtingais pradiniais svoriais, o prognozuojama reikšmė, tai gautų 50-ies tinklų išėjimo sluoksniu neuronų vidurkis.

2.6.3. Prognozės įvertinimo metrikos

Iš 1.2.2. sekcijoje pristatytų prognozavimo modelių kokybės įvertinimo metrikų šiame darbe taikysime šias metrikas:

- šaknis iš vidutinės kvadratinės paklaidos (angl. *Root Mean Squared Error*, RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{Y}_t - Y_t)^2}{n}}; \quad (18)$$

- anksčiau aprašytame straipsnyje pristatytas prognozavimo kokybės įvertinimas remiantis prognozės krypties pokyčiu (angl. *Prediction On Change In Direction*, POCID) [14]:

$$POCID = 100 \cdot \frac{\sum_{t=1}^n D_t}{n}; \quad (19)$$

$$D_t = \begin{cases} 1, & (\hat{Y}_t - \hat{Y}_{t-1}) \cdot (Y_t - Y_{t-1}) > 0, \\ 0, & \text{kitu atveju} \end{cases}; \quad (20)$$

Ši metrika, parodo kiek procentų prognozuotų akcijų gražos yra teisinga kryptimi, t. y. kiek teisingai suprognozuota akcijos gražos kilimų ir nuosmukių.

- įvedama nauja ženklų sutapimo metrika:

$$POS = 100 \cdot \frac{\sum_{t=1}^n S_t}{n} \quad (21)$$

$$S_t = \begin{cases} 1, & \text{sgn}(\hat{Y}_t) = \text{sgn}(Y_t), \\ 0, & \text{kitu atveju} \end{cases} \quad (22)$$

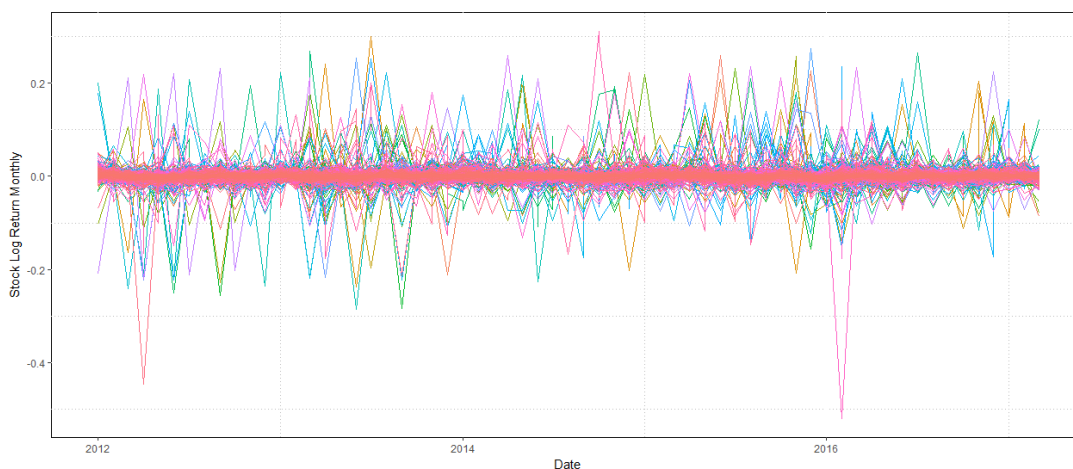
Ši metrika, parodo kiek procentų prognozuotų akcijų gražos yra teisingo ženklo, t. y. kiek teisingai suprognozuota akcijos teigiama ir neigiama gražos (pelnas ar nuostoliai).

3. TYRIMŲ REZULTATAI IR JŲ APTARIMAS

Šioje dalyje sprendžiant 1.3. poskyryje išsikeltą uždavinį ištirsime analizuojamas akcijų duomenų, *Google* tendencijos, makroekonominių rodiklių laiko eilutes. Po to atliksime akcijų klasterizavimų algoritmų lyginamąją analizę išskirdami optimalų skaičių akcijų klasterių. Kiekviename klasteryje ypatingų reikšmių dekompozicijos metodu atrinksime reikšmingiausią laiko eilutę, kurioms atskirai atliksime prognozės modelių lyginamąją analizę, taip sudarydami geriausią modelį kiekvienos grupės akcijų vidutinei grąžai prognozuoti.

3.1. Pirminė duomenų analizė

Analizuosime 3 206 akcijų laiko eilutes Y_t , kur laiko periodas $t = 1, 2, \dots, 63$ lygus vienam mėnesiui (nuo 2012 01 01 iki 2017 03 31). Visų akcijų vidutinės mėnesinės grąžos kitimas pavaizduotas 3.1 grafike. Iš šio grafiko matome, kad akcijų grąža kinta gana chaotiškai, o daugumos akcijų grąža koncentruojasi apie nulinę grąžos ašį, t. y. grąža artima nuliui, yra išskirtinių akcijų, kurių maksimali/ minimali vidutinė grąža yra lygi apytiksliai ± 0.1 ir ± 0.2 . Taip pat iš grafiko matome, kad iš visos imties išsiskiria keletas akcijų grąžų, kurių maksimali/ minimali grąža didesnė/ mažesnė už ± 0.3 .

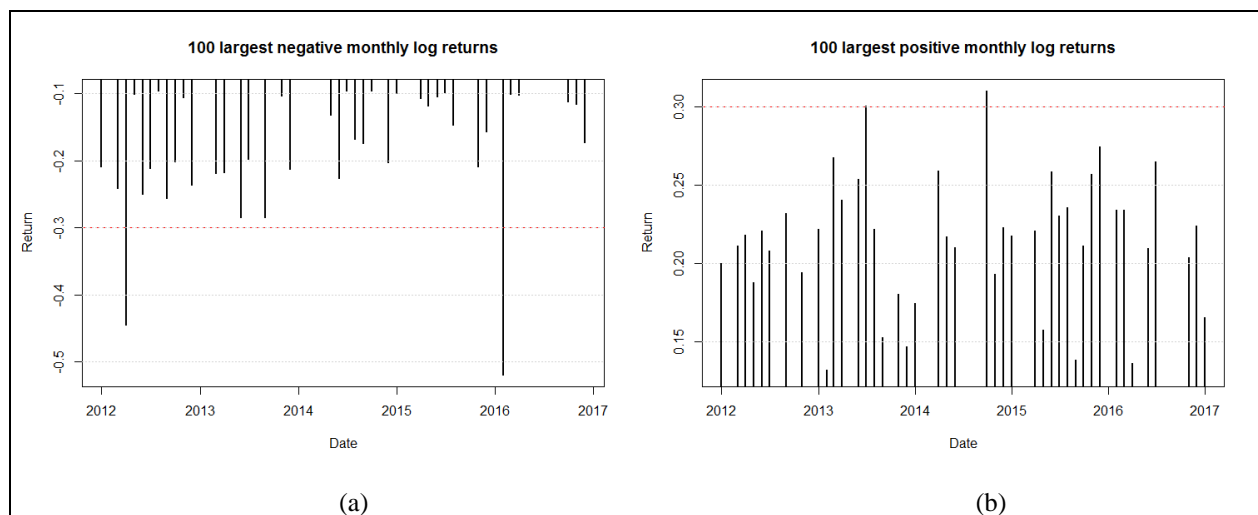


3.1. pav. Analizuojamų akcijų laiko eilučių kitimas

3.1.1. Ekstremalių reikšmių tyrimas

Iš akcijų grąžų kitimo laike, matome, kad duomenyse egzistuoja daug išskirčių. Tačiau pašalinti visas išskirtis būtų nekorektiška, kadangi tolimesniame etape grupuosime akcijas į klasterius ir išskirtys gali sudaryti atskirus klasterius. Akcijas, kurių momentinės grąžos ypatingai didelės ar ypatingai mažos naudinga pašalinti iš imties, kadangi tokių momentinių grąžos šuolių yra tik keletas.

Ekstremalių reikšmių nustatymui atliekame šimto didžiausių ir mažiausių momentinių akcijų grąžų tyrimą, šios reikšmės atvaizduotos 3.2. grafikuose. Iš jų (a) dalyje matome žymiai išsiskiriančias dvi reikšmes, o (b) dalyje taip pat išsiskiriančias dvi reikšmes.

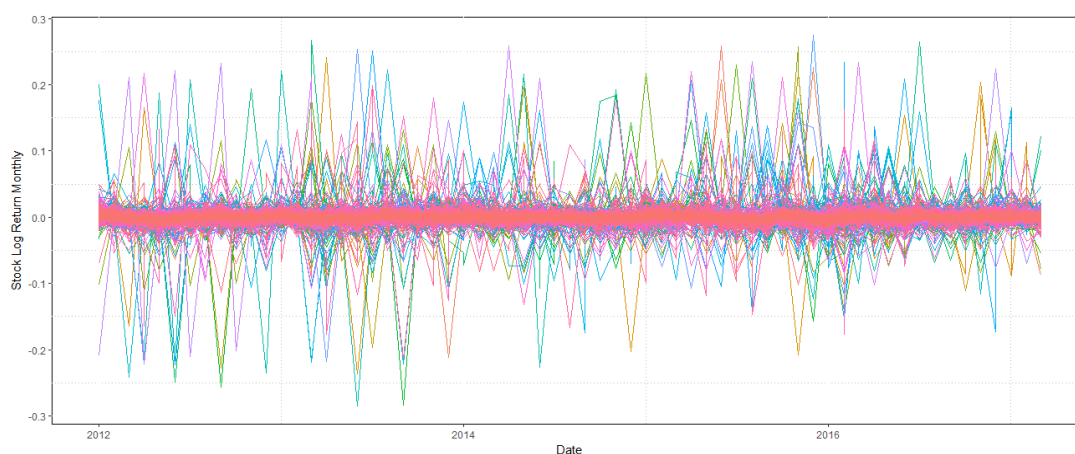


3.2. pav. Ekstremalių reikšmių tyrimas (slenkstis raudona kontūrinė linija). (a) dalyje 100-as mažiausių gražų reikšmių; (b) – 100-as didžiausių gražų reikšmių;

Peržiūrėjus šių keturių ypatingų gražų šuolių laiko eilutes, galime padaryti išvadas:

- Įmonė „Migme“ – akcijų kainos nuosmukis 2016 02 mėnesį;
- Įmonė „Zeeland Family Oyj“ – akcijų kainos nuosmukis 2012 04 mėnesį;
- Įmonė „Biome Technologies plc“ – akcijų kainos šuolis 2013 07 mėnesį;
- Įmonė „Marenica Energy“ – akcijų kainos šuolis 2014 10 mėnesį;

Šių keturių įmonių laiko eilutes pašaliname iš bendro duomenų rinkinio ir tolimesniame tyrime jų neanalizuosime (liko 3 202 akcijų, bendrai įrašų 201 726). Likusių akcijų vidutinės mėnesinės gražos kitimas pavaizduotas 3.3 grafike. Jų minimali reikšmė lygi -0.29, mediana (dažniausiai pasikartojanti reikšmė) 0.00015, vidurkis 0.00025, maksimali reikšmė 0.27.



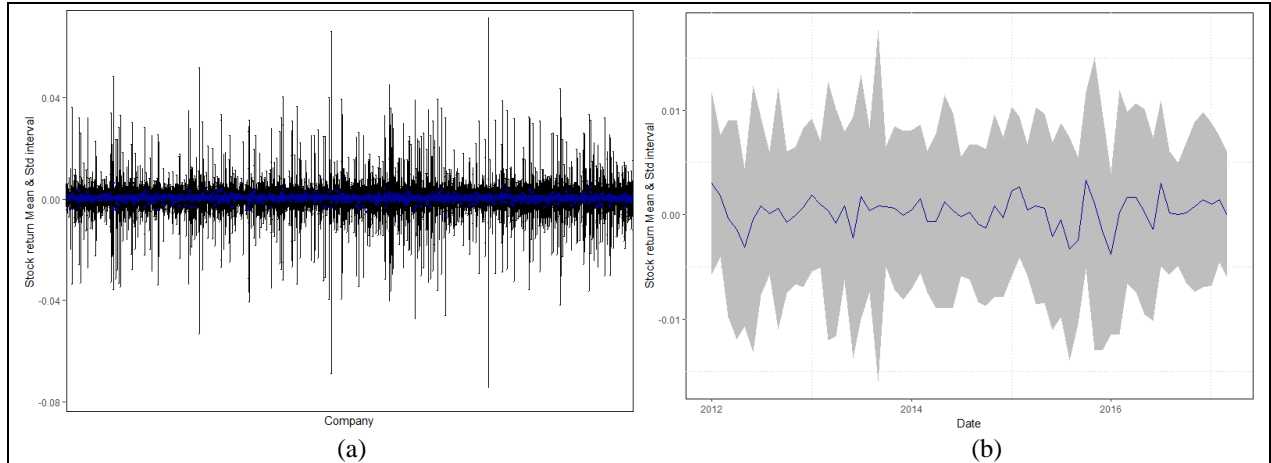
3.3. pav. Analizuojamų akcijų laiko eilučių kitimas pašalinus ekstremalias laiko eilutes

3.1.2. Vidurkio ir sklaidos apie jį tyrimas

Vienas iš akcijų nepastovumo (kintamumo) dydžių, kuris matuoja investavimo riziką yra akcijų gražos standartinis nuokrypis σ – sklaida apie vidutinę akcijų vertę. Analizuojame vidurkio

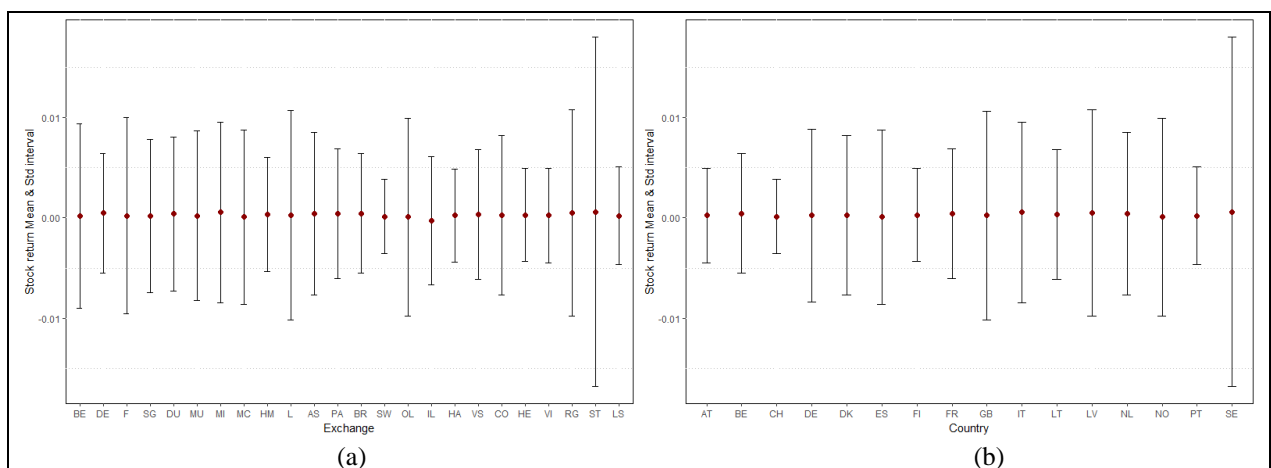
ir standartinio nuokrypio intervalo ($\bar{Y}_{level} - \sigma; \bar{Y}_{level} + \sigma$) kitimą įmonių, šalių, biržų ir mėnesių lygiuose.

Iš 3.4 paveikslo (a) dalies, matome, kad vidutinė akcijų per įmones grąža svyruoja apie nulinę vertę, o sklaida kinta gana įvairiai, tai lemia duomenų rinkinio akcijų įvairovė. Todėl toliau atliksime klasterizavimą. Iš (b) dalies mėnesinių visų įmonių vidurkių – vidurkis ir sklaida apie jį gana nestabiliai kinta 2012, 2015, 2016 metais. Tikėtina, kad šį kitimą įtakoja išoriniai faktoriai.



3.4. pav. Analizuojamų akcijų grąžos vidurkio ir standartinio nuokrypio kitimas. (a) – įmonių lygyje; (b) – mėnesių lygyje;

Platesniame biržų lygyje vidurkio vidutinė vertė išlieka tokia pat kaip ir įmonių lygyje (žr. 3.5 pav.), sklaidos atžvilgiu išsiskiria Stokholmo vertybinių popierių biržos (viso 40 įmonių) akcijų sklaida, (šioje rinkoje daugiausia rizikingų akcijų), mažiausiai akcijų grąža kinta Šveicarijos vertybinių biržoje (viso 51 įmonė). Analogiškai ir analizuojant platesniame šalių lygyje išsiskiria Šveicarija ir Kopenhaga.



3.5. pav. Analizuojamų akcijų grąžos vidurkio ir standartinio nuokrypio kitimas biržų lygyje. (a) – biržų lygyje; (b) – šalių lygyje;

3.1.3. Autokoreliacijos tyrimas.

Autokoreliacija – tai laiko eilutės ir tos pačios eilutės su laiko vėlinimu koreliacijos dydis, nustatantis ateities reikšmės priklausomumą nuo istorinių reikšmių. Nustatyti ar laiko eilučių duomenys yra nepriklausomi atliekame *Ljung-Box* testą kiekvienos akcijos gražos laiko eilutei atskirai, čia tikrinamos hipotezės:

H_0 : duomenys yra nepriklausomi (nėra autokoreliacijos);

H_a : duomenyse yra daline koreliacija.

Atlikus testą 3 202 laiko eilutėms, gauti rezultatai su 95 % reikšmingumo lygmeniu pateikti 3.1 lentelėje.

3.1 lentelė. *Ljung-Box* testo rezultatai.

	Akcijų skaičius
H_0 neatmesta	2 953
H_0 atmesta	249

Iš autokoreliacijos tyrimo rezultatų gauta, kad 92.22 % visų analizuojamų laiko eilučių autokoreliacija yra nereikšminga.

3.1.4. Stacionarumo tyrimas

Laiko eilutė laikoma stacionariu procesu, kai jos vidurkis pastovus, dispersija pastovi, stebėjimų kovariacija nepriklauso nuo laiko, o priklauso tik nuo vėlinimo. Integruotumo eilei nustatyti dažniausiai naudojamas vienietinės šaknies išplėstinis *Dickey-Fuller* (angl. *augmented Dickey-Fuller*) testas, tikrinama hipotezė kiekvienos akcijos laiko eilutei atskirai:

H_0 : kintamasis nėra stacionarus ir turi būti integruotas bent pirma eile;

H_a : kintamasis yra stacionarus.

Atlikus testą 3 202 laiko eilutėms, gauti rezultatai su 95 % reikšmingumo lygmeniu pateikti 3.2 lentelėje.

3.2 lentelė. *Dickey-Fuller* testo rezultatai.

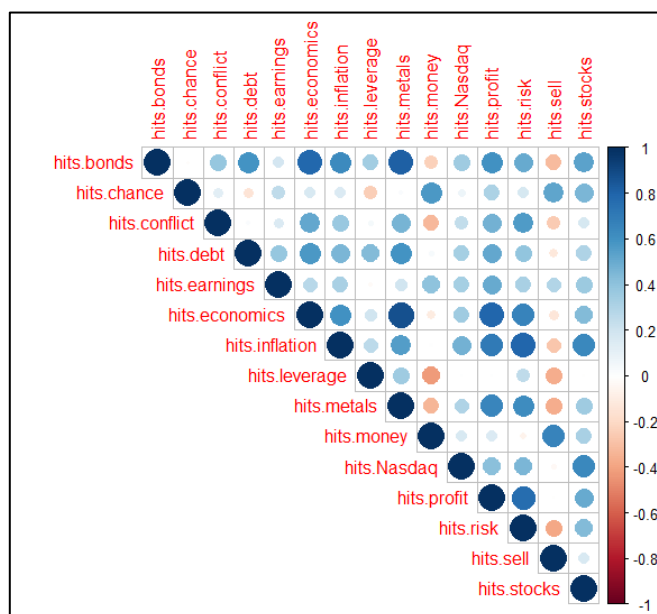
	Akcijų skaičius
H_0 neatmesta	19
H_0 atmesta	3183

Gauta, kad 99.41 % visų analizuojamų duomenų yra stacionarūs procesai.

3.1.5. Multikolinearumo tyrimas

Multikolinearumas – tai koreliacinio ryšio tyrimas tarp akcijų gražą galimai įtakančių veiksnių.

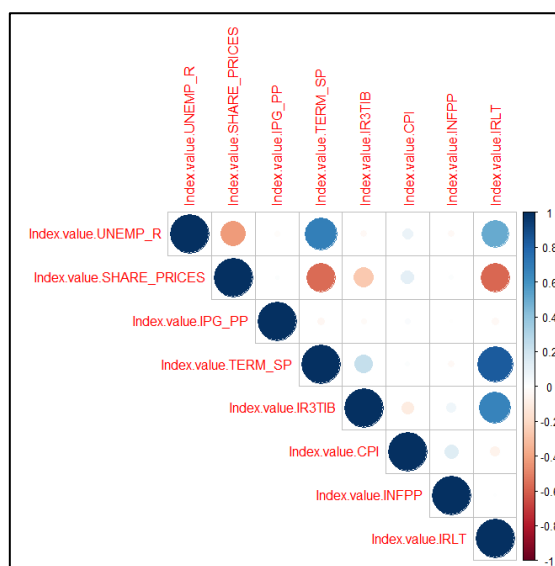
Pirma, tiriame multikolinearumą tarp finansinių terminų *Google* tendencijos duomenų. Šiam tikslui randame Spirmeno ranginės koreliacijos koeficientus, kurie apibūdina ryšio tarp dviejų kintamųjų stiprumą monotoniškumo prasme, t. y. didėjant vienam kintamajam, monotoniškai (nebūtinai tiesiškai) didėja kitas kintamasis. Koreliacijos koeficientų matrica vizualine forma pateikta 3.6 paveiksle, čia kuo stipresnis ryšis tuo apskritimas kintamųjų sankirtoje didesnis (mėlyni atspalviai – teigiama koreliacija, raudoni – neigiama).



3.6. pav. Finansinių terminų *Google* tendencijos koreliacinė matrica

Kaip matome iš koreliacijos matricos, egzistuoja gana stiprus ryšis ($|\rho_{X,Y}| \geq 0.5$) tarp paieškos terminų: *bonds* ir *economics* ($|\rho_{X,Y}| = 0.79$), *bonds* ir *metals* ($|\rho_{X,Y}| = 0.81$), *economics* ir *metals* ($|\rho_{X,Y}| = 0.87$), *economics* ir *profit* ($|\rho_{X,Y}| = 0.79$), *economics* ir *risk* ($|\rho_{X,Y}| = 0.67$), *inflation* ir *risk* ($|\rho_{X,Y}| = 0.79$), *inflation* ir *stocks* ($|\rho_{X,Y}| = 0.65$), *money* ir *sell* ($|\rho_{X,Y}| = 0.66$), *profit* ir *risk* ($|\rho_{X,Y}| = 0.76$). Tad multikolinearumo susilpninimui iš duomenų rinkio pašaliname šiuos kintamuosius: *hits.metals*, *hits.economics*, *hits.risk*, *hits.profit*, *hits.bonds*, *hits.stocks*, *hits.sell*.

Antra, analogiškai tiriame multikolinearumą tarp makroekonominių rodiklių duomenų. Spirmeno koreliacijos koeficientų matrica vizualine forma pateikta 3.7 paveiksle. Stiprus ryšis egzistuoja tarp kintamųjų: *TERM_SP* ir *UNEMP_R* ($|\rho_{X,Y}| = 0.68$), *IRLT* ir *UNEMP_R* ($|\rho_{X,Y}| = 0.50$), *SHARE_PRICES* ir *TERM_SP* ($|\rho_{X,Y}| = -0.56$), *SHARE_PRICES* ir *IRLT* ($|\rho_{X,Y}| = -0.57$), *TERM_SP* ir *IRLT* ($|\rho_{X,Y}| = 0.83$), *IR3TIB* ir *IRLT* ($|\rho_{X,Y}| = 0.66$). Multikolinearumo susilpninimui iš duomenų rinkio pašaliname šiuos kintamuosius: *Index.value.IRLT*, *Index.value.TERM_SP*.



3.7. pav. Makroekonominių rodiklių koreliacinė matrica

3.1.6. Koreliacijos tyrimas tarp gražos ir nepriklausomų kintamųjų

Atliekame koreliacijos tyrimą akcijų gražos ir nepriklausomų kintamųjų – *Google* finansinių terminų tendencijų, *Google* įmonės pavadinimo tendencijų ir makroekonominių rodiklių.

Pirma, atliekama analizė tarp kiekvienos akcijos duomenų ir analizuojamų *Google* finansinių terminų tendencijų. Tam tikslui atliekame Spirmeno koreliacijos koeficiento reikšmingumo testą, tikrinama hipotezė tarp kiekvienos akcijos duomenų ir analizuojamų finansinių terminų paieškos tendencijų:

$$H_0: \rho_{X,Y} = 0$$

H_a : duomenyse yra daline koreliacija.

Atlikus testą, gauti rezultatai su 95 % reikšmingumo lygmeniu pateikti 3.3 lentelėje. Iš kurių galime teigti, kad tarp beveik visų šių kintamųjų porų, koreliacija yra arba nereikšminga arba silpna.

3.3 lentelė. Spirmeno koreliacijos testo rezultatai (finansinių terminų).

	Porų skaičius		Porų skaičius
H_0 neatmesta	45 112	$ \rho_{X,Y} \geq 0.5$	3
H_0 atmesta	2 918	$ \rho_{X,Y} < 0.5$	48 027

Antra, analogiškai atliekama analizė tarp kiekvienos akcijos duomenų ir makroekonominių rodiklių, rezultatai pateikti 3.4 lentelėje. Čia matome, kad tarp beveik visų šių kintamųjų porų, koreliacija yra arba nereikšminga arba silpna.

3.4 lentelė. Spirmeno koreliacijos testo rezultatai (makroekonominių rodiklių).

	Porų skaičius		Porų skaičius
H_0 neatmesta	24 358	$ \rho_{X,Y} \geq 0.5$	8
H_0 atmesta	1 258	$ \rho_{X,Y} < 0.5$	25 608

Antra, analogiškai atliekama analizė tarp kiekvienos akcijos duomenų ir *Google* įmonės pavadinimo tendencijų, rezultatai pateikti 3.5 lentelėje. Čia matome, kad tarp beveik visų šių kintamųjų porų, koreliacija yra arba nereikšminga arba silpna.

3.5 lentelė. Spirmeno koreliacijos testo rezultatai (makroekonominių rodiklių).

	Porų skaičius		Porų skaičius
H_0 neatmesta	2 970	$ \rho_{X,Y} \geq 0.5$	0
H_0 atmesta	232	$ \rho_{X,Y} < 0.5$	3 202

3.2. Klasterizavimo metodų tyrimas

Iš ankstesniame skyrelyje atliktos pirminės duomenų analizės nustatėme, kad duomenys tarp 3 202 akcijų yra gana skirtingi, todėl norint sudaryti prognozės modelį reikia akcijas sugrupuoti pagal tam tikrus panašumo matavimus. Tam tikslui atliekame klasterizavimo analizę panaudojant antrame skyriuje aprašytus metodus, jungimo technikas, atstumo metrikas ir kokybės įvertinimo metrikas.

Pradžioje duomenis klasterizuojame taikydami padalijimu pagrįstą ir hierarchiniu metodus. Kiekvienam jų nustatome optimalų klasterių skaičių pagal *Silhouette*, *Davies-Bouldin* ir *Dunn* indeksus. Tam tikslui atliekame kiekvieno metodo realizaciją kintant klasterių skaičiui nuo 2 iki 15.

3.2.1. Padalijimu pagrįstų metodų tyrimas

Padalijimu paremtų metodų optimalaus klasterių skaičiaus parinkimo tyrimo rezultatai pateikti 3.8 paveikslo grafikuose (mėlynos kreivės nurodo, kad optimalus klasterių skaičius yra minimizuojant kokybės indeksą, o raudonos – maksimizuojant).

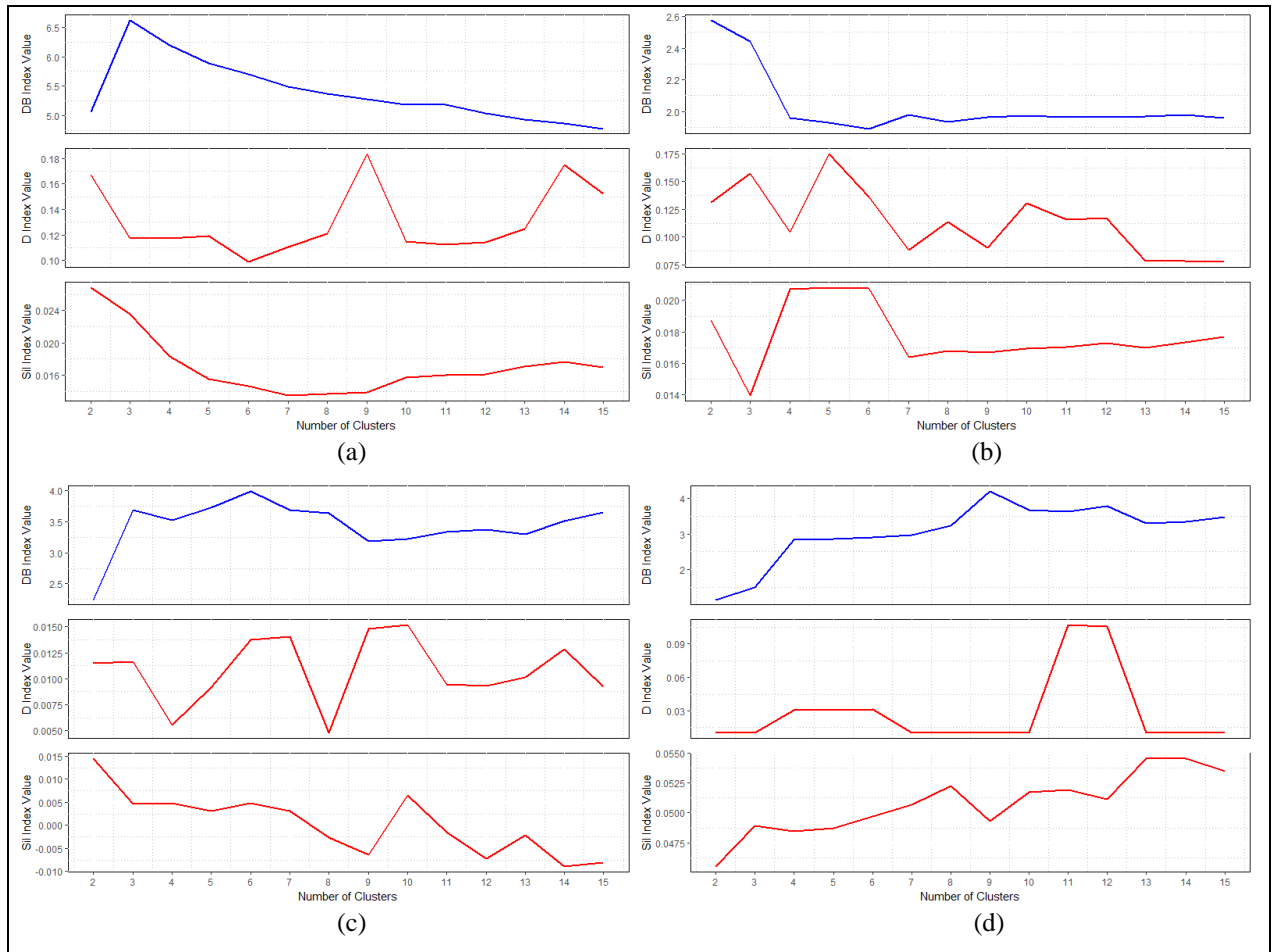
Iš 3.8 pav. (a) dalies, kai naudojama Euklido atstumo metrika, jungimo metodas – vidurkis, gautas optimalus klasterių skaičius lygus pagal $DB_{index} - 14$, pagal $D_{index} - 9$, pagal $Sil_{index} - 2$. Šiuo atveju parenkame optimalų skaičių lygų 2 pagal Sil_{index} indeksą (kadangi jis įvertina atstumus kartu klasterio viduje ir tarp klasterių), čia $Sil_{index} = 0.027$, $D_{index} = 0.167$, $DB_{index} = 5.062$. Sudaryti du klasteriai, kur viename 1 115 akcijos, kitame 2 087.

Iš (b) dalies, kai naudojama Euklido atstumo metrika, jungimo metodas – dalijimo apie „medoidus“ (PAM), gautas optimalus klasterių skaičius lygus pagal $DB_{index} - 6$, pagal $D_{index} - 5$, pagal $Sil_{index} - 5$. Parenkame optimalų skaičių lygų 5 pagal Sil_{index} indeksą, čia $Sil_{index} = 0.021$, $D_{index} = 0.175$, $DB_{index} = 1.972$. Sudaryti penki klasteriai, kur pirmame 511 akcijos, antrame – 383, trečiame – 462, ketvirtame – 1 401, penktame – 445.

Iš (c) dalies, kai naudojama forma paremta atstumo metrika (SBD), centrų skaičiavimo funkcija – vidurkis, gauta optimalus klasterių skaičius lygus pagal $DB_{index} - 2$, pagal $D_{index} - 10$,

pagal $Sil_{index} - 2$. Parenkame optimalų skaičių lygų 2, čia $Sil_{index} = 0.015$, $D_{index} = 0.012$, $DB_{index} = 2.246$. Sudaryti du klasteriai, kur pirmame 1 005 akcijos, antrame – 2 197.

Iš (d) dalies, kai naudojama forma paremta atstumo metrika (SBD), centrų skaičiavimo funkcija – k -formos, gautas optimalus klasterių skaičius lygus pagal $DB_{index} - 2$, pagal $D_{index} - 11$, pagal $Sil_{index} - 14$. Parenkame optimalų skaičių lygų 14, čia $Sil_{index} = 0.055$, $D_{index} = 0.010$, $DB_{index} = 3.355$. Sudaryta 14 klasterių, kuriuose atitinkamas akcijų skaičius: 190, 168, 143, 120, 176, 163, 223, 186, 568, 137, 120, 648, 151, 209.



3.8. pav. *Silhouette*, *Davies-Bouldin* ir *Dunn* indeksų priklausomybė nuo klasterių skaičiaus. (a) dalyje dalijimo klasterizavime padalijimo metodu naudojama Euklido atstumo metrika, centrų skaičiavimo funkcija – k -vidurkis; (b) – Euklido atstumas, centrų skaičiavimo funkcija – dalijimo apie „medoidus“ (PAM); (c) dalyje klasterizavime padalijimo metodu naudojama forma paremta atstumo metrika (SBD), centrų skaičiavimo funkcija – k -vidurkis; (d) – forma paremta atstumo metrika (SBD), centrų skaičiavimo funkcija – k -formos

3.2.2. Hierarchinių metodų tyrimas

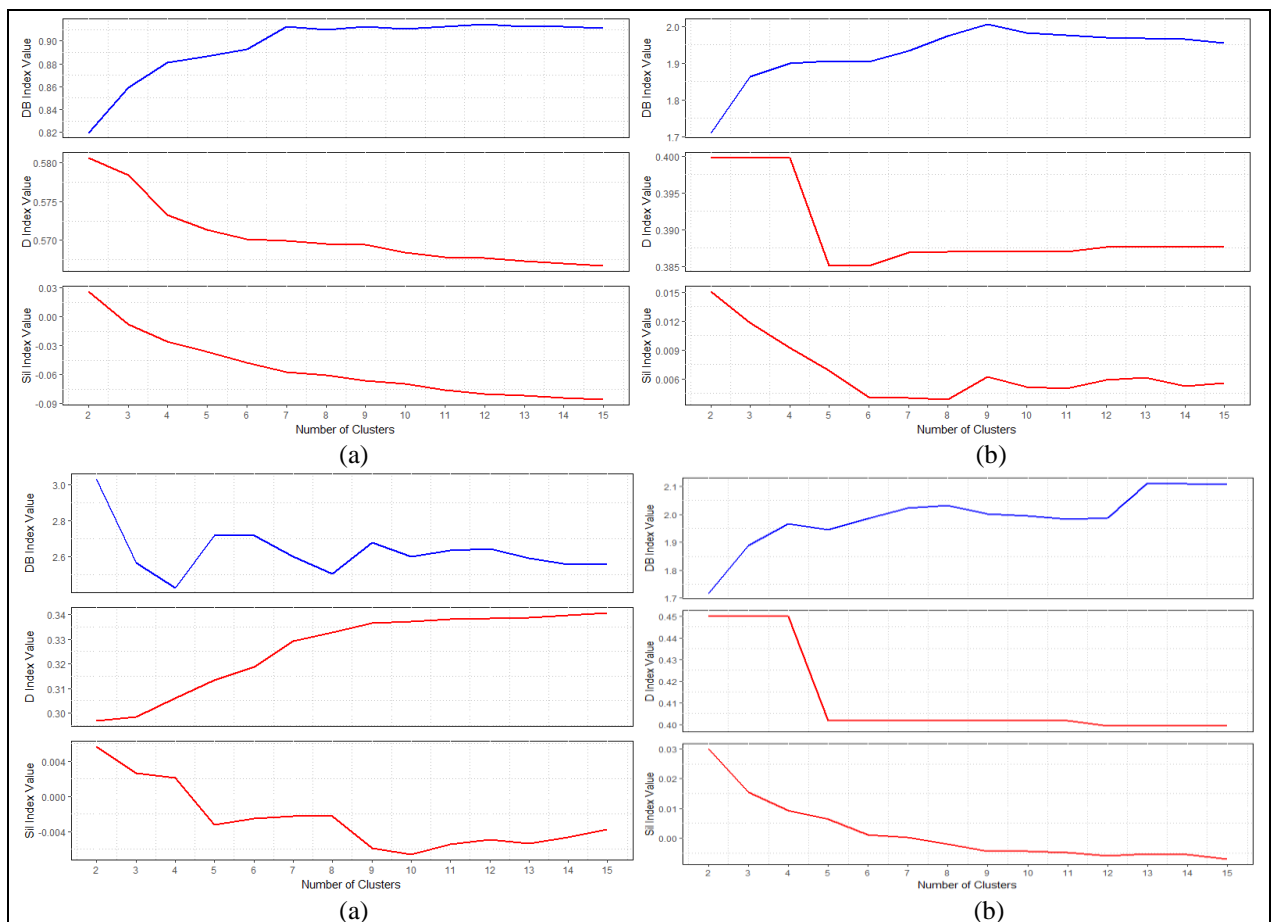
Hierarchinių metodų optimalaus klasterių skaičiaus parinkimo tyrimo rezultatai pateikti 3.9 paveikslo grafikuose. Čia (a), (b) ir (c) atvejais klasterių centrai skaičiuojami dalijimo aplink „medoidus“ metodu – klasterių centras yra viena laiko eilutė iš duomenų rinkinio.

Kai klasterizavime hierarchiniu metodu naudojama Euklido atstumo metrika, jungimo metodas – artimiausio kaimyno, gauti netinkami klasteriai, kadangi daugumą jų sudaro po vieną laiko eilutę.

Iš 3.9 pav. (b) dalies, kai naudojama Euklido atstumo metrika, jungimo metodas – vidutinės jungties, gautas optimalus klasterių skaičius lygus pagal $DB_{index} - 2$, pagal $D_{index} - 4$, pagal $Sil_{index} - 2$. Šiuo atveju parenkame optimalų skaičių lygų 2, $Sil_{index} = 0.015$, $D_{index} = 0.4$, $DB_{index} = 1.711$, sudaryti klasteriai, kur viename 3 085 akcijos, kitame 117.

Iš 3.9 pav. (c) dalies, kai naudojama Euklido atstumo metrika, jungimo metodas – tolimiausio kaimyno, gauta optimalus klasterių skaičius lygus pagal $DB_{index} - 4$, pagal $D_{index} - 9$ ir daugiau, pagal $Sil_{index} - 2$. Šiuo atveju parenkame optimalų skaičių lygų 2, $Sil_{index} = 0.006$, $D_{index} = 0.297$, $DB_{index} = 3.03$, sudaryti klasteriai, kur viename 1 175 akcijos, kitame 2 027.

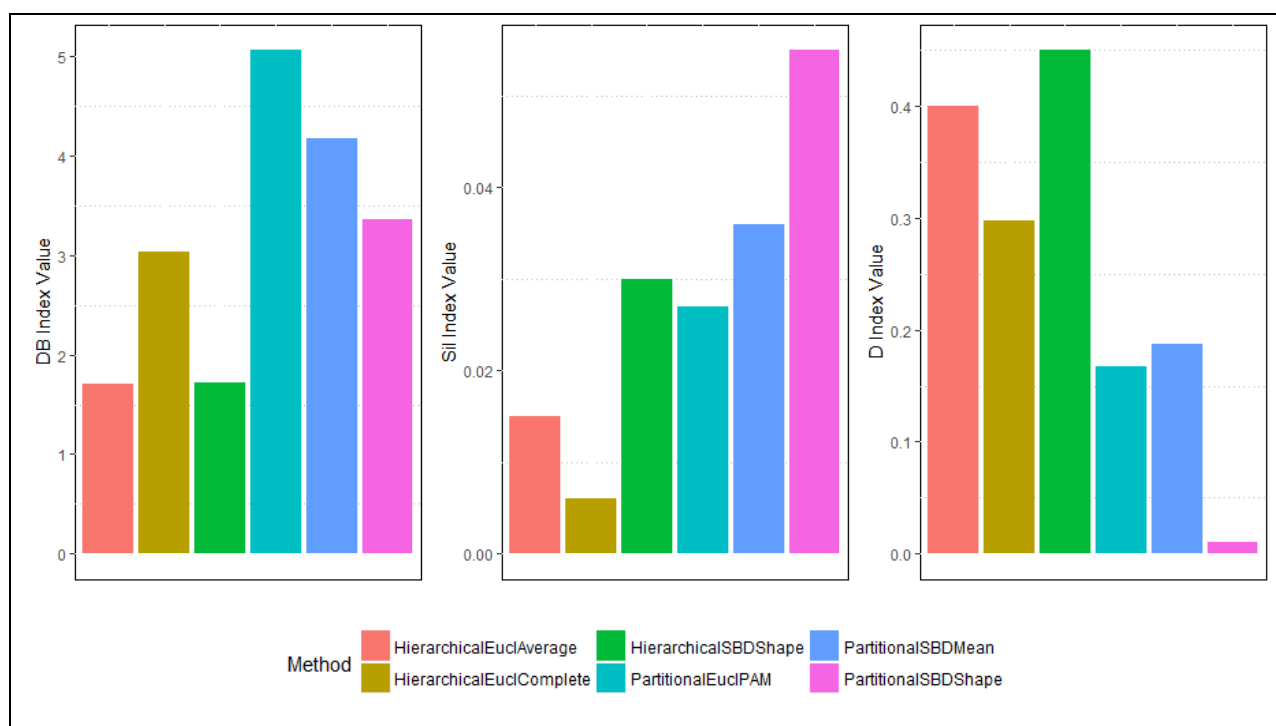
Iš 3.9 pav. (d) dalies, kai naudojama eilutės forma paremta atstumo metrika, metodas eilutės formos vidurkis, gauta optimalus klasterių skaičius lygus pagal $DB_{index} - 2$, pagal $D_{index} - 4$ ir daugiau, pagal $Sil_{index} - 2$. Šiuo atveju parenkame optimalų skaičių lygų 2, $Sil_{index} = 0.03$, $D_{index} = 0.45$, $DB_{index} = 1.718$, sudaryti klasteriai, kur viename 3 190 akcijos, kitame 12.



3.9. pav. Silhouette, Davies-Bouldin ir Dunn indeksų priklausomybė nuo klasterių skaičiaus. (a) dalyje klasterizavime hierarchiniu metodu naudojama Euklido atstumo metrika, metodas – artimiausio kaimyno; (b) – Euklido atstumo metrika, metodas – vidutinės jungties; (c) – Euklido atstumo metrika, metodas – tolimiausio kaimyno; (d) – eilutės forma paremta atstumo metrika, metodas – k -formos

3.2.3. Klasterizavimo metodų palyginimas

Palyginime visus taikytus klasterizavimo metodus su gautu optimaliu klasterių skaičiumi. Palyginimas *Silhouette*, *Davies-Bouldin* ir *Dunn* indeksų atžvilgiu pateiktas 3.10 paveikslo diagramose. Čia *Davies-Bouldin* metrikos atžvilgiu geriausiai mūsų akcijų duomenims klasterizuoti tinka klasterizavimas į du klasterius hierarchiniu metodu naudojant Euklido atstumo metriką, vidutinės jungties metodą, *Dunn* metrikos atžvilgiu geriausiai tinka klasterizavimas į du klasterius hierarchiniu metodu naudojant eilutės forma paremtą atstumo metriką, *k*-formos jungties metodą, *Silhouette* metrikos atžvilgiu – klasterizuojant į 14 grupių padalijimo metodu forma paremta atstumo metrika (SBD), centrų skaičiavimo funkcija – *k*-forma. Parenkame pagal *Silhouette* geriausiai tinkantį klasterizavimą į 14 klasterių, taip pat logine prasme išskaidyti 3 202 akcijų gražos laiko eilutes tik į du klasterius būtų nenaudinga.



3.10. pav. *Silhouette*, *Davies-Bouldin* ir *Dunn* indeksų reikšmių palyginimas skirtingais klasterizavimo metodais.

3.2.4. Sudarytų klasterių aprašymas

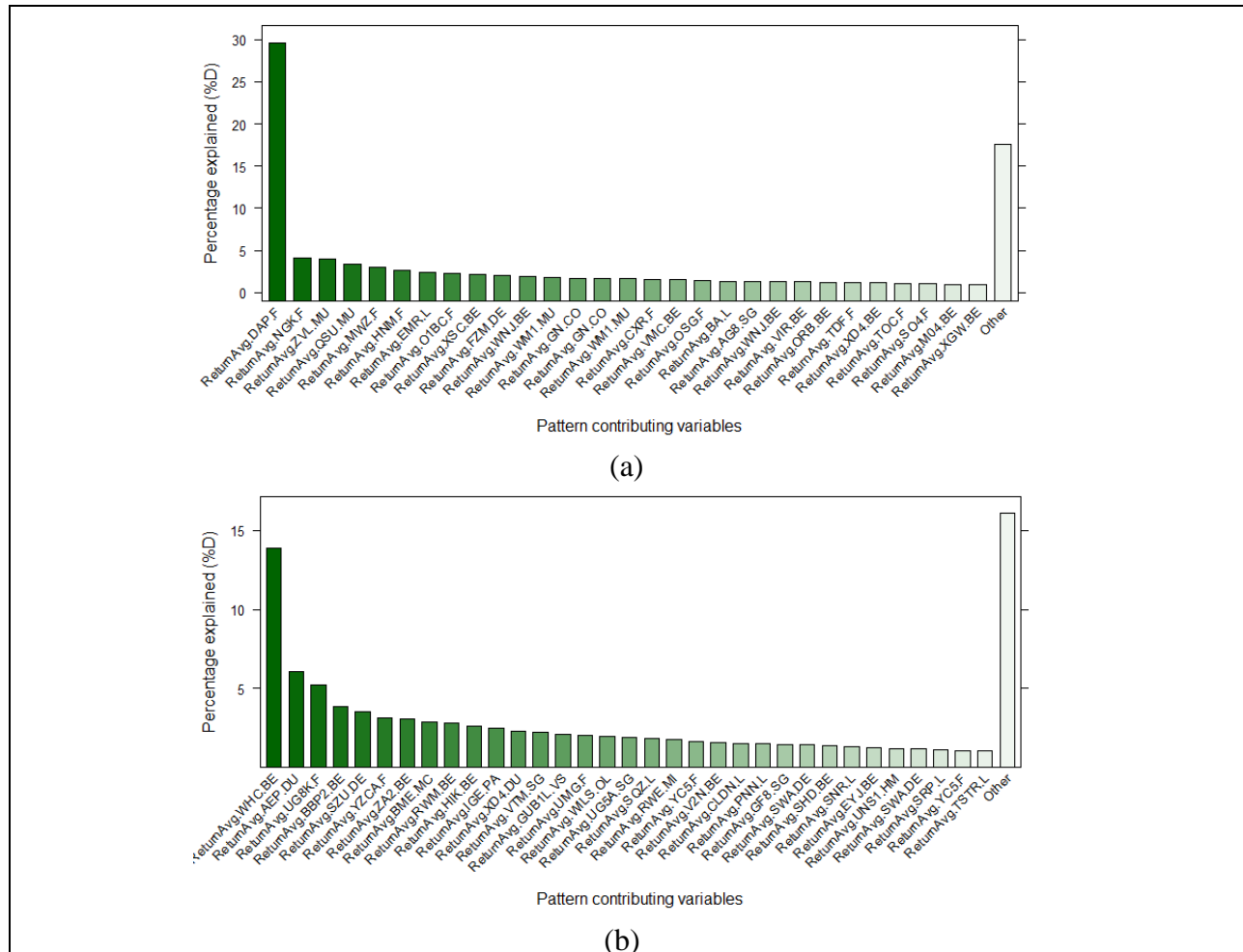
Gautus 14 klasterių aprašant randame skiriamuosius požymius pagal jų aprašomąją statistiką ir skiriamuosius požymius pagal šalis, ši detali informacija pateikta 4 priedo lentelėje:

- 1 klasteris – sudaro 190 akcijų, skirtingų šalių 9, maksimali vidutinė teigiama grąža, dominuojanti šalis Šveicarija (10 % visų duomenų šios šalies duomenų yra šiame klasteryje);
- 2 klasteris – sudaro 168 akcijos, skirtingų šalių 10, dominuojanti šalis Austrija (13% visų duomenų);
- 3 klasteris – sudaro 143 akcijos, skirtingų šalių 12, dominuojančios šalys Lietuva (17%) ir Švedija (10%);
- 4 klasteris – sudaro 120 akcijų, skirtingų šalių 10, dominuojančios šalys Belgija (13%), Nyderlandai (11%) ir Švedija (10%);
- 5 klasteris – sudaro 176 akcijos, skirtingų šalių 12, minimali vidutinė teigiama grąža, dominuojančios šalys Latvija (17%) ir Portugalija (17%);
- 6 klasteris – sudaro 163 akcijos, skirtingų šalių 15, dominuojančios šalys Latvija (17%) ir Portugalija(17%);
- 7 klasteris – sudaro 223 akcijos, skirtingų šalių 15, dominuojančios šalys Ispanija (30%) Prancūzija (28%) ir Italija (26%);
- 8 klasteris – sudaro 186 akcijos, skirtingų šalių 12, dominuojančios šalys Belgija (20%) ir Suomija (14%);
- 9 klasteris – sudaro 568 akcijos, skirtingų šalių 14, vidutinė neigiama grąža (nuostolingos akcijos), dominuojančios šalys Vokietija (19%), Danija (19%) ir Portugalija (16%);
- 10 klasteris – sudaro 137 akcijos, skirtingų šalių 10, dominuojančios šalys Latvija (17%) ir Suomija (14%);
- 11 klasteris – sudaro 120 akcijų, skirtingų šalių 11, dominuojanti šalis Nyderlandai (10%) ir Norvegija (35%);
- 12 klasteris – sudaro 648 akcijos, skirtingų šalių 16, maksimali vidutinė teigiama grąža, dominuojančios šalys Latvija (50%) ir Lietuva (47%);
- 13 klasteris – sudaro 151 akcijos, skirtingų šalių 10, minimali vidutinė teigiama grąža, dominuojanti šalis Portugalija (17%);
- 14 klasteris – sudaro 209 akcijos, skirtingų šalių 11, minimali vidutinė teigiama grąža, dominuojanti šalis Ispanija (11%);

Taip pat akcijų grąžų kitimas kiekviename klasteryje kas mėnesį pateiktas 4 priede.

3.3. Ypatingųjų reikšmių dekompozicija

Anksčiau gautuose 14 klasterių randame dominuojančias akcijų gražos laiko eilutes ypatingųjų reikšmių dekompozicijos metodu, pagal kurias toliau sudarysime akcijų gražos prognozės modelius.



3.11. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaiškinama dispersijos procentinė dalis; (a) dalyje – 1-ojo klasterio; (b) dalyje – 2-ojo klasterio

Kaip matome iš 3.11 pav. (a) dalies 1-ame klasteryje net apie 30% pagrindinių komponentių dispersijos paaiškina ir prognozės modeliui sudaryti parenkame „Danaher“ įmonės akcijas Frankfurto biržoje (akcijų simbolis DAP.F), iš (b) dalies – 2-ame klasteryje apie 14% įmonės „Weyerhaeuser Reit“ akcijas Berlyno biržoje (WHC.BE).

Tokiu pačiu principu išskiriame ir kitų klasterių dominuojančias akcijų laiko eilutes (rezultatai pateikti 5 darbo priede). Iš jų gauta kad, 3-ame klasteryje net apie 12% pagrindinių komponentių dispersijos paaiškina ir prognozės modeliui sudaryti parenkame „T&D Holdings“ įmonės akcijas Miuncheno biržoje (akcijų simbolis TZ6.MU), 4-ame klasteryje net 28% įmonės „M&G High Income Geared Unit“ akcijas Londono biržoje (MGHU.L), 5-ame klasteryje apie 14% dispersijos paaiškina ir prognozės modeliui sudaryti parenkame „Goldman Sachs Group“ įmonės akcijas Vokietijos Diuseldorfo biržoje (akcijų simbolis GOS.DU), 6-ame klasteryje apie

12% įmonės „Mondi -NPV-“ akcijos Vokietijos Štutgarto biržoje (BTKC.SG), 7-ame klasteryje net apie 34% „Schroders plc“ įmonės akcijos Londono biržoje (akcijų simbolis SDR.L), 8-ame klasteryje apie 11% įmonės „TDK Corp“ akcijos Vokietijos Diuseldorfo biržoje (TDK.DU), 9-ame klasteryje apie 11% „Deutsche Post N“ įmonės akcijos Frankfurto biržoje (akcijų simbolis DPW.F), 10-ame klasteryje apie 15% įmonės „Mondi -NPV-“ akcijos Frankfurto biržoje (BTKC.F), 11-ame klasteryje apie 19% „Bayerische Motoren Werke Aktiengesellschaft“ įmonės akcijos Vokietijos XETRA biržoje (akcijų simbolis BMW3.DE), 12-ame klasteryje apie 7% įmonės „SAP SE“ akcijas Vokietijos XETRA biržoje (SAP.DE), 13-ame klasteryje apie 29% „Petrochina“ įmonės akcijos Berlyno biržoje (akcijų simbolis PC6A.BE), 14-ame klasteryje apie 17% pagrindinių komponentų dispersijos paaiškina ir prognozės modeliui sudaryti parenkame įmonės „Deutsche Post AG“ akcijas Vokietijos XETRA biržoje (DPW.DE).

3.4. Akcijų grąžos prognozių modelių sudarymas ir palyginimas

Sudarome prognozių modelius akcijų grąžai kiekvienai laiko eilutei atskirai. Modelio parametrus parenkame pritaikydami modelį kiekviename klasteryje išskirtai reikšmingiausiai laiko eilutei ir gautą modelį taikome kiekvienos įmonės vidutinei mėnesinei akcijų grąžai prognozuoti atskirai, t. y. sudarome 3 202 modelius su 15 nepriklausomų kintamųjų:

$$\hat{Y}_t = f(Y_{t-1}, X_{t-1}^{(1)}, X_{t-1}^{(2)}, X_{t-1}^{(3)}, X_{t-1}^{(4)}, X_{t-1}^{(5)}, X_{t-1}^{(6)}, X_{t-1}^{(7)}, X_{t-1}^{(8)}, X_{t-1}^{(9)}, X_{t-1}^{(10)}, X_{t-1}^{(11)}, X_{t-1}^{(12)}, X_{t-1}^{(13)}, X_{t-1}^{(14)}, X_{t-1}^{(15)}) + \varepsilon_t; \quad (23)$$

čia \hat{Y}_t vidutinės logaritminė mėnesinė akcijų grąžos prognozė laiko momentu t , Y_{t-1} akcijų grąžos reikšmė prieš tai buvusiu laiko momentu, f – funkcinė priklausomybė, ε – atsitiktinė paklaida, $X_{t-1}^{(1)}$ – įmonės pavadinimo *Google* tendencija prieš tai buvusiu laiko momentu, $X_{t-1}^{(2)}$ – termino *chance Google* tendencija, $X_{t-1}^{(3)}$ – termino *conflict Google* tendencija, $X_{t-1}^{(4)}$ – termino *debt Google* tendencija, $X_{t-1}^{(5)}$ – termino *earnings Google* tendencija, $X_{t-1}^{(6)}$ – termino *leverage Google* tendencija, $X_{t-1}^{(7)}$ – termino *inflation Google* tendencija, $X_{t-1}^{(8)}$ – termino *money Google* tendencija, $X_{t-1}^{(9)}$ – termino *Nasdaq Google* tendencija, $X_{t-1}^{(10)}$ – vartotojų kainų indeksas, $X_{t-1}^{(11)}$ – pramonės produkcijos augimas, $X_{t-1}^{(12)}$ – infliacija, $X_{t-1}^{(13)}$ – trumpalaikė palūkanų norma, $X_{t-1}^{(14)}$ – akcijų kainų indeksas, $X_{t-1}^{(15)}$ – nedarbo lygis.

Kiekvieną priklausomo ir nepriklausomo kintamojo laiko eilutę išskiriame į apmokymo ir testavimo imtis tokiais principu: apmokome su 46 mėnesių nuo 2012 02 01 iki 2015 12 31 duomenis (viso 75 % procentai visos imties), testuojame su 15 mėnesių nuo 2016 01 01 iki 2017 03 31 duomenis (viso 25 % procentai visos imties).

Kad kintamųjų matavimo skalė būtų suvienodinta standartizuojame atskirai visų eilučių reikšmes centruodami ir pakeisdami jų mastelį pagal šią formulę:

$$x_{stand_t} = \frac{x_t - \mu}{\sigma} \quad (24)$$

čia x_{stand_t} – standartizuota kintamojo x reikšmė laiko momentu t , x_t – kintamojo x reikšmė laiko momentu t , t. y. laiko eilutė $x = (x_1, \dots, x_n)$, μ – kintamojo x vidurkis, σ – kintamojo x standartinis nuokrypis.

3.4.1. 1-ojo klasterio akcijų prognozės modeliai

Regresijos modelis. Sudarome apibendrintą tiesinį modelį, kai ryšio funkcija tiesinė, o priklausomo kintamojo skirstinys normalusis. Modelį pritaikome „Danaher“ įmonės akcijos duomenims. Modelį sudarant naudojama žingsninė regresija, kai „žingsniuojama atgal“, t. y. iš modelio pašalinami vis kiti statistiškai nereikšmingi regresoriai (angl. *backward stepwise regression*). Procesas baigiamas, kai įtraukti lieka tik statistiškai reikšmingi regresoriai. Gauti rezultatai 1-o klasterio reikšmingiausiai laiko eilutei pateikti 3.6 lentelėje.

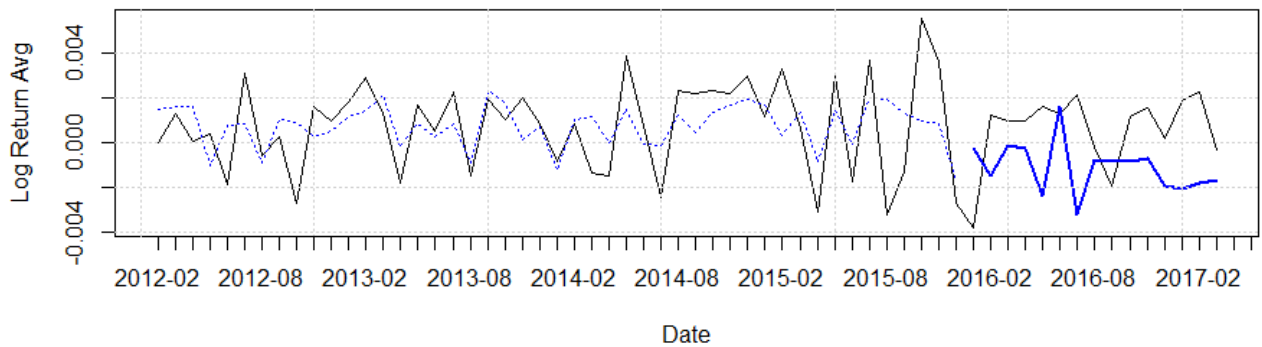
3.6 lentelė. 1-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.5103	0.1114	4.58	4.11E-05	***
debt_hits_lag1	0.2967	0.1998	1.485	0.145	
earnings_hits_lag1	-0.3576	0.1529	-2.338	0.0242	*
inflation_hits_lag1	0.2265	0.1415	1.601	0.1169	
IPG_PP_lag1	-0.2528	0.1583	-1.597	0.1178	

Iš rezultatų lentelės gauname, kad hipotezė apie regresijos lygties koeficientų reikšmingumą ($H_0: \beta_i = 0$, $H_a: \beta_i$ reikšmingas) su 95 % reikšmingumo lygmeniu atmesta atkarpos koeficientui ir termino *earnings* tendencijos koeficientui. Gautos modelio kokybės vertinimo metrikos pateiktos 3.7 lentelėje ir akcijų gražos kitimas laike lyginant su gautos prognozės kitimu pateiktas 3.12 paveiksle. Čia matome, kad paklaida testavimo imčiai gauta yra gana didelė lyginant su duomenų masteliu, taip pat nei pusė akcijų gražos kitimo krypčių ir gražos kitimo nuostolio ar pelno atžvilgiu nėra suprognozuota teisingai.

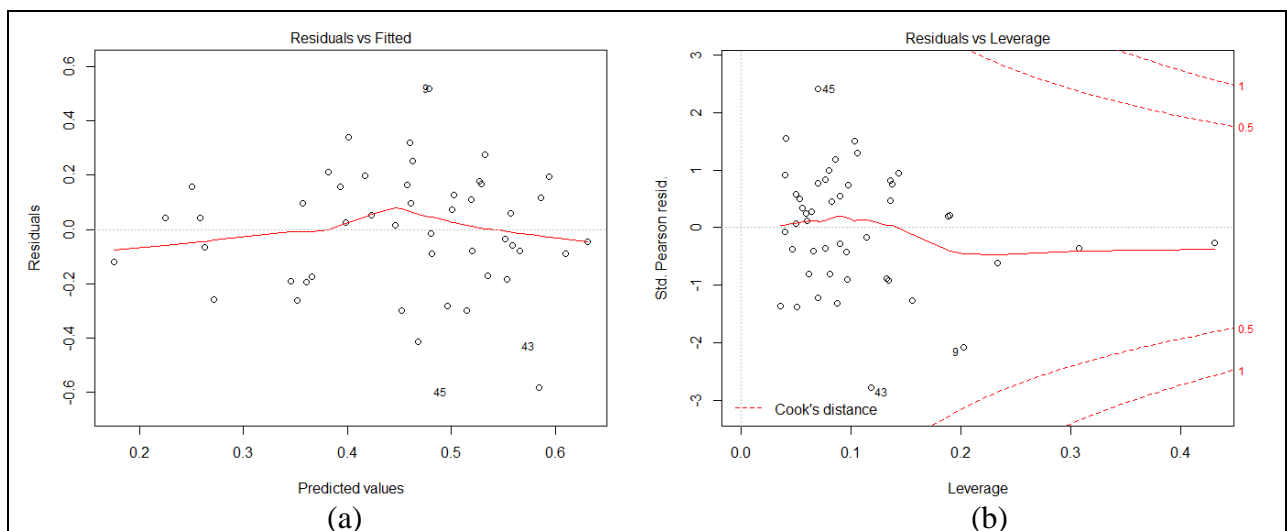
3.7 lentelė. 1-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-1.79E-18	0.00185	57.14	73.33
Testavimo imtis	0.00172	0.00279	35.71	33.33



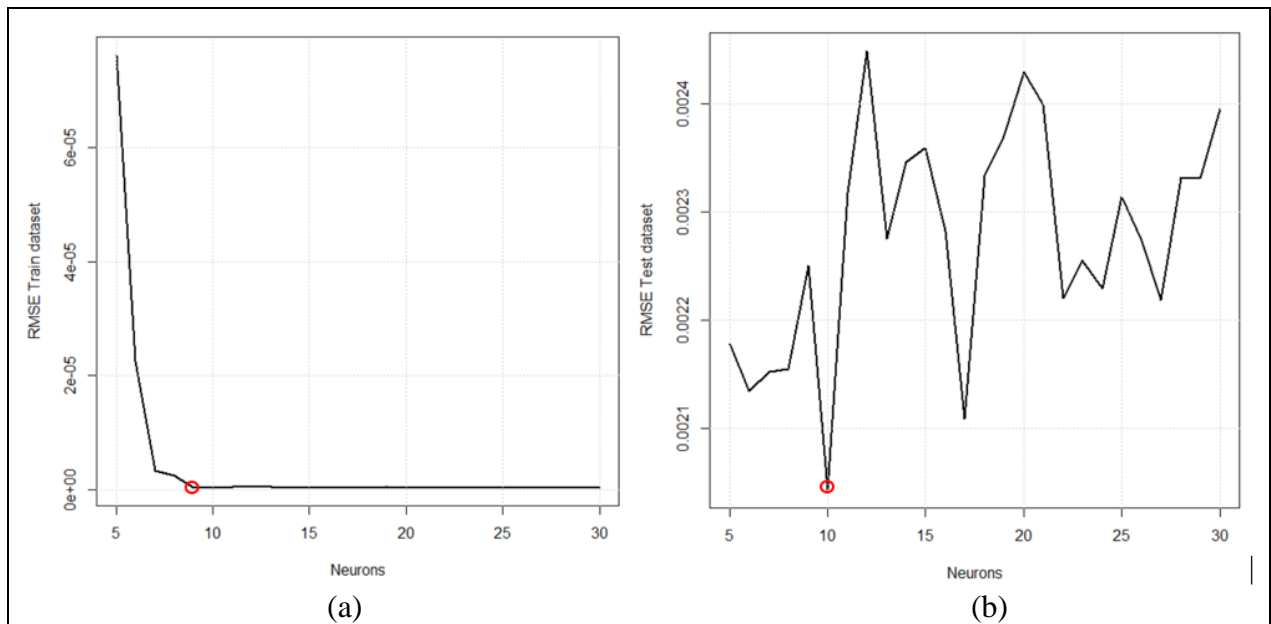
3.12. pav. 1-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliui (mėlyna punktyrinė – apmokymo imtis, mėlyna ištisinė – testavimo imtis).

Patikriname sudaryto modelio prielaidas atsitiktinėms paklaidoms ε_t . Patikrinimui ar atsitiktinės paklaidos ε_t yra normaliai pasiskirstę atsitiktiniai dydžiai, atliekame *Shapiro-Wilk* normalumo testą, kur tikrinama hipotezė: $H_0: \varepsilon_t \sim N(0, \sigma)$, $H_a: \varepsilon_t$ nėra normalusis dydis. Gauta p-reikšmė lygi 0.9331 (statistika $W = 0.98898$), tad su 95 % reikšmingumo lygmeniu hipotezė, kad $\varepsilon_t \sim N(0, \sigma)$ neatmetama – prielaida tenkinama. Norint nustatyti, ar skirtingų stebėjimų liekamosios paklaidos koreliuoja (yra autokoreliacija), randame *Durbin-Watson* statistiką, jei ji yra tarp 1.5 ir 2.5, tai tariama, kad auto-koreliacijos nėra. Gauta $DW = 2.22$, tad ši prielaida tenkinama. Liekanų homoskedastiškumui patikrinti naudojame prognozuotų reikšmių ir standartizuotų liekanų sklaidos diagramą, kuri pateikta 3.13 (a) dalies grafike. Čia matome, kad homoskedastiškumo prielaida tenkinama (reikšmių priklausomybė tiesinė), taip pat matome, kad vidurkis artimas nuliui. Iš (b) dalies, matome kad Kuko mato atžvilgiu išskirčių modelio liekanose nėra.



3.13. pav. Prielaidų tikrinimo rezultatai. (a) dalis prognozių reikšmių priklausomybė nuo standartizuotų modelio liekanų; (b) – liekanų ir išskirčių sveto, bei Kuko mato priklausomybės

Neuroninių tinklų modelis. Taikome netiesinį autoregresinį su papildomais įėjimais (išoriniai kintamaisiais) neuronų tinklą. Neuroninis tinklas sudaromas 50 kartų su skirtingais atsitiktiniais pradiniais svoriais. Kaip ir 1.1.4 poskyryje apžvelgtuose tyrimuose naudojamas laiko vėlinimas vienu periodu. Atliekame eksperimentinį optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimą, t. y. sudarome 25 neuroninius tinklus kintant neuronų skaičiui nuo 5 iki 30. Gauti rezultatai vertinami *RMSE* paklaidos atžvilgiu ir pateikti 3.12 grafikuose atskirai apmokymo imčiai ir testavimo imčiai. Iš šių rezultatų (a) dalies matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 9 ($RMSE = 2.90E-07$), o testavimo imčiai lygus 10 ($RMSE = 0.002$). Parenkame optimalų neuronų skaičių pagal testavimo imties rezultatus.



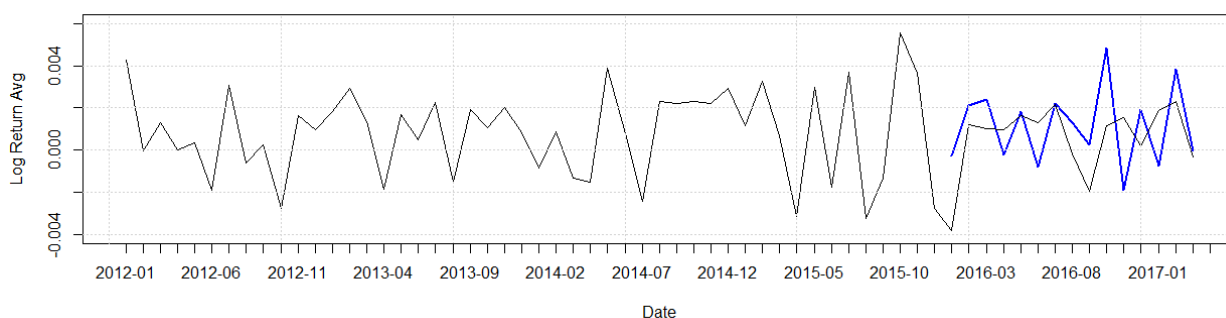
3.14. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

Taigi sudarytas neuroninis tinklas, kurio struktūra 16-10-1, t. y. 16 įėjimo neuronų, 10 paslėptojo sluoksnio neuronai ir 1 išėjimo neuronas, viso apskaičiuota 181 neuronų jungčių svoriai.

Gautos modelio kokybės vertinimo metrikos pateiktos 3.8 lentelėje ir akcijų gražos kitimas laike lyginant su gautos prognozės kitimu pateiktas 3.15 paveiksle. Čia matome, kad paklaida testavimo imčiai gauta kaip ir regresijos modelio atveju yra gana didelė lyginant su duomenų masteliu, tačiau daugiau nei pusė testavimo duomenų akcijų gražos kitimo kryptį ir kitimo nuostolio ar pelno atžvilgiu suprognozuota teisingai.

3.8 lentelė. 1-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-3.01E-08	3.011E-07	100.00	100.00
Testavimo imtis	-5.08E-04	2.09E-03	71.42	60.00



3.15. pav. 1-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

Sudarytą dominuojančiai 1-ojo klasterio laiko eilutei regresijos ir NARX modelių pritaikome visoms 1-ojo klasterio akcijų grąžos laiko eilutėms (viso 190 įmonių). Gautus sudarytų modelių ir prognozės rezultatus apibendriname apskaičiuojant jų kokybės vertinimo metrikų vidurkius, išrinkus minimalias ir maksimalias reikšmes. Šie apibendrinti modelių rezultatai prognozuojant 1-ojo klasterio akcijų grąžas pateikti 3.9 lentelėje.

3.9 lentelė. Apibendrinti visų 1-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-6.239e-18	0.001326	39.13	46.81
		Vidurkis	-4.030e-19	0.003411	60.41	64.07
		Maksimumas	9.916e-18	0.009313	78.26	82.98
	Testavimo imtis	Minimumas	-0.0193504	0.001301	7.143	20.00
		Vidurkis	0.0002531	0.004712	46.391	49.61
		Maksimumas	0.0145187	0.026324	92.308	86.67
NARX	Apmokymo imtis	Minimumas	-5.045e-07	2.538e-07	69.57	80.85
		Vidurkis	2.710e-08	3.202e-06	99.82	99.70
		Maksimumas	1.790e-06	3.467e-05	100.00	100.00
	Testavimo imtis	Minimumas	-0.0135054	0.001810	14.29	20.00
		Vidurkis	0.0005672	0.005063	56.95	54.56
		Maksimumas	0.0110769	0.019839	100.00	80.00

Lyginant regresijos modelio ir autoregresinio neuroninio tinklo prognozės rezultatus testavimo imčiai matome, kad *RMSE* metrikos dydis šiek tiek mažesnis regresijos modelio atveju, tačiau *POCID* ir *POS* metrikų atžvilgiu geresnis neuroninių tinklų modelis. Tad kadangi pastarasis modelis geriau prognozuoja akcijų grąžos kryptį ir grąžos teigiamumą ar neigiamumą, o tai yra labai naudinga informacija investuotojams, galime teigti, kad pirmojo klasterio akcijoms prognozuoti šiek tiek geriau tinka NARX modelis.

3.4.2. Kitų klasterių akcijų prognozės modeliai

Tokiu pačiu principu sudaromi ir kitų klasterių akcijų prognozės modeliai, gauti rezultatai pateikti 6 – 18 prieduose.

2-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 6 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atkarpos koeficientas, termino *money* tendencijos, infliacijos, pramonės produkcijos augimo, trumpalaikės palūkanų normos rodiklių. Regresijos modelio kokybės vertinimo metrikos yra prastos – tik apie 30 % akcijos gražos kitimo kryptį ir kitimo nuostolio ar pelno atžvilgiu suprognozuota teisingai. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 10 ($RMSE = 3.5e-07$), o testavimo imčiai lygus 9 ($RMSE = 0.00402$). Parinkus optimalų neuronų skaičių lygų 11 (testavimo imties $RMSE = 0.00403$, apmokymo imties $RMSE = 4.12e-07$) gautas NARX modelis $RMSE$, POS , $POCID$ metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 2-ojo klasterio akcijų gražos laiko eilutėms (viso 168 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.10 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 2-ojo klasterio akcijų gražoms prognozuoti.

3.10 lentelė. Apibendrinti visų 2-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-9.468e-18	0.0008172	45.65	46.81
		Vidurkis	3.926e-19	0.0043300	62.44	63.63
		Maksimumas	3.541e-17	0.0281154	82.61	82.98
	Testavimo imtis	Minimumas	-4.423e-02	0	0	13.33
		Vidurkis	-2.051e-04	0.005705	48.81	50.39
		Maksimumas	1.242e-02	0.055671	85.71	100.00
NARX	Apmokymo imtis	Minimumas	-3.561e-07	1.713e-07	65.22	72.34
		Vidurkis	1.590e-08	8.811e-07	99.20	99.01
		Maksimumas	6.376e-07	7.045e-06	100.00	100.00
	Testavimo imtis	Minimumas	-2.105e-02	0.001883	0	20.00
		Vidurkis	5.846e-04	0.006627	52.17	52.26
		Maksimumas	1.866e-02	0.044528	92.86	100.00

3-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 7 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra infliacijos, akcijų kainų indekso, nedarbo lygio rodiklių. Regresijos modelio prognozės $RMSE$ kaip ir kitų klasterių atveju gana didelė, tačiau apie 50 % akcijos gražos kitimo kryptį ir kitimo nuostolio ar pelno atžvilgiu suprognozuota teisingai. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 5.42e-07$), o testavimo imčiai lygus 6

($RMSE = 0.00542$). Parinkus optimalų neuronų skaičių lygų 8 gautas NARX modelis visų metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 3-ojo klasterio akcijų gražos laiko eilutėms (viso 143 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.11 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 3-ojo klasterio akcijų gražoms prognozuoti.

3.11 lentelė. Apibendrinti visų 3-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-3.179e-16	0.0006784	19.57	42.55
		Vidurkis	-2.710e-18	0.0049857	59.08	61.30
		Maksimumas	1.433e-17	0.0741863	80.43	80.43
	Testavimo imtis	Minimumas	-0.0293545	0.0007143	14.29	26.67
		Vidurkis	0.0002115	0.0060103	49.09	49.91
		Maksimumas	0.0074046	0.0661420	85.71	80.00
NARX	Apmokymo imtis	Minimumas	-1.210e-06	3.782e-07	32.61	65.96
		Vidurkis	-7.014e-09	5.402e-06	98.86	98.42
		Maksimumas	1.212e-06	5.496e-05	100.00	100.00
	Testavimo imtis	Minimumas	-1.888e-02	0.0009905	7.143	20.00
		Vidurkis	-6.950e-07	0.0069635	52.298	53.47
		Maksimumas	1.711e-02	0.0695301	85.714	86.67

4-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 8 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra terminų *debt*, *infaltion*, *money*, *Nasdaq* tendencijos, vartotojų kainų indekso, pramonės produkcijos augimo, akcijų kainų indekso, nedarbo lygio rodikliai ir praėjusio mėnesio akcijų kainos graža. Regresijos modelio prognozės $RMSE$ kaip ir kitų klasterių atveju gana didelė, tačiau net apie 70 % akcijos gražos kitimo kryptį ir kitimo nuostolio ar pelno atžvilgiu suprognozuota teisingai. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 7 ($RMSE = 5.97e-07$), o testavimo imčiai lygus 7 ($RMSE = 0.00402$). Parinkus optimalų neuronų skaičių lygų 7 gautas NARX modelis visų metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 4-ojo klasterio akcijų gražos laiko eilutėms (viso 120 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.12 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, tačiau galime teigti kad visų metrikų atžvilgiu 4-ojo klasterio akcijų gražoms prognozuoti geresnis NARX modelis.

3.12 lentelė. Apibendrinti visų 4-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-8.85E-18	0.0008007	50	52.17
		Vidurkis	5.87E-20	0.0026166	69.4	69.51
		Maksimumas	7.28E-18	0.0098089	86.96	82.98
	Testavimo imtis	Minimumas	-0.00795	0.001595	14.29	20
		Vidurkis	0.0008213	0.004786	48.61	46.21
		Maksimumas	0.0101198	0.017097	92.86	73.33
NARX	Apmokymo imtis	Minimumas	-2.89E-06	2.98E-07	86.96	85.11
		Vidurkis	-7.88E-09	9.11E-06	99.82	99.65
		Maksimumas	1.11E-06	9.09E-05	100	100
	Testavimo imtis	Minimumas	-0.0128365	0.001714	14.29	26.67
		Vidurkis	-0.0003231	0.004171	54.11	53.28
		Maksimumas	0.0076694	0.015804	85.71	80

5-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 9 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atkarpos koeficiento ir infliacijos rodiklio. Regresijos modelio prognozės *RMSE* kaip ir kitų klasterių atveju gana didelė, virš 50 % akcijos gražos kitimo krypčių ir kitimo nuostolio ar pelno atžvilgiu suprognuota teisingai. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 7 ($RMSE = 5.933579e-06$), o testavimo imčiai lygus 12 ($RMSE = 0.004603$). Parinkus optimalų neuronų skaičių lygų 8 gautas NARX modelis visų metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 5-ojo klasterio akcijų gražos laiko eilutėms (viso 174 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.13 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 5-ojo klasterio akcijų gražoms prognozuoti.

3.13 lentelė. Apibendrinti visų 5-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-2.53E-17	0.0006718	30.43	36.17
		Vidurkis	-3.72E-19	0.0045487	58.03	60.16
		Maksimumas	2.36E-17	0.0223703	76.09	76.6
	Testavimo imtis	Minimumas	-0.0324668	0.00154	14.29	13.33
		Vidurkis	-0.0004296	0.005487	48.86	51.54
		Maksimumas	0.0210404	0.041997	92.86	80
NARX	Apmokymo imtis	Minimumas	-2.27E-06	2.67E-07	45.65	70.21
		Vidurkis	-2.15E-08	4.65E-06	99.36	99.01
		Maksimumas	1.92E-06	7.49E-05	100	100
	Testavimo imtis	Minimumas	-0.0139832	0.001014	14.29	13.33
		Vidurkis	0.0008306	0.006491	52.96	49.85
		Maksimumas	0.0149958	0.029936	92.86	80

6-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 10 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra *debt* ir *inflation* tendencijos, vartotojų kainų indekso, pramonės produkcijos augimo rodiklių. Regresijos modelio prognozė visų metrikų atžvilgių labai prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 5.629e-07$), o testavimo imčiai lygus 10 ($RMSE = 0.003302$). Parinkus optimalų neuronų skaičių lygų 10 gautas NARX modelis visų metrikų atžvilgiu daug geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 6-ojo klasterio akcijų gražos laiko eilutėms (viso 163 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.14 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 6-ojo klasterio akcijų gražoms prognozuoti.

3.14 lentelė. Apibendrinti visų 6-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-5.85E-17	0.001049	34.78	46.81
		Vidurkis	-1.62E-18	0.005675	63.29	65.8
		Maksimumas	5.88E-17	0.071191	84.78	80.85
	Testavimo imtis	Minimumas	-4.20E-02	0.0006736	7.692	13.33
		Vidurkis	-1.63E-04	0.0074075	49.164	51.38
		Maksimumas	2.29E-02	0.0769622	85.714	86.67
NARX	Apmokymo imtis	Minimumas	-8.38E-07	1.86E-07	58.7	74.47
		Vidurkis	1.58E-08	1.71E-06	98.6	98.21
		Maksimumas	2.60E-06	2.87E-05	100	100
	Testavimo imtis	Minimumas	-0.0200994	0.001001	14.29	20
		Vidurkis	0.0007806	0.008424	53.51	51
		Maksimumas	0.05783	0.089531	92.86	86.67

7-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 11 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atkarpos koeficiento ir termino *earnings* tendencijos. Regresijos modelio prognozė $RMSE$ atžvilgiu šiek tiek geresnė nei kitų analizuotų akcijų atveju. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 3.954e-07$), o testavimo imčiai lygus 11 ($RMSE = 0.00285$). Parinkus optimalų neuronų skaičių lygų 11 gautas NARX modelis visų metrikų atžvilgiu šiek tiek geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 7-ojo klasterio akcijų gražos laiko eilutėms (viso 223 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.15 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 7-ojo klasterio akcijų gražoms prognozuoti.

3.15 lentelė. Apibendrinti visų 7-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-8.21E-18	0.0009493	39.13	42.55
		Vidurkis	-4.98E-19	0.0032379	56.84	60.77
		Maksimumas	6.13E-18	0.008662	76.09	78.72
	Testavimo imtis	Minimumas	-0.0052726	0.001043	7.692	26.67
		Vidurkis	0.0006894	0.004107	51.212	50.63
		Maksimumas	0.0115197	0.01314	84.615	78.57
NARX	Apmokymo imtis	Minimumas	-3.56E-07	2.04E-07	86.96	93.62
		Vidurkis	2.19E-09	6.47E-07	99.91	99.71
		Maksimumas	2.06E-07	2.04E-06	100	100
	Testavimo imtis	Minimumas	-7.34E-03	0.001256	14.29	20
		Vidurkis	9.39E-05	0.004704	53.94	51.66
		Maksimumas	9.93E-03	0.014811	92.86	80

8-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 12 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra termino *inflation* tendencijos, infliacijos, trumpalaikės palūkanų normos, nedarbo lygio rodiklių. Regresijos modelio prognozė visų metrikų atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 8.242e-07$), o testavimo imčiai lygus 8 ($RMSE = 0.00414$). Parinkus optimalų neuronų skaičių lygų 8 gautas NARX modelis visų metrikų atžvilgiu šiek tiek geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 8-ojo klasterio akcijų gražos laiko eilutėms (viso 186 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.16 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 8-ojo klasterio akcijų gražoms prognozuoti.

3.16 lentelė. Apibendrinti visų 8-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	0E+00	0.001569	34.78	40.43
		Vidurkis	0E+00	0.004804	65.24	65.15
		Maksimumas	0E+00	0.04069	86.96	80.85
	Testavimo imtis	Minimumas	-0.0143714	0.001263	14.29	20
		Vidurkis	0.0005421	0.006718	49.47	49.55
		Maksimumas	0.0093061	0.044096	85.71	80
NARX	Apmokymo imtis	Minimumas	-1.00E-06	4.00E-07	54.35	63.83
		Vidurkis	0.00E+00	4.90E-06	98.77	98.61
		Maksimumas	5.00E-07	4.47E-05	100	100
	Testavimo imtis	Minimumas	-0.0279361	0.001573	21.43	26.67
		Vidurkis	0.0007949	0.007671	56.7	50.21
		Maksimumas	0.0128462	0.058715	100	80

9-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 13 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atkarpos koeficiento, termino *debt* tendencijos, infliacijos, pramonės produkcijos augimo, trumpalaikės palūkanų normos, nedarbo lygio rodiklių. Regresijos modelio prognozė visų metrikų atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 4.655e-07$), o testavimo imčiai lygus 8 ($RMSE = 0.003008$). Parinkus optimalų neuronų skaičių lygų 8 gautas NARX modelis visų metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 9-ojo klasterio akcijų gražos laiko eilutėms (viso 568 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.17 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 9-ojo klasterio akcijų gražoms prognozuoti.

3.17 lentelė. Apibendrinti visų 9-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-1.30E-16	0.0002206	6.522	34.78
		Vidurkis	-2.37E-18	0.006912	61.945	62.09
		Maksimumas	2.21E-16	0.0403208	86.957	80.43
	Testavimo imtis	Minimumas	-0.040246	0	0	13.33
		Vidurkis	0.0000999	0.008993	49.35	52.09
		Maksimumas	0.0887144	0.104448	85.71	100
NARX	Apmokymo imtis	Minimumas	-5.15E-06	6.61E-08	10.87	46.81
		Vidurkis	-5.23E-09	1.28E-05	98.09	98.15
		Maksimumas	4.15E-06	3.11E-04	100	100
	Testavimo imtis	Minimumas	-0.0168994	0.0008572	0	13.33
		Vidurkis	0.0014497	0.0095154	50.57	52.42
		Maksimumas	0.06705	0.0753166	85.71	93.33

10-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 14 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atmesta termino *inflation* tendencijos, vartotojų kainų indekso ir pramonės produkcijos augimo rodiklių. Regresijos modelio prognozė visų metrikų atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 6.301e-07$), o testavimo imčiai lygus 12 ($RMSE = 0.00335$). Parinkus optimalų neuronų skaičių lygų 12 gautas NARX modelis visų metrikų atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 10-ojo klasterio akcijų gražos laiko eilutėms (viso 137 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.18 lentelė) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 10-ojo klasterio akcijų gražoms prognozuoti.

3.18 lentelė. Apibendrinti visų 10-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-2.66E-17	0.001168	41.3	50
		Vidurkis	-2.32E-18	0.003916	63.07	64.09
		Maksimumas	2.08E-17	0.02	84.78	82.98
	Testavimo imtis	Minimumas	-0.0058377	0.0009597	14.29	13.33
		Vidurkis	-0.0002312	0.0052608	49.05	49.93
		Maksimumas	0.0085168	0.0242726	85.71	80
NARX	Apmokymo imtis	Minimumas	-9.77E-07	2.21E-07	63.04	68.09
		Vidurkis	-7.09E-09	8.57E-07	98.97	98.66
		Maksimumas	5.36E-07	4.76E-06	100	100
	Testavimo imtis	Minimumas	-1.13E-02	0.001548	7.143	26.67
		Vidurkis	4.39E-04	0.006007	52.972	51.09
		Maksimumas	1.36E-02	0.022281	85.714	86.67

11-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 15 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra termino *chance* tendencijos, infliacijos, nedarbo lygio rodiklių. Regresijos modelio prognozė visų metrių atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 9 ($RMSE = 5.07e-07$), o testavimo imčiai lygus 11 ($RMSE = 0.00543$). Parinkus optimalų neuronų skaičių lygų 11 gautas NARX modelis *POS* ir *POCID* metrių atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 11-ojo klasterio akcijų gražos laiko eilutėms (viso 120 įmonių). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.19 lentelę) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 11-ojo klasterio akcijų gražoms prognozuoti.

3.19 lentelė. Apibendrinti visų 11-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-1.04E-17	0.001736	39.13	46.81
		Vidurkis	-5.13E-19	0.003669	64.07	64.72
		Maksimumas	9.63E-18	0.011224	82.61	82.98
	Testavimo imtis	Minimumas	-0.0058553	0.001449	7.143	14.29
		Vidurkis	0.0004699	0.004723	49.4	49.45
		Maksimumas	0.0067163	0.013797	85.714	93.33
NARX	Apmokymo imtis	Minimumas	-2.54E-07	2.95E-07	82.61	85.11
		Vidurkis	-2.03E-09	8.00E-07	99.75	99.57
		Maksimumas	2.57E-07	2.79E-06	100	100
	Testavimo imtis	Minimumas	-0.0052988	0.001854	14.29	20
		Vidurkis	0.0008724	0.005688	54.82	52.17
		Maksimumas	0.0147089	0.020125	85.71	86.67

12-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 16 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra atkarpos koeficiento, termino *conflict*, *inflation* tendencijos, trumpalaikės palūkanų normos, nedarbo lygio rodiklių. Regresijos modelio prognozė visų metrikų atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 10 ($RMSE = 3.560019e-07$), o testavimo imčiai lygus 10 ($RMSE = 0.004753$). Parinkus optimalų neuronų skaičių lygų 10 gautas NARX modelis tik *POS* metrikos atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 12-ojo klasterio akcijų gražos laiko eilutėms (viso 648 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.20 lentelę) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 12-ojo klasterio akcijų gražoms prognozuoti.

3.20 lentelė. Apibendrinti visų 12-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-5.44E-17	0.0004308	2.174	29.79
		Vidurkis	-8.60E-19	0.0086703	59.395	61.15
		Maksimumas	4.72E-17	0.0435107	82.222	82.98
	Testavimo imtis	Minimumas	-0.161151	0	0	6.667
		Vidurkis	-1.06E-03	0.011777	46.99	52.368
		Maksimumas	4.23E-02	0.176902	92.86	100
NARX	Apmokymo imtis	Minimumas	-1.62E-06	7.32E-08	2.174	48.94
		Vidurkis	-1.28E-10	2.80E-06	97.098	96.9
		Maksimumas	1.93E-06	3.67E-05	100	100
	Testavimo imtis	Minimumas	-7.05E-02	0.0001511	0	0
		Vidurkis	-9.08E-05	0.0127434	50.49	51.97
		Maksimumas	4.02E-02	0.1136641	100	100

13-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 17 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingas koeficientas yra tik infliacijos rodiklio. Regresijos modelio prognozė visų metrikų atžvilgiu gana nebloga. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 6.631587e-07$), o testavimo imčiai lygus 9 ($RMSE = 0.004863176$). Parinkus optimalų neuronų skaičių lygų 9 gautas NARX modelis tik *POS* metrikos atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 13-ojo klasterio akcijų gražos laiko eilutėms (viso 151 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.21 lentelę) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 13-ojo klasterio akcijų gražoms prognozuoti.

3.21 lentelė. Apibendrinti visų 13-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-8.25E-18	0.001342	46.67	46.81
		Vidurkis	-5.83E-19	0.003951	61.36	63.37
		Maksimumas	1.08E-17	0.011656	82.61	78.72
	Testavimo imtis	Minimumas	-0.0098735	0.001506	0	14.29
		Vidurkis	-0.0002582	0.006002	46.9	51.43
		Maksimumas	0.00504	0.03266	92.31	85.71
NARX	Apmokymo imtis	Minimumas	-6.78E-07	2.79E-07	91.3	89.36
		Vidurkis	-6.85E-10	1.35E-06	99.91	99.56
		Maksimumas	3.07E-07	1.99E-05	100	100
	Testavimo imtis	Minimumas	-0.0052231	0.002133	14.29	20
		Vidurkis	0.0025222	0.00672	52.55	49.45
		Maksimumas	0.0144799	0.030836	85.71	80

14-ojo klasterio akcijų prognozės modelių tyrimas. Iš rezultatų pateiktų 18 priede regresijos modelio rezultatų lentelės, matome, kad modelio reikšmingi koeficientai yra termino *debt* tendencijos, pramonės produkcijos augimo, akcijų kainų indekso, nedarbo lygio rodiklių. Regresijos modelio prognozė visų metrikų atžvilgiu prasta. Šiame priede pateikto NARX modelio optimalaus paslėptojo sluoksnio neuronų skaičiaus parinkimo rezultatų matome, kad apmokymo imčiai geriausi rezultatai gauti jau su neuronų skaičiumi lygiu 8 ($RMSE = 7.487490e-07$), o testavimo imčiai lygus 8 ($RMSE = 0.003068220$). Parinkus optimalų neuronų skaičių lygų 8 gautas NARX modelis tik *POS* metrikos atžvilgiu geresnis nei regresijos modelis. Sudarytus modelius pritaikome visoms 14-ojo klasterio akcijų gražos laiko eilutėms (viso 209 įmonės). Lyginant prognozės rezultatus testavimo imčiai (žr. 3.22 lentelę) matome, kad abiejų modelių gauti prognozės rezultatai labai nežymiai skiriasi, todėl abu modeliai tinkami 14-ojo klasterio akcijų gražoms prognozuoti.

3.22 lentelė. Apibendrinti visų 14-ojo klasterio akcijų modelių kokybės įvertinimo rezultatai.

Modelis	Imtis	Statistika	ME	RMSE	POCID, %	POS, %
Regresija	Apmokymo imtis	Minimumas	-3.61E-17	0.0006466	43.48	46.81
		Vidurkis	-7.56E-19	0.0049488	61.41	63.38
		Maksimumas	2.87E-17	0.0464825	82.61	80.85
	Testavimo imtis	Minimumas	-1.56E-02	0	0	20
		Vidurkis	-6.60E-05	0.00665	47.12	50.77
		Maksimumas	4.96E-02	0.080123	85.71	100
NARX	Apmokymo imtis	Minimumas	-1.25E-06	1.74E-07	63.04	55.32
		Vidurkis	3.17E-09	4.73E-06	98.79	98.13
		Maksimumas	3.97E-06	8.97E-05	100	100
	Testavimo imtis	Minimumas	-0.0363323	0.0005134	0	6.667
		Vidurkis	0.0006468	0.0077693	52.63	50.845
		Maksimumas	0.0344324	0.0649644	85.71	80

IŠVADOS

1. Atlikus Europos šalių akcijų duomenų analizę įvairiais lygiais, nustatyta, kad egzistuoja panašių ir skirtingų akcijų gražų tendencijų kitimas laike, kuris reikšmingai neišsiskiria nei šalių, nei biržų lygiuose. Todėl norint sudaryti sėkmingą prognozės modelį reikia suskirstyti akcijas į grupes pagal jų panašumą.

2. Patikrinus analizuojamų 3 202 akcijų gražos laiko eilučių stacionarumo ir autokoreliacijos testus nustatyta, kad dauguma laiko eilučių yra stacionarus procesas ir kiekvienos eilutės duomenys yra tarpusavyje nepriklausomi.

3. Ištyrus analizuojamų finansinių *Google* tendencijų tarpusavio priklausomybes, rasti septyni stipriai koreliuoti kintamieji, kurie pašalinti iš duomenų rinkinio, t. y. netraukti į prognozės modelį. Analogiškai rastos ir pašalintos dviejų makroekonominių rodiklių priklausomybės.

4. Atlikus padalijimu paremto ir hierarchinio klasterizavimo metodų, atstumo matų ir jungimo algoritmų lyginamąją analizę, nustatyta, kad akcijoms klasterizuoti geriausiai tinka padalijimu paremtas k -formos metodas. Tokiu būdu akcijos sugrupuotos į 14-a grupių.

5. Ypatingųjų reikšmių dekompozicijos metodu rastos kiekvieno klasterio reikšmingiausios akcijų gražos laiko eilutės. Šioms eilutėms sudaryti regresijos ir netiesinių autoregresinių neuroninių tinklų modeliai, kurie pritaikyti visų klasterių Europos šalių akcijų gražai prognozuoti. Nustatyta, kad abiejų metodų prognozės kokybė reikšmingai nesiskiria, tačiau galima teigti, kad neuroninio tinklo prognozė yra šiek tiek kokybiškesnė akcijų kitimo tendencijos ir pelno/nuosmukio prognozės atžvilgiu.

6. Atlikus gautų tiesinės regresijos modelių analizę, nustatyta, kad daugumos grupių akcijų gražos prognozės modelyje terminų *debt* (skola), *inflation* (infliacija), *earnings* (pajamos) paieškos dydžiai *Google* sistemoje ir pramonės produkcijos augimo, trumpalaikės palūkanų normos, nedarbo lygio bei infliacijos rodikliai yra statistiškai reikšmingi kintamieji.

7. Tikėtina, kad netiesiniu autoregresiniu neuroninių tinklų modeliu panaudojant daugiau duomenų (kasdienius, o ne mėnesinius) tinklo apmokymui būtų gauta kokybiškesnė akcijų gražos prognozė. Taip pat naudojant realaus laiko išorinių veiksnių duomenis būtų galima sudaryti labai kokybišką nuolat atnaujinimą modelį.

LITERATŪROS ŠARŠAS

- [1] Stock market prediction [interactive]. Wikipedia, the free encyclopedia. 2017 [viewed 2017 03 01]. Available from: https://en.wikipedia.org/wiki/stock_market_prediction
- [2] YATES T. 4 Ways To Predict Market Performance [interactive]. Investopedia, 2017 [viewed 2017 03 01]. Available from: http://www.investopedia.com/articles/07/mean_reversion_martingale.asp
- [3] Why Log Returns [interactive]. Uncommon Returns through Quantitative and Algorithmic Trading, 2012 [viewed 2017 03 01]. Available from: <https://quantivity.wordpress.com/2011/02/21/why-log-returns/>
- [4] RAPACHA D. E., WOHARB M. E., RANGVIDC J. Macro variables and international stock return predictability [interactive]. *Elsevier Science Direct International Journal of Forecasting*, 2005, 21, 137– 166. Available from: Science Direct
- [5] SHUE-JEN W., WEI-MING L. Predicting severe simultaneous bear stock markets using macroeconomic variables as leading indicators [interactive]. *Elsevier Science Direct Finance Research Letters*, 2015, 21, 196–204. Available from: Science Direct
- [6] NYBERG H., PÖNKÄ H. International Sign Predictability of Stock Returns: The Role of the United States [interactive]. *Center for Research in Econometrics Analysis of Time Series*, 2015 [viewed 2017 03 01]. Available from: http://econ.au.dk/fileadmin/site_files/filer_oekonomi/Working_Papers/CREATES/2015/rp15_20.pdf
- [7] SURESH BABU M., GEETHANJALI N., SATYANARAYANA B. Clustering Approach to Stock Market Prediction [interactive]. *International Journal of Advanced Networking and Applications*, 2012 3(4), 1281-1291. Available from: <http://www.ijana.in/papers/V3I4-10.pdf>
- [8] PREIS T., REITH D., AND EUGENE H. Complex dynamics of our economic life on different scales: insights from search engine query data [interactive]. *The Royal Society*: 2010, 368 (1933). Available from: doi: 10.1098/rsta.2010.0284
- [9] PREIS T., MOAT H. S., and EUGENE H. Quantifying Trading Behavior in Financial Markets Using Google Trends [interactive]. *Scientific Reports*: 2013, 1684 (2013). Available from: doi: 10.1038/srep01684
- [10] ROSEN F. Correlation based clustering of the Stockholm Stock Exchange [interactive]. Stockholm university, 2006. Available from: <http://www.diva-portal.org/smash/get/diva2:196577/FULLTEXT01.pdf>
- [11] SUGANTHI R., KAMALAKANNAN P. Analyzing Stock Market Data Using Clustering Algorithm [interactive]. *International Journal of Future Computer and Communication*, 2015 4(2). Available from: <http://www.ijfcc.org/vol4/366-IT307.pdf>
- [12] DOHERTY K. A. J., ADAMS R. G., DAVEY N., and PENSUWON W. Hierarchical Topological Clustering Learns Stock Market Sectors [interactive]. Available from: <http://uhra.herts.ac.uk/bitstream/handle/2299/776/900869.pdf?sequence=1>
- [13] BINI B.S., TESSY M. Clustering and Regression Techniques for Stock Prediction. *Procedia Technology* [interactive]. Elsevier 2016, 24, 1248-1255. Available from: Science Direct doi: <https://doi.org/10.1016/j.protcy.2016.05.104>

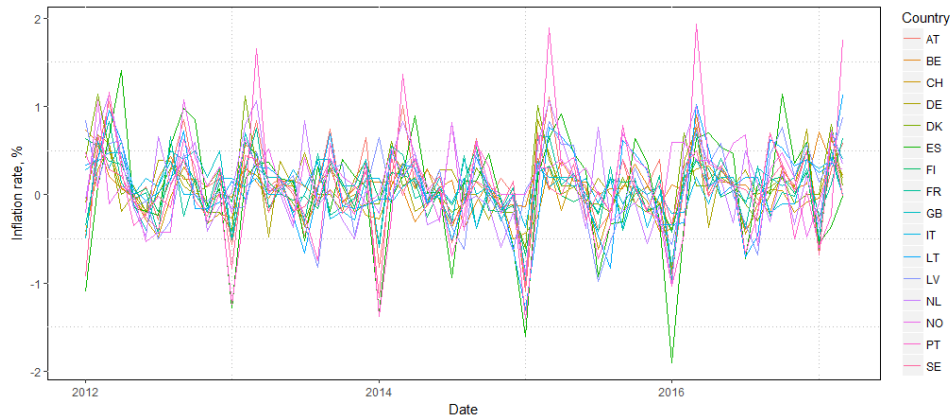
- [14] AMORIM NETO M. C. and others. Stock Price Forecasting Using Exogenous Time Series and Combined Neural Networks [interactive]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.358.1149&rep=rep1&type=pdf>
- [15] GRIGORYAN H. Stock Market Prediction using Artificial Neural Networks. Case Study of TALIT, Nasdaq OMX Baltic Stock [interactive]. *Database systems journal* ISSN:2069-3230. Available from: http://www.dbjournal.ro/archive/20/20_2.pdf
- [16] COCIANU C. L., GRIGORYAN H. An Artificial Neural Network for Data Forecasting Purposes [interactive]. *Informatica Economică*, 2015, 19(2) [viewed 2017 03 01]. Available from: doi: 10.12948/issn14531305/19.2.2015.04
- [17] SARD'A-ESPINOSA A. Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package [interactive]. Cran R-Project Vignettes. 2017 [viewed 2017 03 20]. Available from: <https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf>
- [18] MONTERO P., VILAR J. A. TSclust: An R Package for Time Series Clustering [interactive]. *Journal of Statistical Software*. 2014, 62(1). Available from: <http://www.jstatsoft.org/>
- [19] PAPARRIZOS J., GRAVANO L. k-Shape: Efficient and Accurate Clustering of Time Series [interactive]. 2015 [viewed 2017 03 20]. Available from: doi: <http://dx.doi.org/10.1145/2723372.2737793>
- [20] DRAGAŠIUS E., LIPNICKAS A. *Intelektualios mechatroninės sistemos: mokomoji knyga* [interaktyvus]. Vilnius: Vilniaus pedagoginio universiteto leidykla, 2008, 29 – 71 [žiūrėta 2017 04 01] UDK 621.313.3(075.8). Prieiga per: <https://www.ebooks.ktu.lt/eb/437/intelektualios-mechatronines-sistemos/>
- [21] TSAY R. S. *Analysis of Financial Time Series*. 2nd Edition. Wiley- Interscience, 2005. ISBN-13 978-0-471-69074-0
- [22] Analysis of Discrete Data: Introduction to Generalized Linear Models [interactive]. *STAT 504 – Analysis of Discrete Data*, 2017 [viewed 2017 04 20]. Available from: <https://onlinecourses.science.psu.edu/stat504/node/216>
- [23] CASANUEVA A. A. Google google on the wall, who is ... [interactive]. Kinea, 2015 [viewed 2017 03 20]. Available from: <http://kinea.media/en/politics-society/google-trends-on-the-wall>
- [24] BERRAR D.P., DUBITZKY W., GRANZOW M. A Practical Approach to Microarray Data Analysis: Singular value decomposition and principal component analysis [interactive]. Kluwer: Norwell, MA, 2003, 91-109 [viewed 2017 04 01] LANL LA-UR-02-4001. Available from: <http://www.cs.wustl.edu/~zhang/teaching/cs517/Spring12/CourseProjects/SVD.pdf>
- [25] GORDON M. Using the SVD to find the needle in the haystack [interactive]. G-forge a blog about orthopaedic surgery, R, research and more, 2013 [viewed 2017 04 01]. Available from: <http://gforge.se/2013/04/using-the-svd-to-find-the-needle-in-the-haystack/>

1 PRIEDAS. ANALIZUOJAMŲ BIRŽŲ SĄRAŠAS

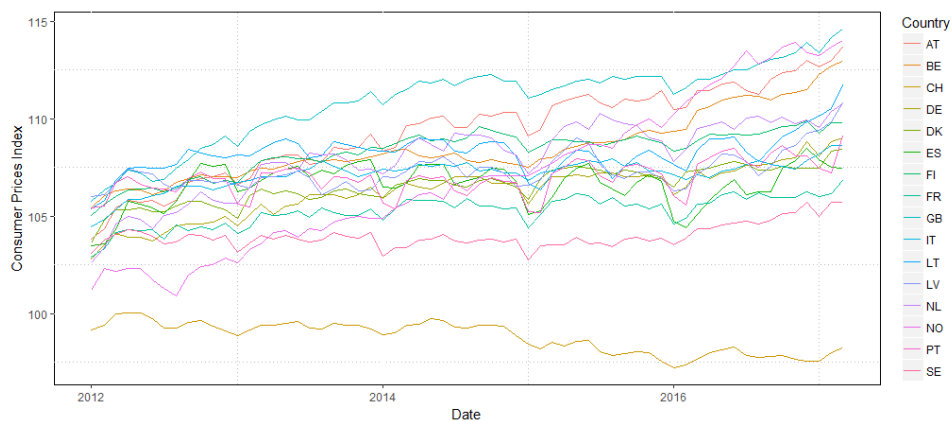
1 lentelė. Yahoo Finance svetainėje pateiktų ir šiame darbe analizuojamų biržų rinkų sąrašas

Šalis	Rinkos sufiksas	Rinkos pavadinimas	Rinkos pavadinimas angl. kalba
Nyderlandai	AS	NYSE Euronext – Amsterdamas	NYSE Euronext – Amsterdam
Italija	MI	Milano vertybinių popierių birža	Milan Stock Exchange
Ispanija	MC	Madrido SE C.A.T.S.	Madrid SE C.A.T.S.
Danija	CO	Kopenhagos vertybinių popierių birža	Copenhagen Stock Exchange
Belgija	BR	NYSE Euronext - Briuselis	NYSE Euronext - Brussels
Portugalija	LS	NYSE Euronext – Lisabona	NYSE Euronext – Lisbon
Prancūzija	PA	NYSE Euronext – Paryžius	NYSE Euronext – Paris
Vokietija	DE	XETRA vertybinių popierių birža	XETRA Stock Exchange
Vokietija	F	Frankfurto vertybinių popierių birža	Frankfurt Stock Exchange
Suomija	HE	Nasdaq OMX Helsinkis	Nasdaq OMX Helsinki
Didžioji Britanija	L	Londono vertybinių popierių birža	London Stock Exchange
Latvija	RG	Nasdaq OMX Ryga	Nasdaq OMX Riga
Švedija	ST	Stokholmo vertybinių popierių birža	Stockholm Stock Exchange
Šveicarija	SW	Šveicarijos birža	Swiss Exchange
Austrija	VI	Vienos vertybinių popierių birža	Vienna Stock Exchange
Lietuva	VS	Nasdaq OMX Vilnius	Nasdaq OMX Vilnius
Vokietija	DU	Diuseldorfo vertybinių popierių birža	Dusseldorf Stock Exchange
Vokietija	BE	Berlyno vertybinių popierių birža	Berlin Stock Exchange
Vokietija	MU	Miuncheno vertybinių popierių birža	Munich Stock Exchange
Vokietija	SG	Štutgarto vertybinių popierių birža	Stuttgart Stock Exchange
Vokietija	HA	Hanoverio vertybinių popierių birža	Hanover Stock Exchange
Vokietija	HM	Hamburgo vertybinių popierių birža	Hamburg Stock Exchange
Italija	TI	EuroTLX	EuroTLX
Norvegija	OL	Oslo vertybinių popierių birža	Oslo Stock Exchange
Švedija	ST	Stokholmo vertybinių popierių birža	Stockholm Stock Exchange
Airija	IR	Airijos vertybinių popierių birža	Irish Stock Exchange
Didžioji Britanija	IL	Londono vertybinių popierių birža	London Stock Exchange

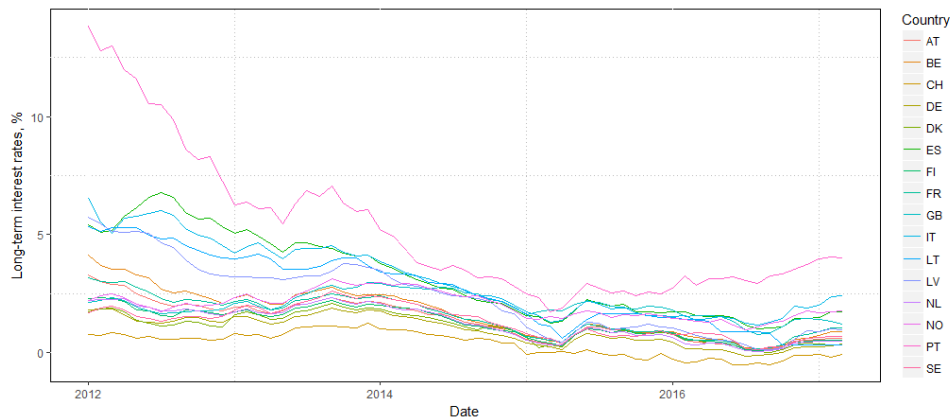
2 PRIEDAS. MAKROEKONOMINIŲ RODIKLIŲ KITIMAS ANALIZUOJAMOSE ŠALYSE



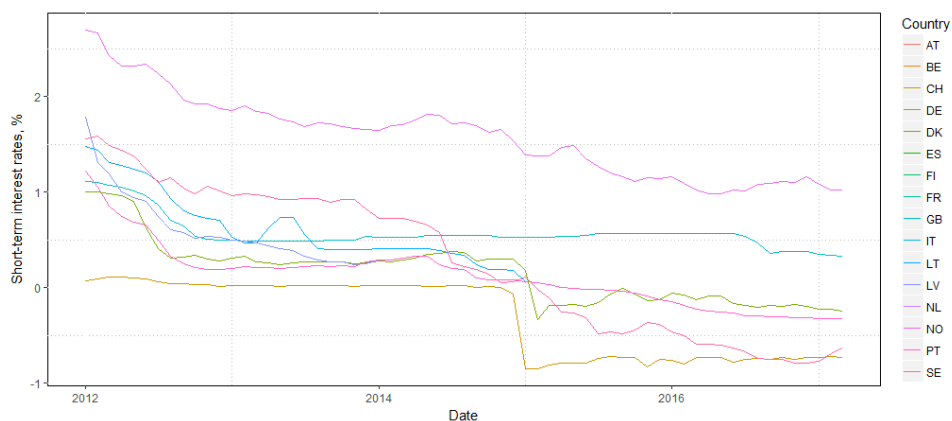
1. pav. Infliacijos kitimas pamėnesiui analizuojamose šalyse



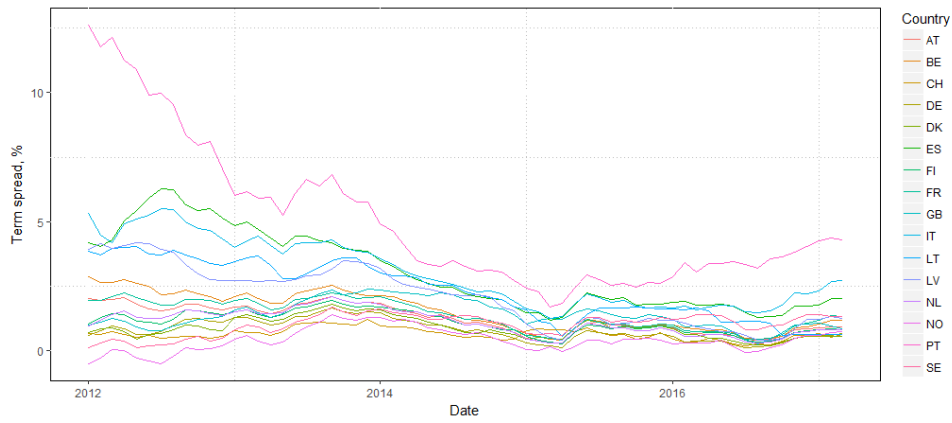
2. pav. Vartotojų kainų indekso kitimas pamėnesiui analizuojamose šalyse



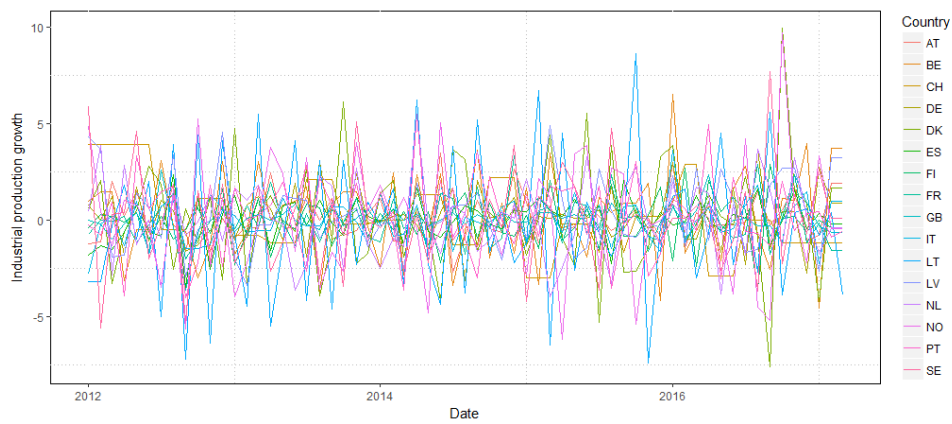
3. pav. Ilgalaikių palūkanų normų kitimas pamėnesiui analizuojamose šalyse



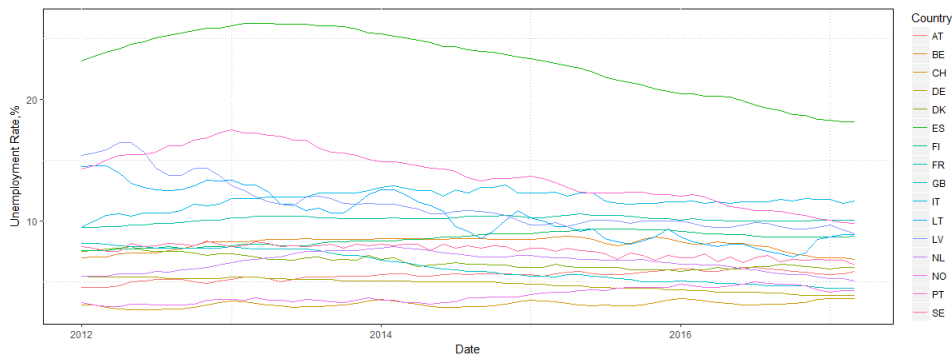
4. pav. Trumpalaikių palūkanų normų kitimas pamėnesiui analizuojamose šalyse



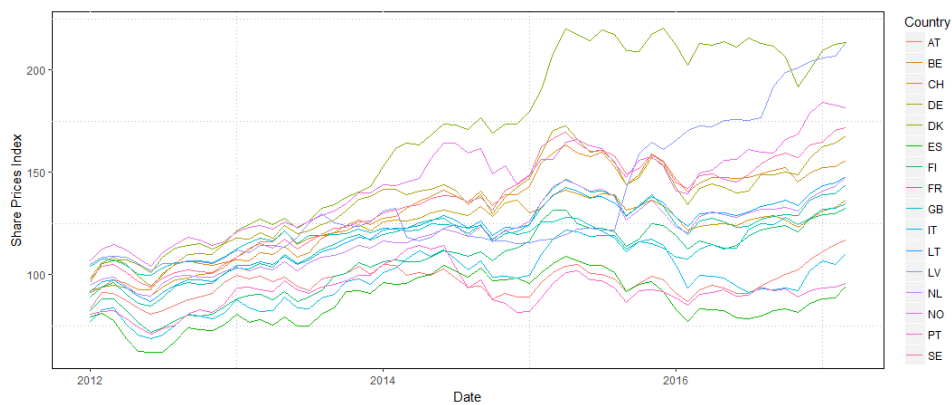
5. pav. Palūkanų normų apimties kitimas pamėnesiui analizuojamose šalyse



6. pav. Pramonės produkcijos augimo kitimas pamėnesiui analizuojamose šalyse

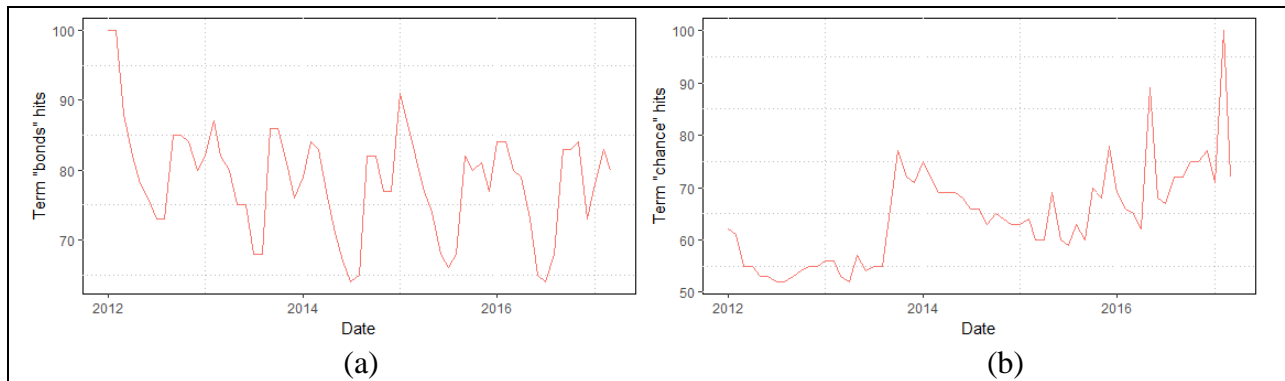


7. pav. Nedarbo lygio kitimas pamėnesiui analizuojamose šalyse

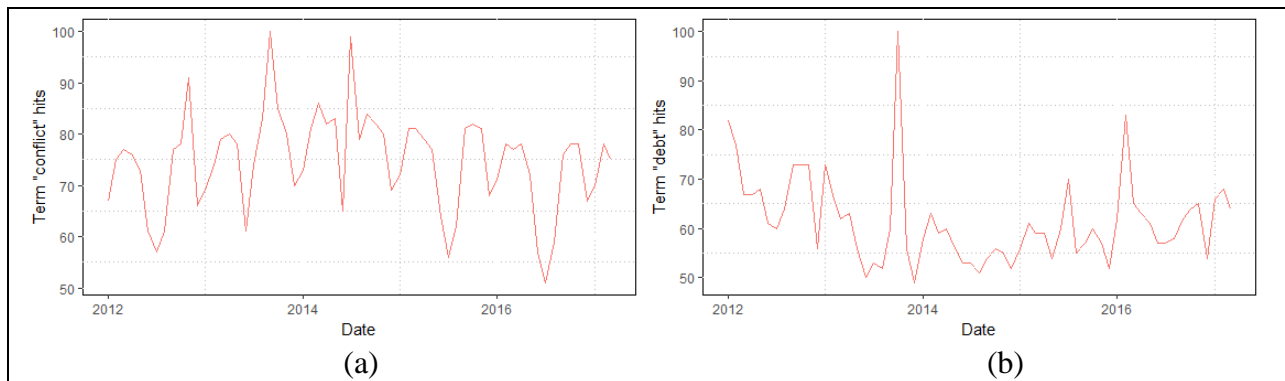


8. pav. Akcijų kainų indekso kitimas pamėnesiui analizuojamose šalyse

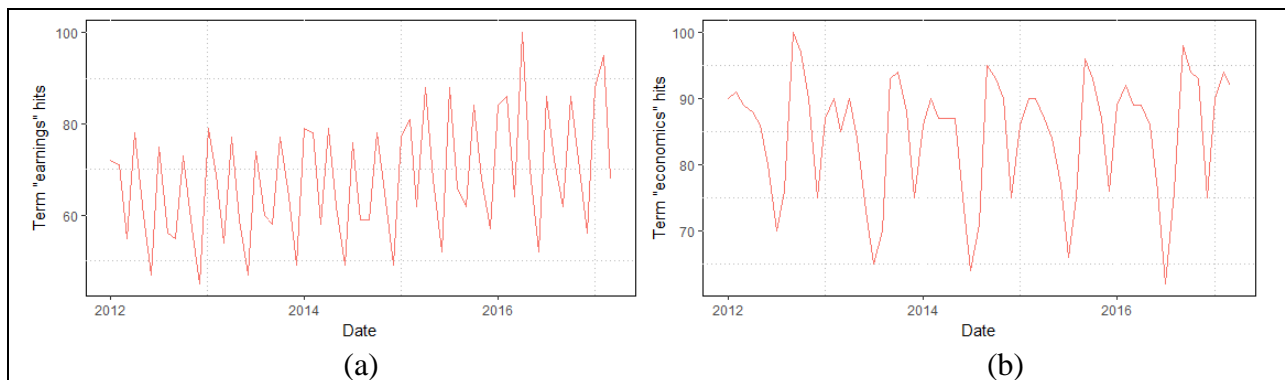
3 PRIEDAS. FINANSINIŲ TERMINŲ PAIEŠKOS APIMČIŲ *GOOGLE* SISTEMOJE KITIMAS



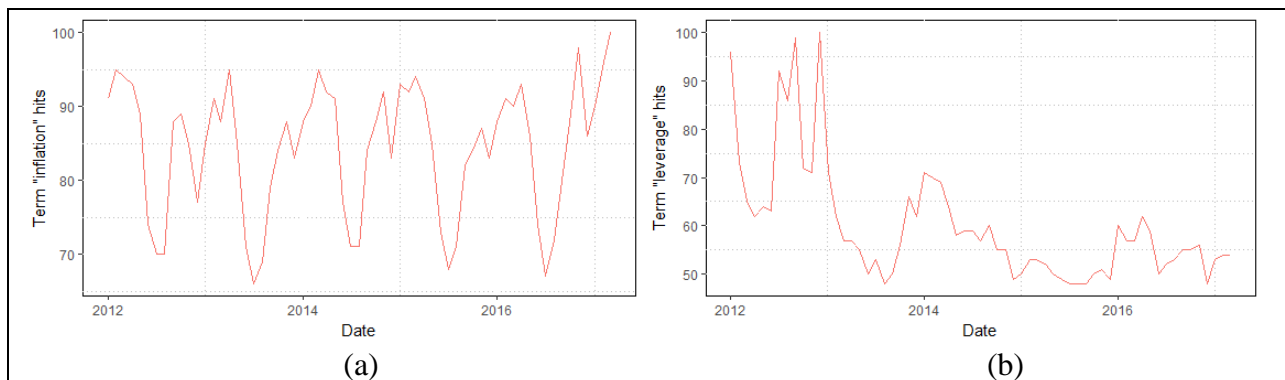
9. pav. (a) – termino „bonds“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „chance“



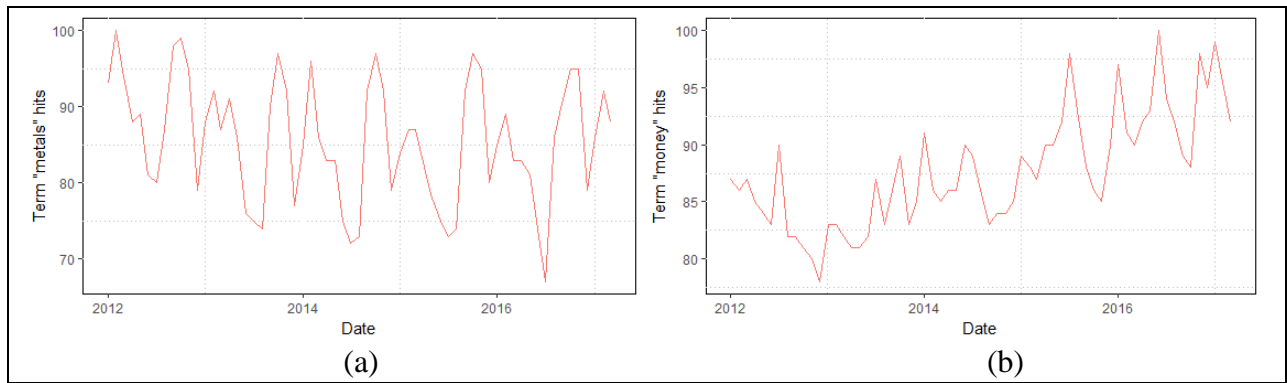
10. pav. (a) – termino „conflict“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „debt“



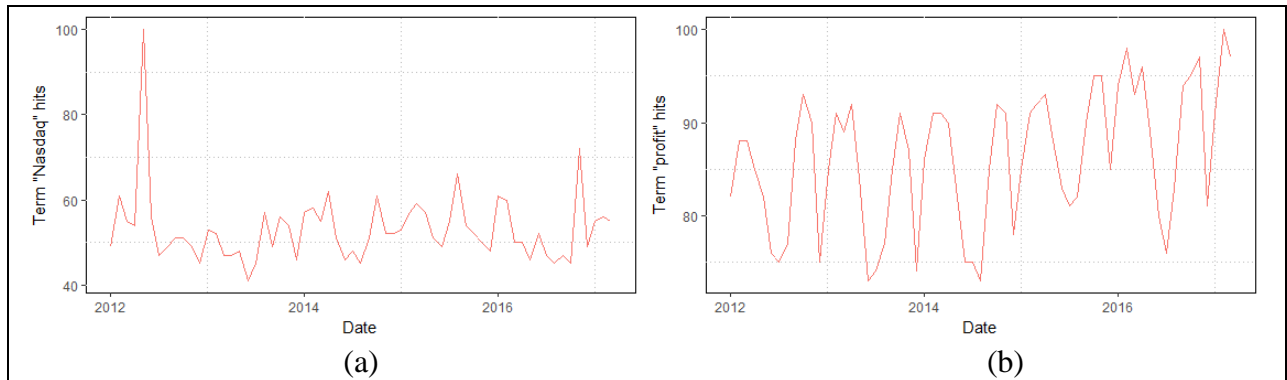
11. pav. (a) – termino „earnings“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „economics“



12. pav. (a) – termino „inflation“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „leverage“



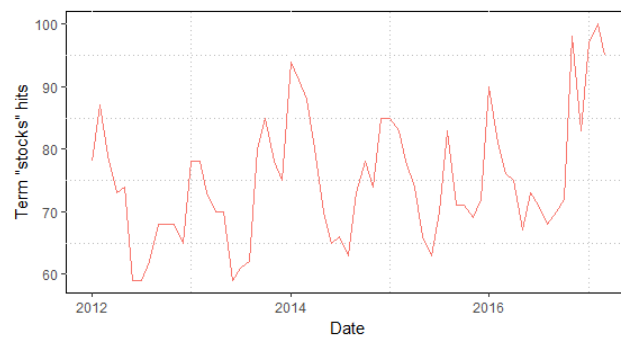
13. pav. (a) – termino „metals“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „money“



14. pav. (a) – termino „Nasdaq“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „profit“



15. pav. (a) – termino „risk“ paieškos *Google* sistemoje kitimas pamėnesiui, (b) – termino „sell“

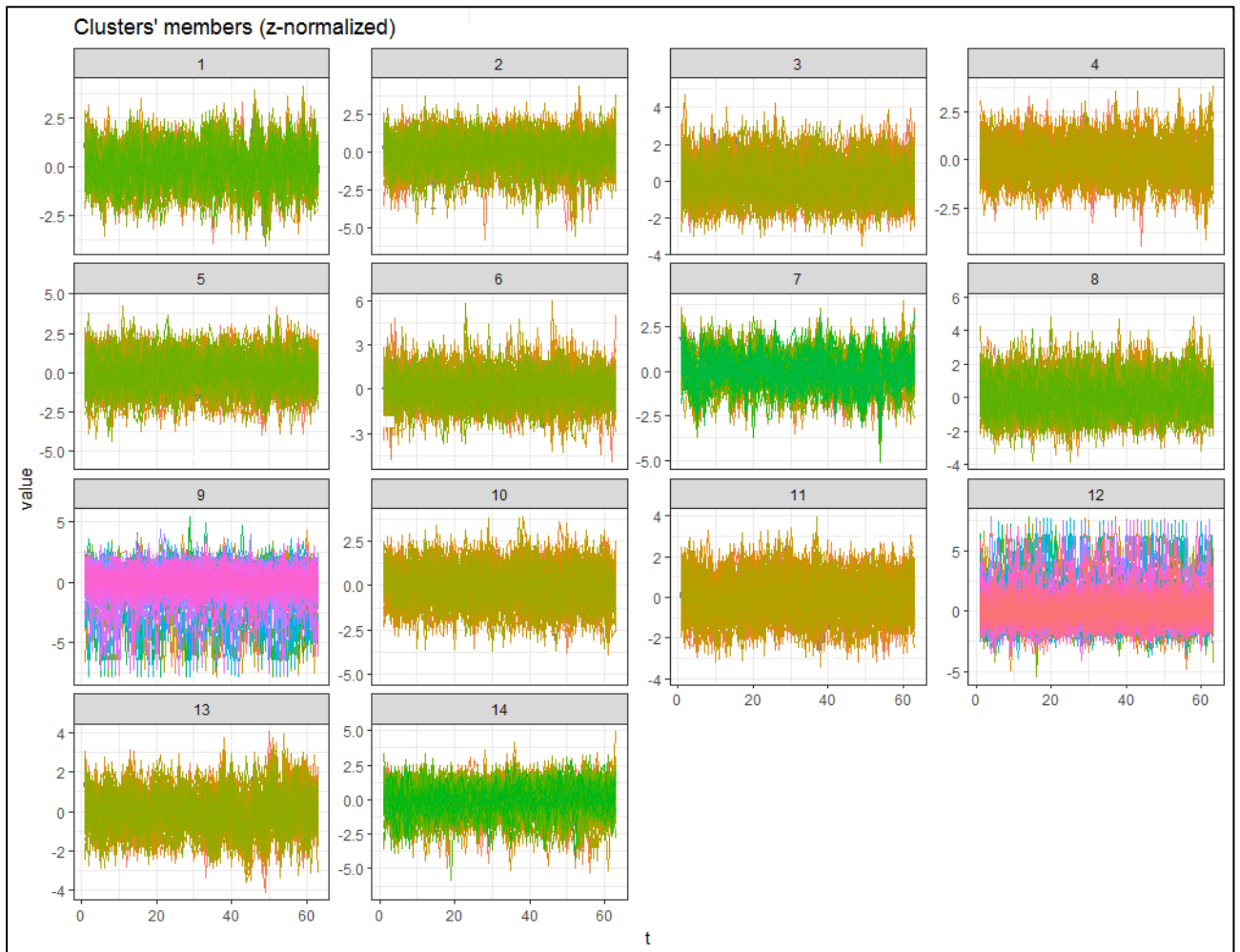


16. pav. Termino „stocks“ paieškos *Google* sistemoje kitimas pamėnesiui

4 PRIEDAS. KLASTERIŲ APRAŠOMOJI STATISTIKA IR POŽYMIAI

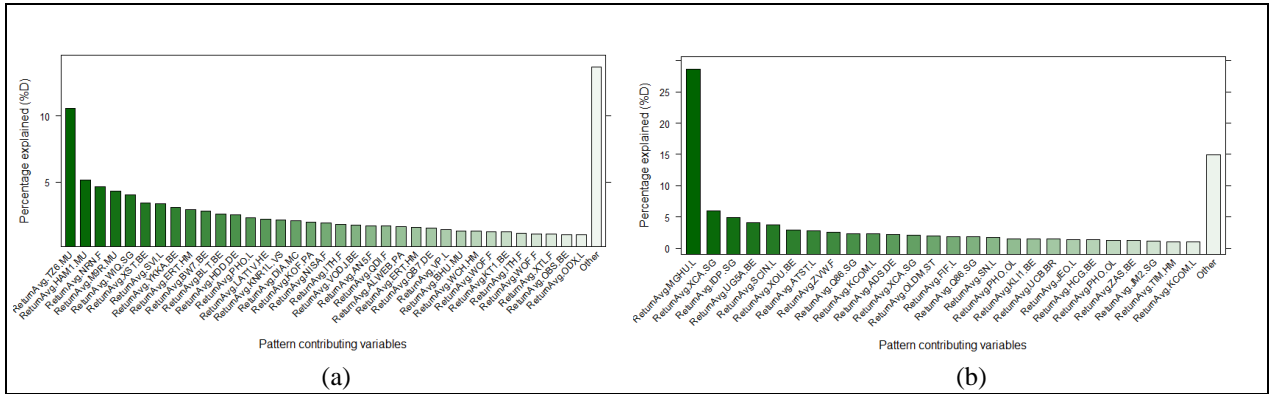
2 lentelė. Gautų klasterių aprašomoji statistika ir požymiai

Nr.	Įmonių sk.	Šalių sk.	Biržų sk.	Uždarymo kaina			Apyvarta			Logaritmuota grąža			Pagrindinė šalis	Šalies dažnis klasteryje / Šalies dažnis duomenų rinkinyje	Pagrindinės biržos	Biržos dažnis klasteryje / Biržos dažnis duomenų rinkinyje
				min.	vid.	maks.	vid.	maks.	min.	vid.	maks.					
1	190	9	16	1.00E-02	7.45E+02	1.67E+05	8.85E+04	1.40E+07	-3.98E-02	6.26E-04	3.47E-02	DE (84%) GB(9%)	CH(10%)	F (30%) BE(20%) MU(15%)	SW(10%) F(10%)	
2	168	10	17	6.60E-04	4.43E+01	1.31E+03	5.95E+05	4.98E+08	-1.18E-01	1.33E-04	1.18E-01	DE (77%) GB(12%)	AT(13%)	BE (22%) F (18%) L(12%)	V(13%)	
3	143	12	19	4.80E-03	4.93E+01	1.42E+03	1.40E+05	1.52E+07	-2.21E-01	2.75E-04	2.32E-01	DE (70%) GB (15%)	LT(17%) SE(10%)	F(20%) BE(17%) L(15%)	VS(17%) ST(10%)	
4	120	10	17	2.21E-02	3.78E+02	8.16E+03	7.29E+05	1.63E+07	-2.34E-02	4.28E-04	2.89E-02	DE(45%) GB(37%)	BE(13%) NL(11%) SE(10%)	L(36%) F(12%)	BR(13%) AS(11%) PA(8.53%)	
5	176	12	19	1.00E-03	9.48E+01	3.83E+03	3.21E+05	1.18E+08	-8.95E-02	8.96E-05	8.67E-02	DE(77%) GB (14%)	LV(17%) PT(17%)	BE(21%) F(18%) L(14%)	LS(17%) RG(17%)	
6	163	15	22	1.00E-03	1.35E+02	5.43E+03	2.11E+05	3.99E+07	-2.36E-01	2.56E-04	2.22E-01	DE(70%) GB(11%)	LV(17%) PT(17%) CH(11%)	F(18%) BE(16%) L(11%)	LS(17%) RG(17%) SW(11%)	
7	223	15	21	5.12E-02	2.95E+03	8.37E+05	4.36E+06	2.64E+08	-2.72E-02	5.09E-04	2.47E-02	GB(37%) DE(23%) FR(10%)	ES(30%) FR(28%) IT(26%) ...	L(36%) PA(10%) DE(7%)	MC(30%) PA(28%) MI(26%)	
8	186	12	19	2.77E-03	1.55E+02	2.11E+04	1.23E+05	3.01E+07	-1.12E-01	2.47E-04	1.19E-01	DE(72%) GB(12%)	BE(20%) FI(14%) CH(10%)	F(18%) MU(15%) BE(15%) L(11%)	BR(20%) HE(14%) SW(10%)	
9	568	14	22	1.00E-03	1.30E+03	1.06E+07	3.23E+05	1.07E+08	-2.85E-01	-3.09E-04	2.17E-01	DE(77%) GB(14%)	DE(19%) DK(19%) PT(16%) ..	F (19%) BE(17%) MU(13%) L(12%)	IL(40%) HA(29%) HM(24%)	
10	137	10	18	3.76E-03	1.23E+03	2.40E+05	1.06E+05	1.48E+07	-6.19E-02	3.32E-04	6.02E-02	DE(81%) GB(10%)	LV(17%) FI(14%)	F(26%) BE(18%) DU(12%)	RG(17%) HE(14%)	
11	120	11	18	1.01E-02	1.08E+02	5.45E+03	4.24E+05	3.68E+07	-3.47E-02	4.59E-04	4.49E-02	DE(80%) GB(8%)	NL(10%)	BE(20%) MU(17%) F(12%) DU(12%)	HA(12.5%) AS(10.8%)	
12	648	16	24	9.00E-04	2.95E+02	2.51E+05	5.95E+05	1.34E+08	-1.09E-01	6.07E-04	2.74E-01	DE(63%) GB(20%)	LV(50%) LT(47%) NO(35%)	L(20%) F(17%) BE(16%)	RG(50%) VS(47%) OL(35%)	
13	151	10	17	4.33E-03	6.58E+02	9.01E+04	1.03E+06	1.32E+08	-4.60E-02	3.35E-05	5.23E-02	DE(67%) GB(22%)	PT(17%)	BE(26%) L(18%) F(16%)	IL(33%) LS(17%)	
14	209	11	18	1.00E-03	1.13E+02	8.98E+03	3.04E+05	2.83E+07	-1.73E-01	6.40E-05	1.65E-01	DE(75%) GB(16%)	ES(11%)	BE(20%) F(19%) L(16%) MU(12%) SG(10%)	HM(11%) MC(11%)	

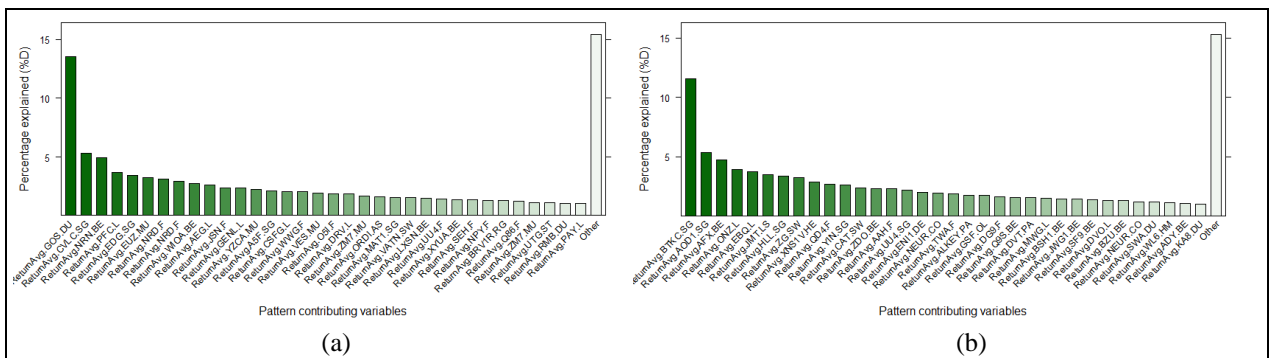


17. pav. Akcijų duomenų išskaidymas pagal gautus 14-a klasterių

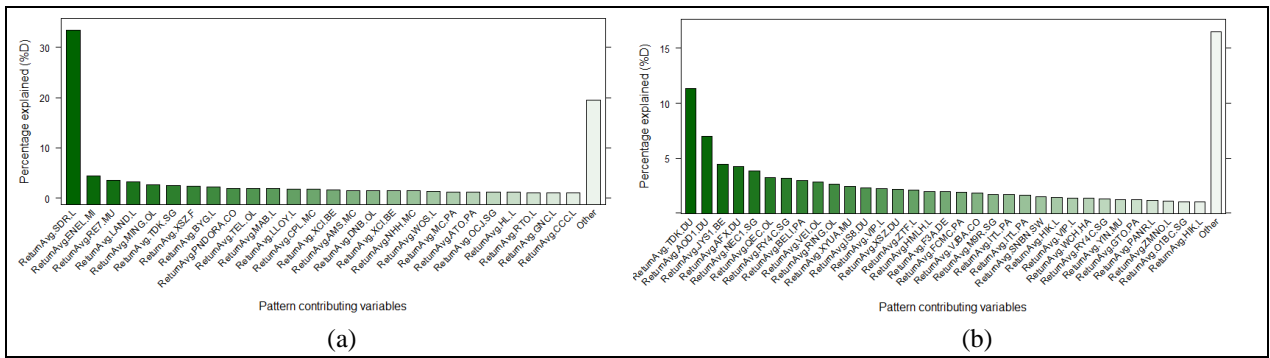
5 PRIEDAS. KLASTERIŲ REIŠMINGŲŲ AKCIJŲ NUSTATYMO REZULTATAI



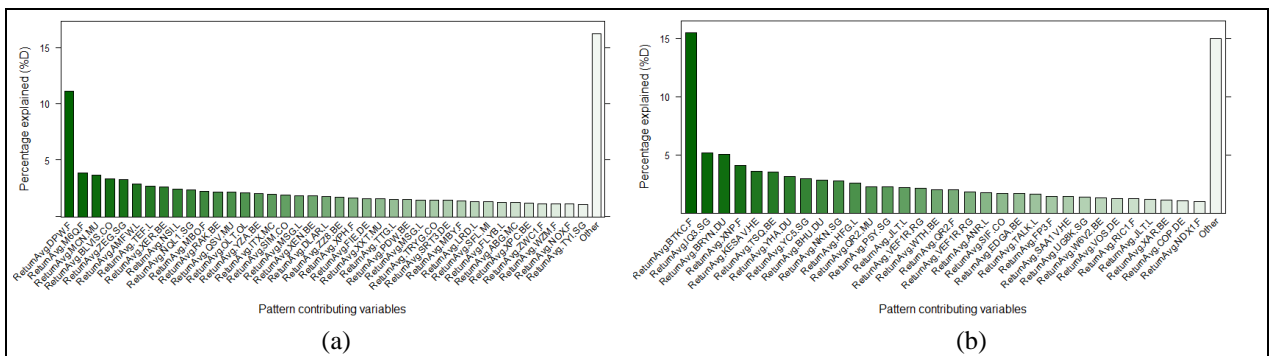
18. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 3-ojo klasterio; (b) dalyje – 4-ojo klasterio



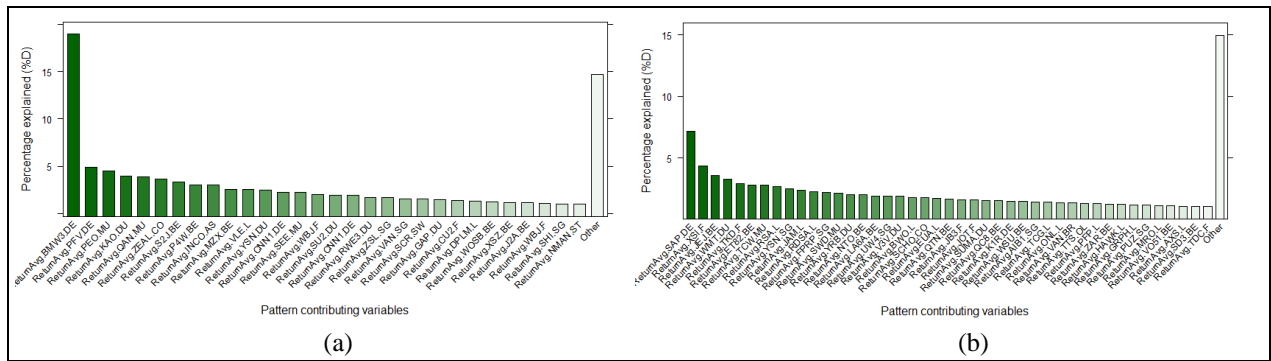
19. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 5-ojo klasterio; (b) dalyje – 6-ojo klasterio



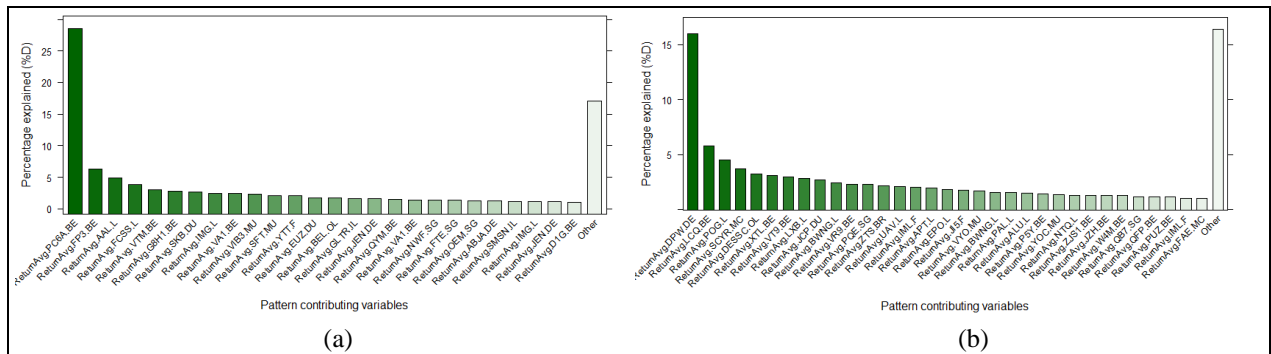
20. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 7-ojo klasterio; (b) dalyje – 8-ojo klasterio



21. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 9-ojo klasterio; (b) dalyje – 10-ojo klasterio



22. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 11-ojo klasterio; (b) dalyje – 12-ojo klasterio



23. pav. Klasterio kintamųjų (akcijų) pagrindinėse komponentėse paaishkinama dispersijos procentinė dalis; (a) dalyje – 13-ojo klasterio; (b) dalyje – 14-ojo klasterio

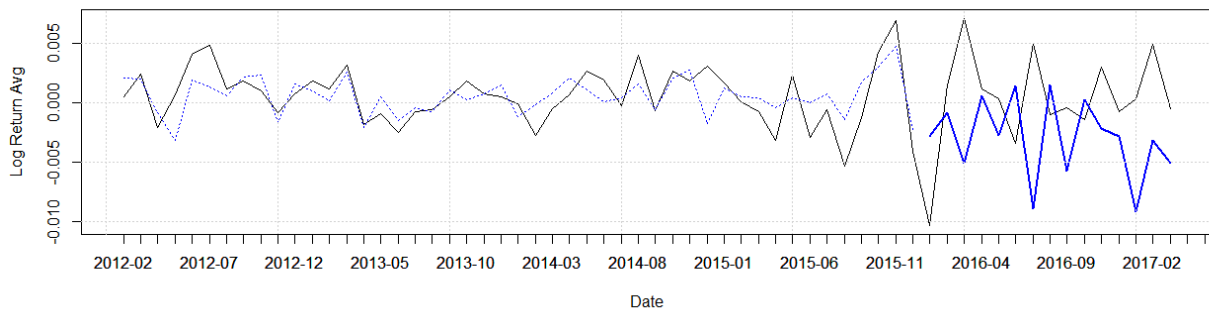
6 PRIEDAS. 2-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

1. lentelė 2-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

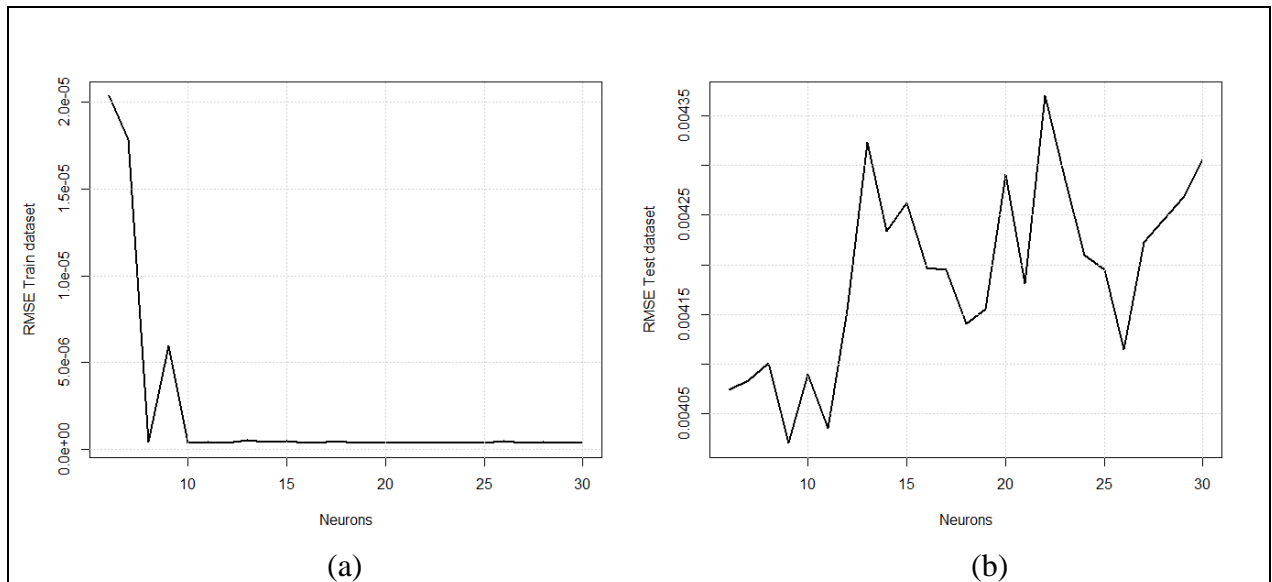
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	1.0363	0.1253	8.269	3.43E-10	***
earnings_hits_lag1	-0.1817	0.1131	-1.607	0.11584	
inflation_hits_lag1	0.1952	0.1006	1.94	0.05943	.
money_hits_lag1	-0.2551	0.1195	-2.135	0.03897	*
INFPP_lag1	-0.4315	0.1237	-3.487	0.0012	**
IPG_PP_lag1	-0.3395	0.1129	-3.007	0.00454	**
IR3TIB_lag1	-0.2265	0.1206	-1.878	0.0677	.

2. lentelė 2-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

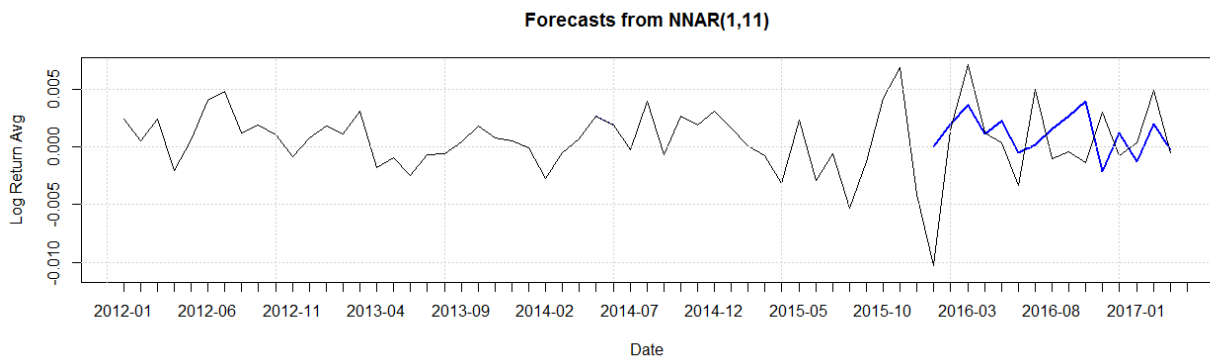
	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-1.77e-18	0.0018	93.33	64.28
Testavimo imtis	0.00335	0.00674	35.71	33.33



24. pav. 2-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos regresijos (mėlyna punktyrinė – apmokymo imtis, mėlyna išsistinė – testavimo imtis).



25. pav. RMSE paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties



26. pav. 1-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

3. lentelė 2-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-2.26e-08	3.41e-07	100	100
Testavimo imtis	-7.27e-04	4.01e-03	57.14	53.33

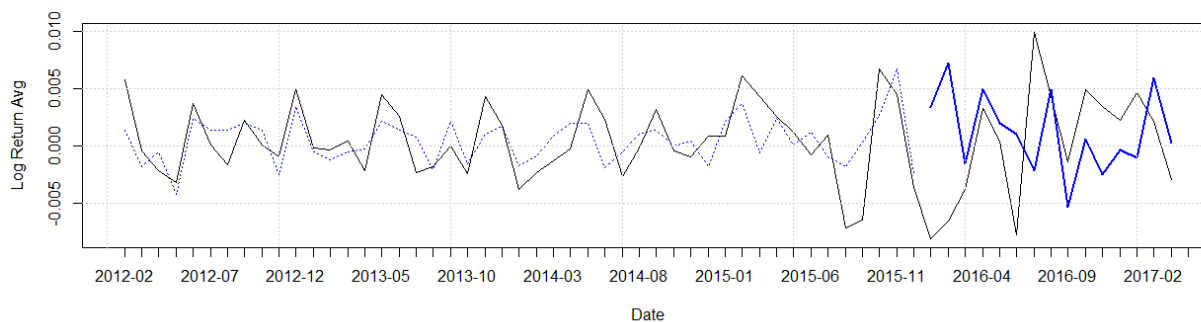
7 PRIEDAS. 3-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

4. lentelė 3-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

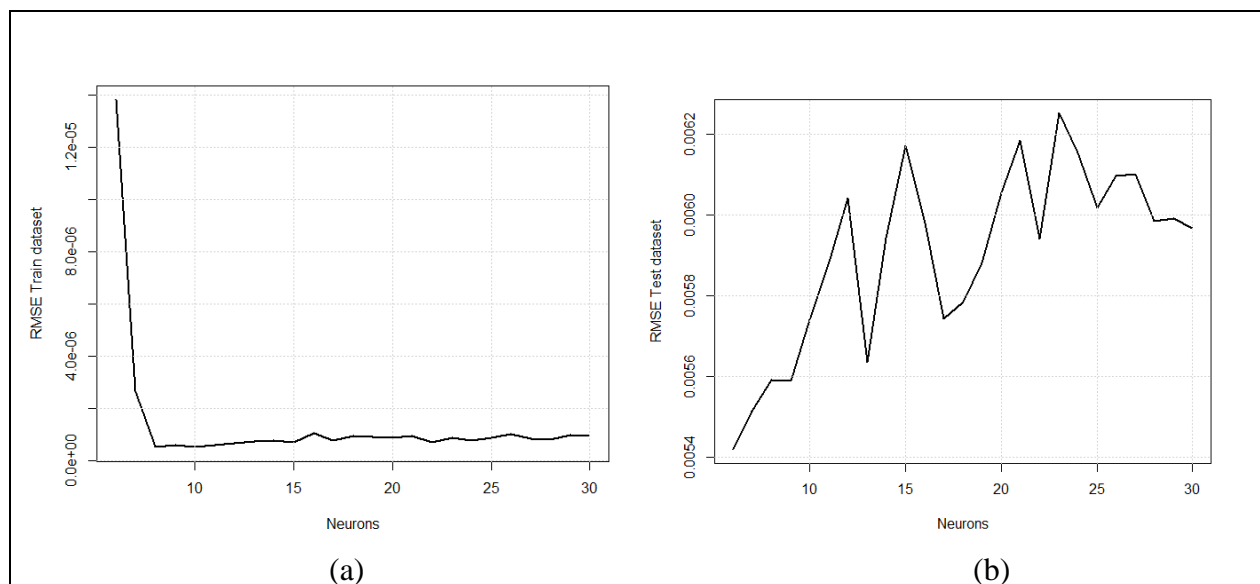
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.2325	0.2045	1.137	0.26216	
inflation_hits_lag1	0.2105	0.1069	1.97	0.05566	.
INFPP_lag1	-0.4297	0.1423	-3.019	0.00435	**
IPG_PP_lag1	-0.1857	0.1331	-1.395	0.17052	
SHARE_PRICES_lag1	0.6042	0.1752	3.449	0.00131	**
UNEMP_R_lag1	0.3778	0.1377	2.744	0.00895	**

5. lentelė 3-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-5.55e-18	0.0025	78.57	86.66
Testavimo imtis	-0.00084	0.0067	57.14	46.66



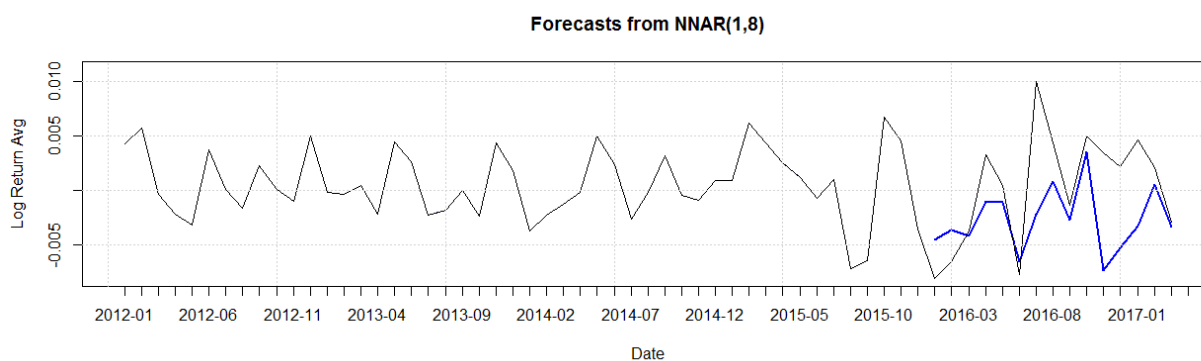
27. pav. 3-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



28. pav. RMSE paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

6. lentelė 3-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	3.433e-08	5.822e-07	100	100
Testavimo imtis	3.025e-03	5.436e-03	71.428	60.00



29. pav. 3-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

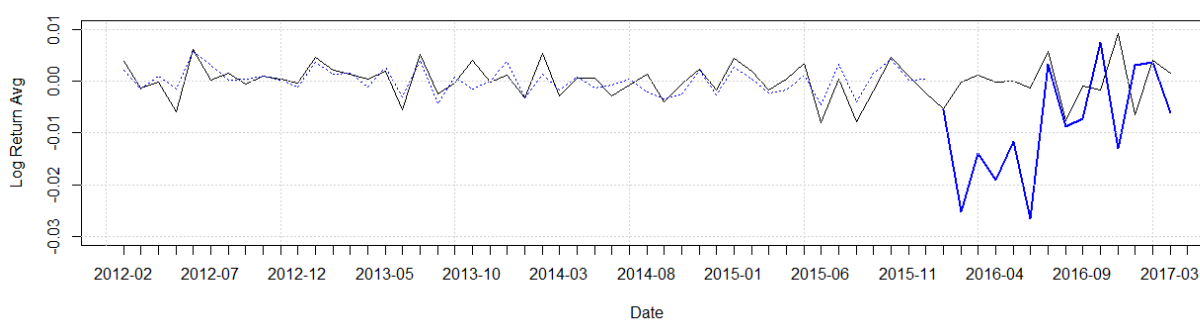
8 PRIEDAS. 4-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

7. lentelė 4-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

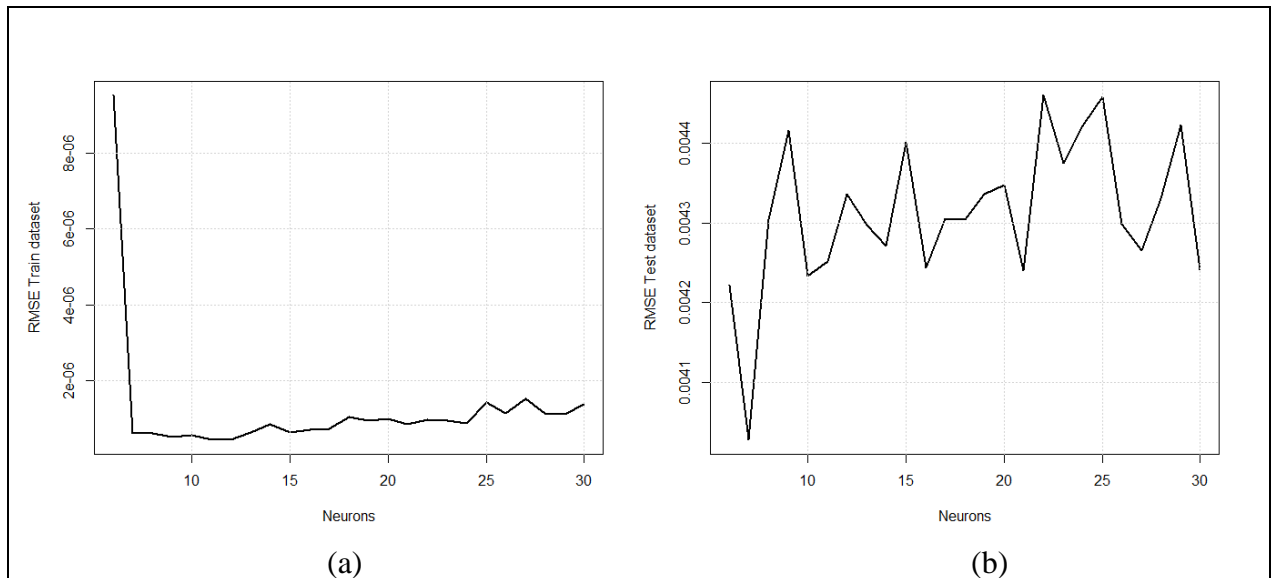
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	-0.4696	0.475	-0.989	0.32979	
hits_lag1	0.2424	0.1559	1.555	0.12912	
chance_hits_lag1	-0.479	0.2423	-1.977	0.05623	.
debt_hits_lag1	-0.509	0.2154	-2.363	0.024	*
inflation_hits_lag1	0.3865	0.1333	2.901	0.00649	**
leverage_hits_lag1	0.1934	0.1544	1.253	0.21891	
money_hits_lag1	1.3597	0.2758	4.93	2.12E-05	***
Nasdaq_hits_lag1	-0.577	0.2086	-2.765	0.00912	**
CPI_lag1	2.006	0.598	3.355	0.00196	**
IPG_PP_lag1	-0.5081	0.1978	-2.568	0.01479	*
SHARE_PRICES_lag1	-1.1404	0.3174	-3.592	0.00102	**
UNEMP_R_lag1	1.001	0.3383	2.959	0.00558	**
ReturnAvg_lag1	-0.6347	0.1329	-4.777	3.34E-05	***

8. lentelė 4-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	4.306e-18	0.00208	78.57	80.00
Testavimo imtis	0.00838	0.0140	69.23	64.28



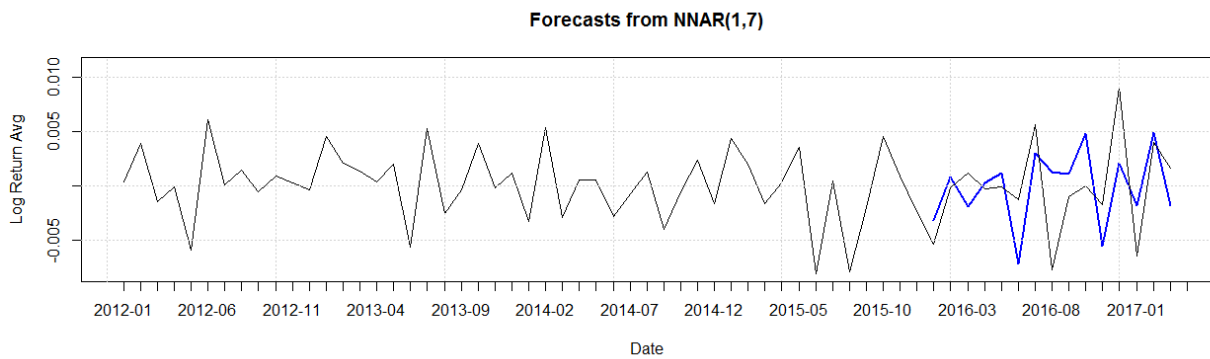
30. pav. 4-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modelių (mėlyna punktyrinė – apmokymo imtis, mėlyna išsistinė – testavimo imtis).



31. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

9. lentelė 4-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	5.178e-08	6.421066e-07	100.00	100.00
Testavimo imtis	-2.897e-05	4.191632e-03	78.57	53.33



32. pav. 4-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

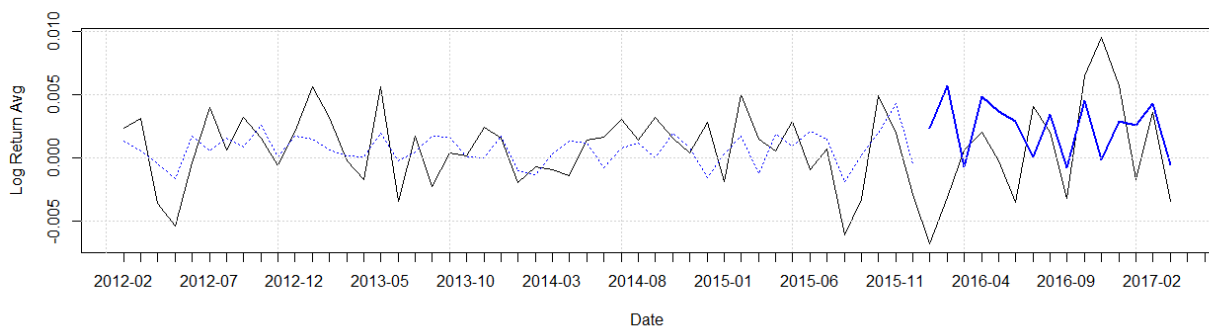
9 PRIEDAS. 5-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

10. lentelė 5-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

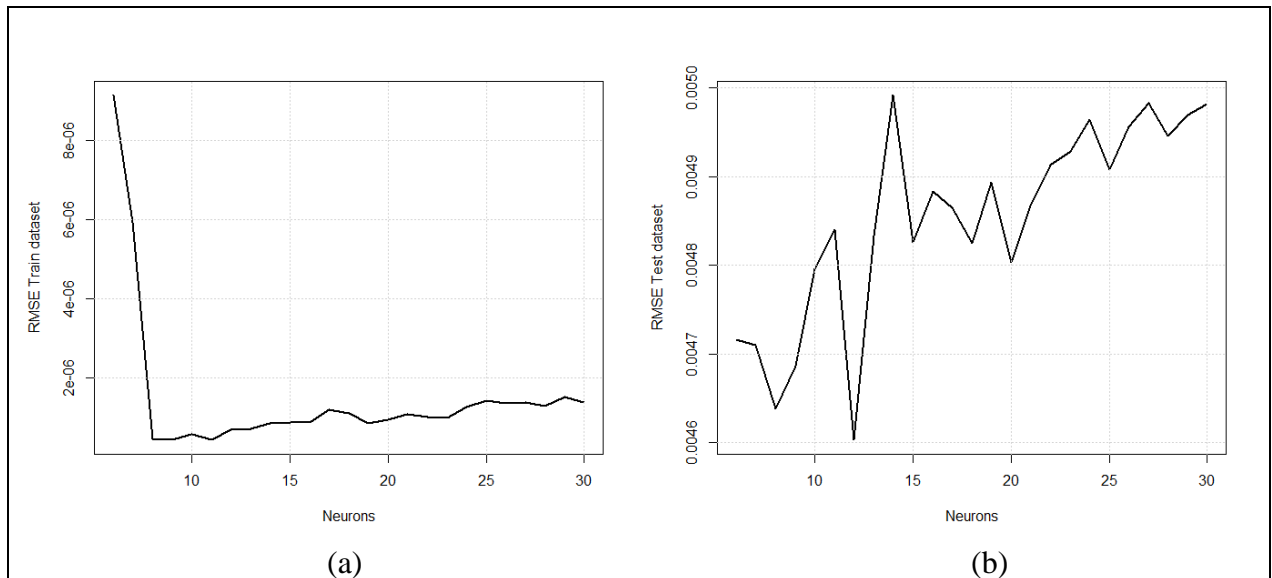
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.816	0.1418	5.755	8.28E-07	***
INFPP_lag1	-0.3524	0.1606	-2.194	0.0337	*
IPG_PP_lag1	-0.3028	0.1511	-2.004	0.0514	.
SHARE_PRICES_lag1	0.2133	0.1198	1.781	0.0819	.

11. lentelė 5-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	2.10348e-18	0.00240	45.65	70.21
Testavimo imtis	-0.001538	0.00507	64.28	53.33



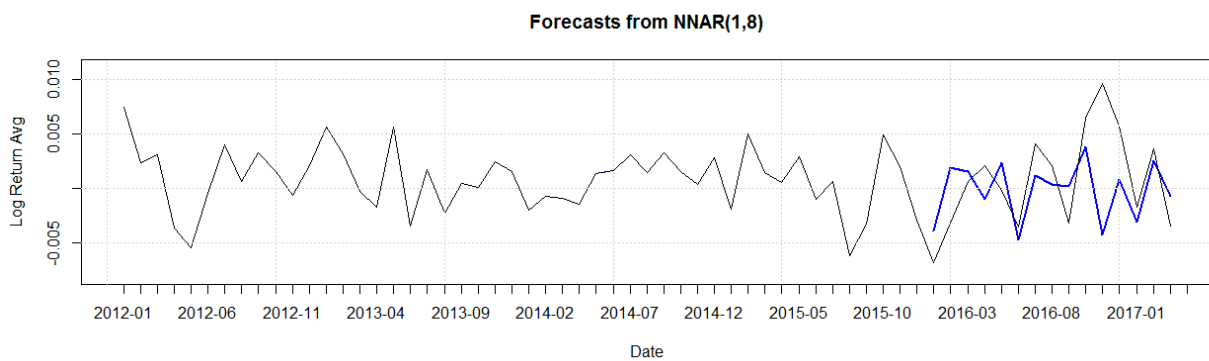
33. pav. 5-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



34. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

12. lentelė 5-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-1.502e-07	9.332e-07	100.00	100.00
Testavimo imtis	1.009e-03	4.535e-03	64.28	66.67



35. pav. 5-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

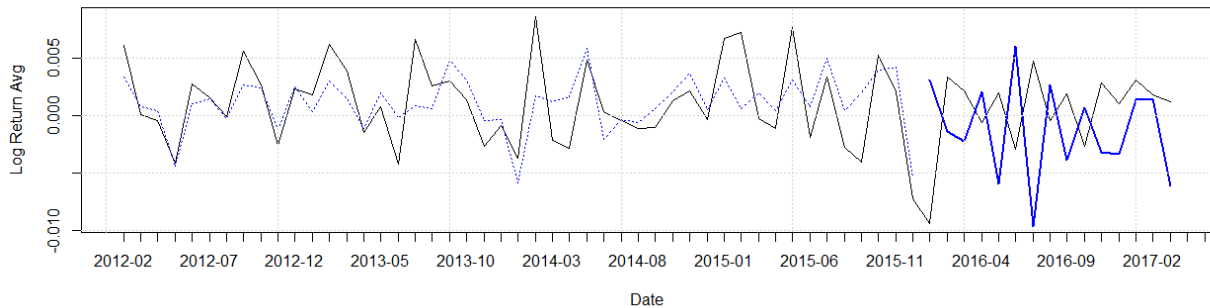
10 PRIEDAS. 6-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

13. lentelė 6-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

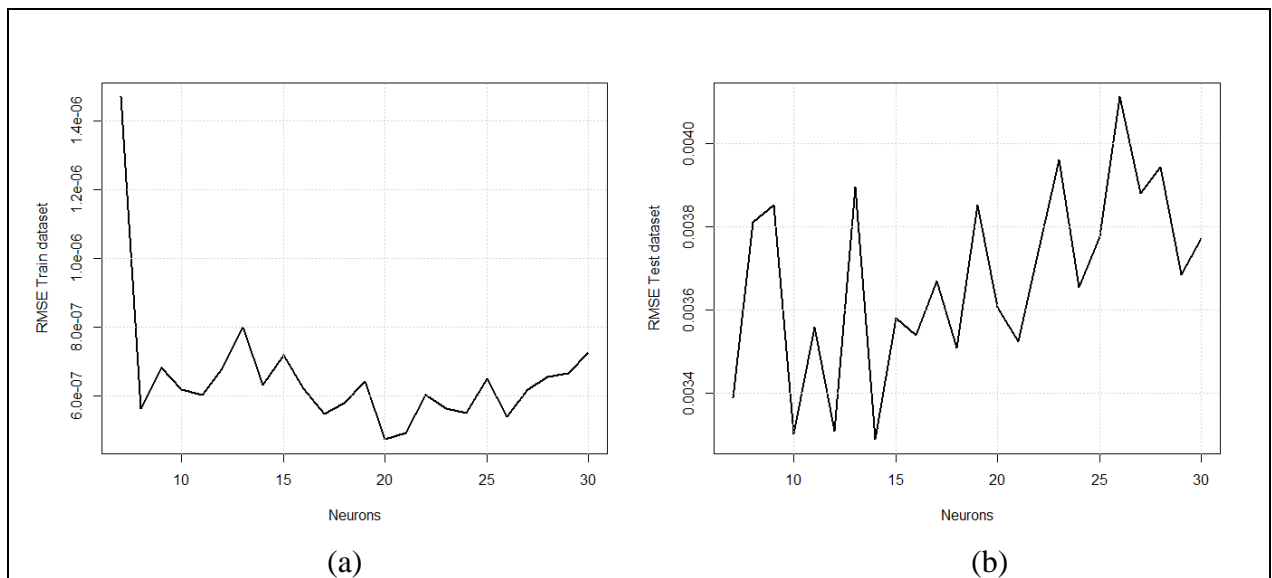
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.1526	0.342	0.446	0.657961	
chance_hits_lag1	-0.3216	0.2313	-1.39	0.172779	
conflict_hits_lag1	-0.3618	0.2219	-1.63	0.111533	
debt_hits_lag1	0.4624	0.2213	2.09	0.043565	*
earnings_hits_lag1	-0.2304	0.1506	-1.53	0.134642	
inflation_hits_lag1	0.4935	0.1509	3.269	0.002335	**
CPI_lag1	0.8875	0.384	2.311	0.026501	*
INFPP_lag1	-0.2274	0.1551	-1.466	0.150999	
IPG_PP_lag1	-0.5033	0.1408	-3.574	0.000999	***
UNEMP_R_lag1	0.3656	0.2499	1.463	0.151867	

14. lentelė 6-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-7.563e-18	0.00275	84.78	76.59
Testavimo imtis	0.00178	0.00697	28.57	13.33



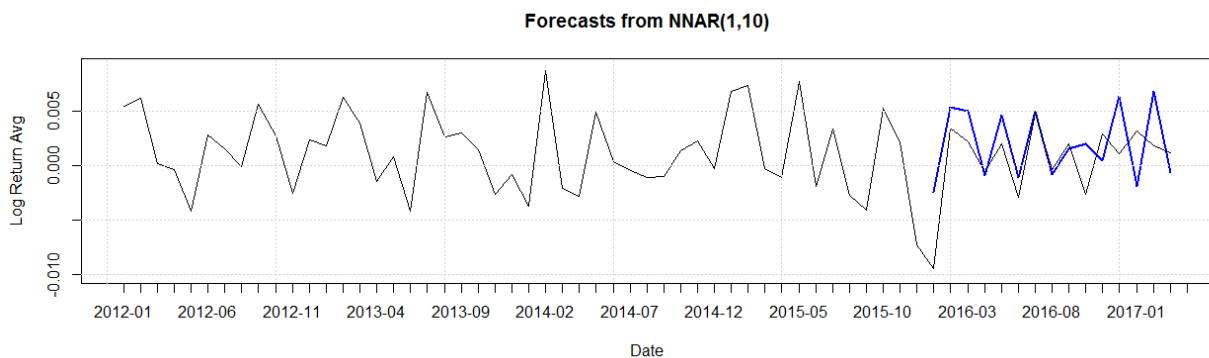
36. pav. 6-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modelių (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



37. pav. RMSE paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

15. lentelė 6-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	1.759e-08	6.219e-07	100.00	100.00
Testavimo imtis	-1.386e-03	3.452e-03	64.28	80.00



38. pav. 6-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

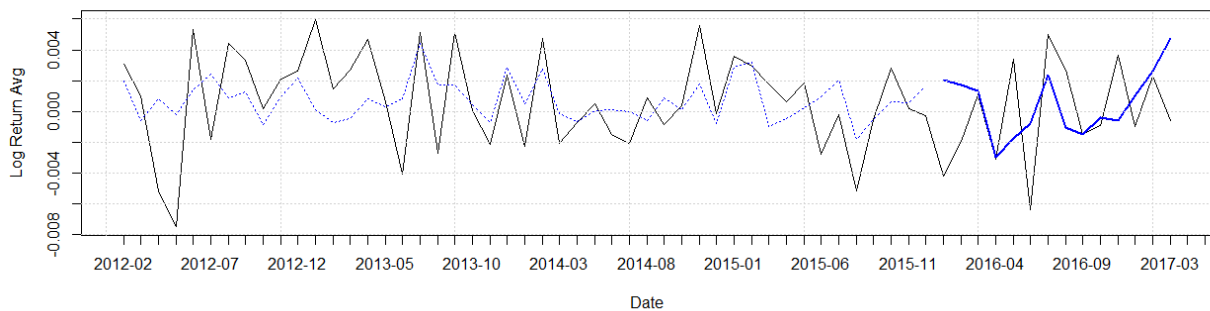
11 PRIEDAS. 7-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

16. lentelė 7-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

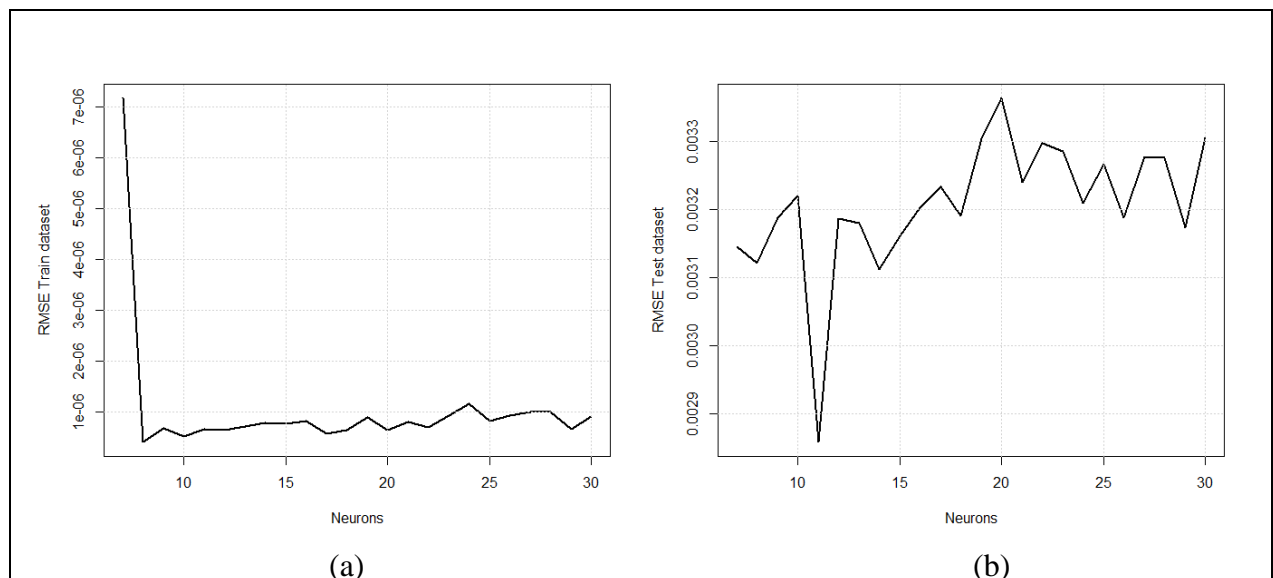
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.6129	0.1004	6.104	2.37E-07	***
conflict_hits_lag1	-0.3234	0.1772	-1.825	0.07477	.
earnings_hits_lag1	0.3617	0.1321	2.738	0.00889	**

17. lentelė 7-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	6.943e-19	0.00277	67.39	61.70
Testavimo imtis	-0.000581	0.00360	53.84	50.00



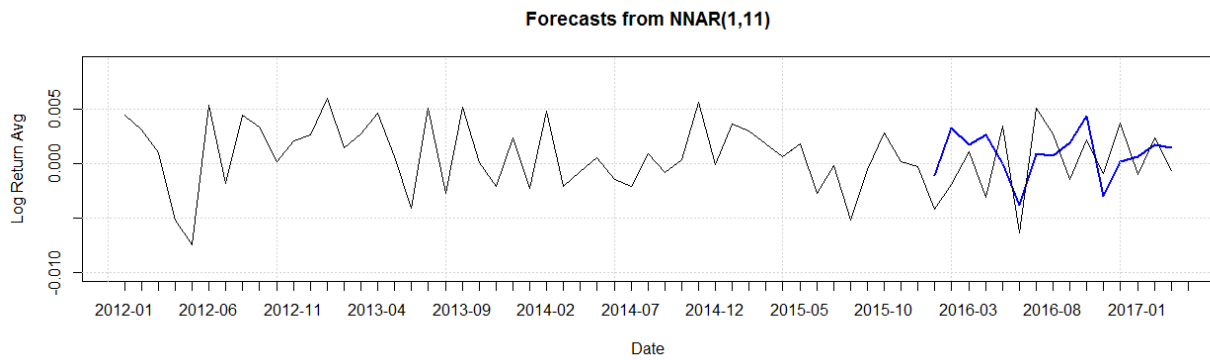
39. pav. 7-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



40. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

18. lentelė 7-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	1.597e-08	5.514e-07	100.00	100.00
Testavimo imtis	-7.101e-04	3.137e-03	64.28	66.66



41. pav. 7-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

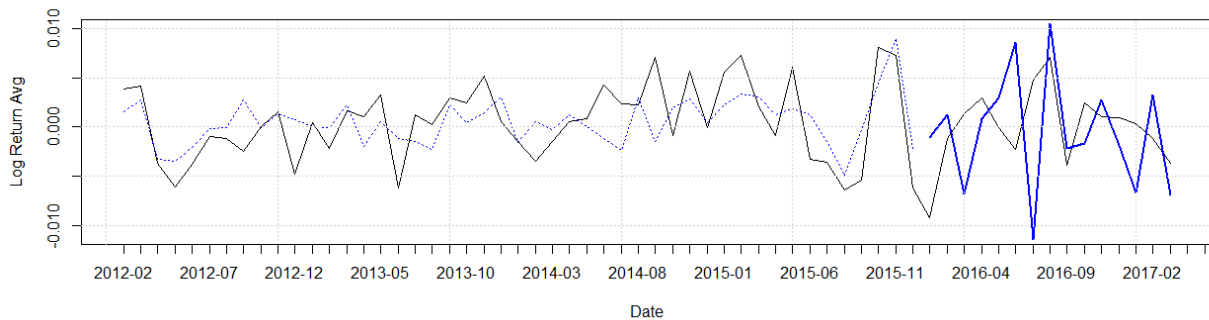
12 PRIEDAS. 8-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

19. lentelė 8-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

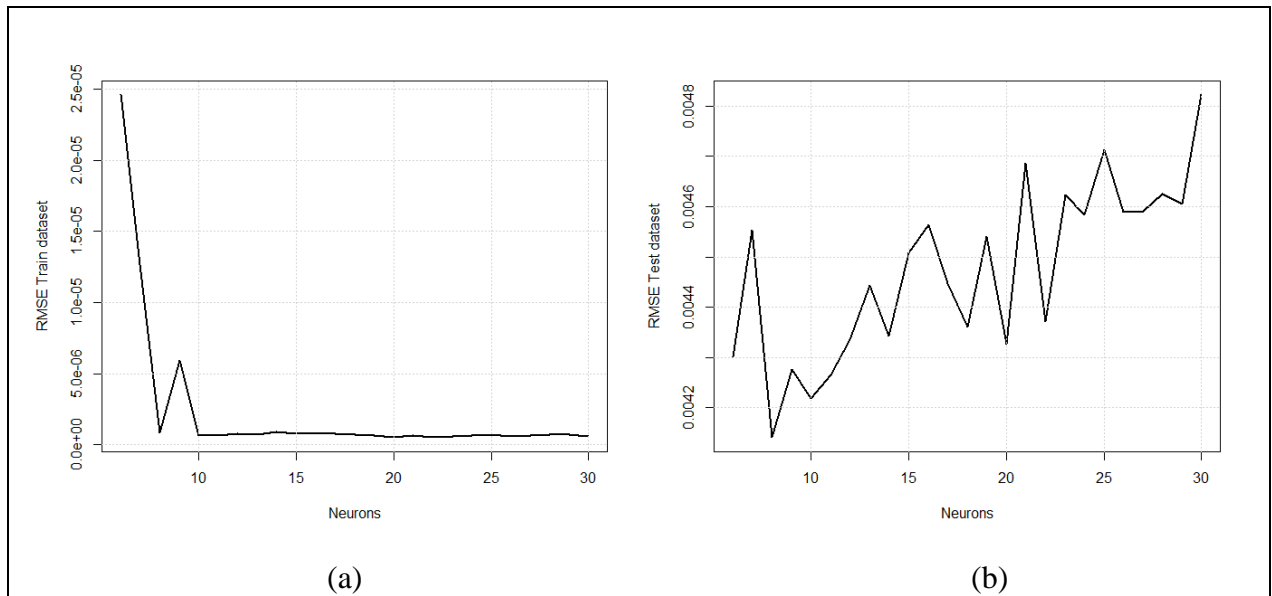
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.2712	0.333	0.815	0.42056	
hits_lag1	0.4712	0.3054	1.543	0.13133	
earnings_hits_lag1	-0.2635	0.1741	-1.514	0.13864	
inflation_hits_lag1	0.3876	0.1658	2.338	0.02489	*
money_hits_lag1	0.3866	0.2946	1.312	0.19748	
INFPP_lag1	-0.4946	0.1928	-2.565	0.01451	*
IPG_PP_lag1	-0.2689	0.1757	-1.531	0.13436	
IR3TIB_lag1	-1.2674	0.4422	-2.866	0.00682	**
UNEMP_R_lag1	0.9545	0.3859	2.473	0.0181	*
ReturnAvg_lag1	0.2478	0.1499	1.653	0.10678	

20. lentelė 8-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	8.92538e-20	0.003111042	69.56	70.21
Testavimo imtis	0.000539327	0.006565923	57.14	40



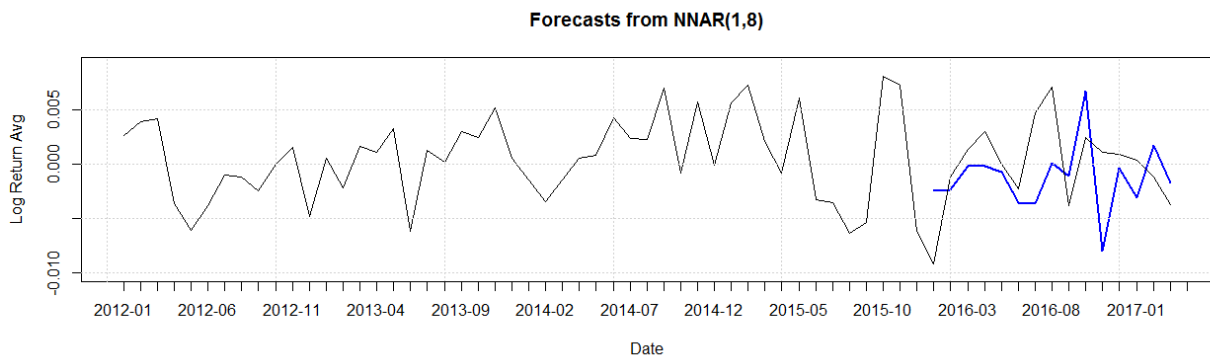
42. pav. 8-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištisinė – testavimo imtis).



43. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

21. lentelė 8-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	8.385e-08	2.082e-06	100.00	100.00
Testavimo imtis	1.242e-03	4.570e-03	78.57	53.33



44. pav. 8-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

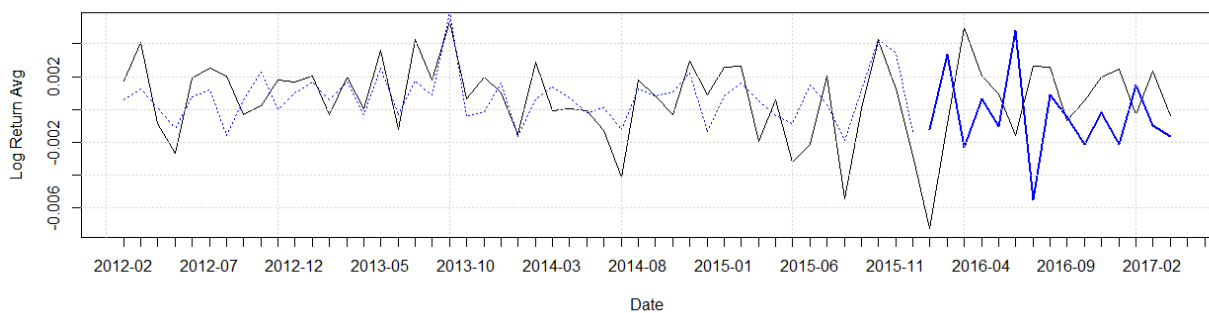
13 PRIEDAS. 9-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

22. lentelė 9-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

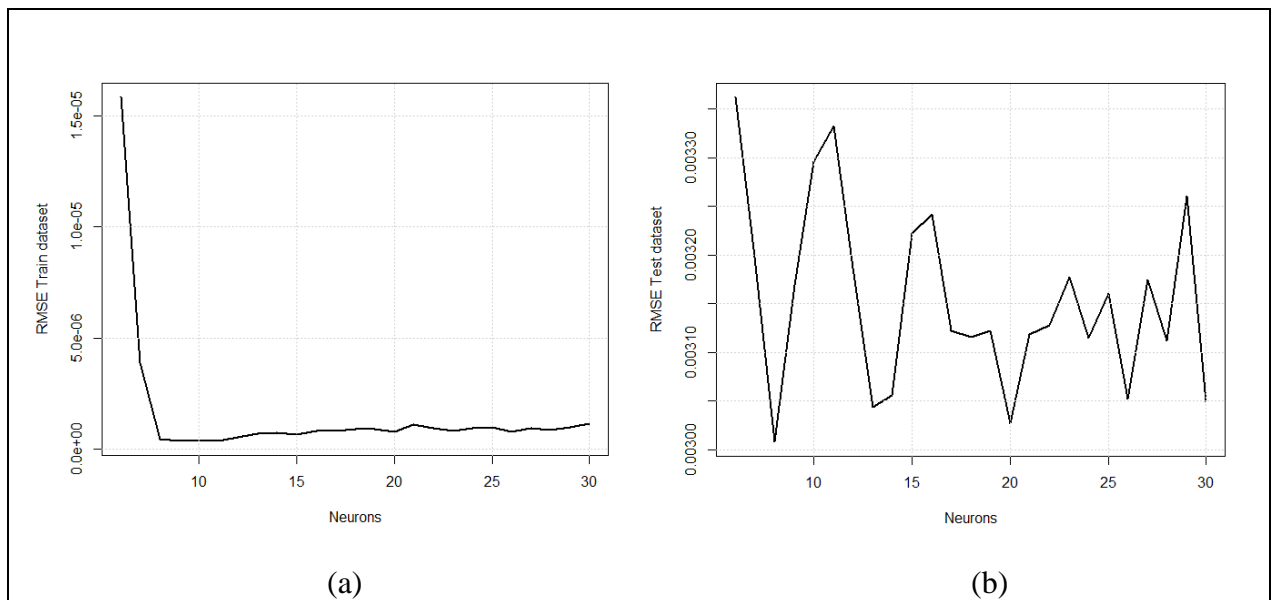
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.4801	0.2193	2.19	0.03494	*
chance_hits_lag1	0.2864	0.21	1.364	0.18084	
conflict_hits_lag1	-0.2576	0.1584	-1.627	0.11226	
debt_hits_lag1	0.4646	0.1846	2.517	0.0163	*
earnings_hits_lag1	-0.1961	0.1314	-1.492	0.14423	
INFPP_lag1	-0.3244	0.1378	-2.354	0.02401	*
IPG_PP_lag1	-0.3389	0.1263	-2.684	0.01081	*
IR3TIB_lag1	-0.5051	0.2464	-2.05	0.0475	*
SHARE_PRICES_lag1	0.3614	0.2049	1.764	0.08606	.
UNEMP_R_lag1	0.6889	0.2136	3.225	0.00263	**

23. lentelė 9-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-1.406514e-18	0.001682747	78.26	68.08
Testavimo imtis	0.001048483	0.004269148	35.71	33.33



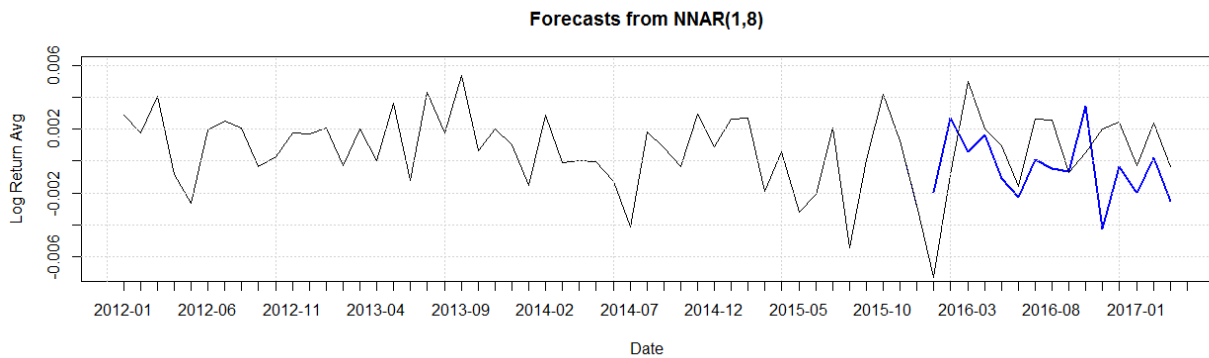
45. pav. 9-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



46. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

24. lentelė 9-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-8.221e-08	5.979e-07	100.00	100.00
Testavimo imtis	1.095e-03	3.150e-03	78.57	66.67



47. pav. 9-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

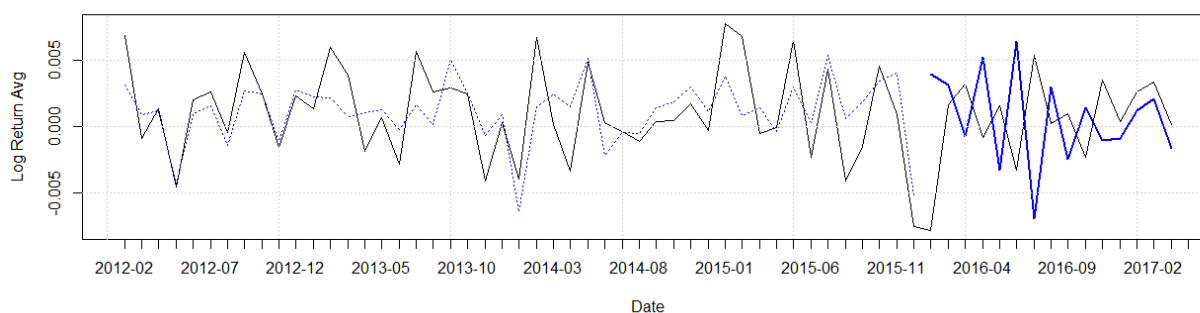
14 PRIEDAS. 10-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

25. lentelė 10-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

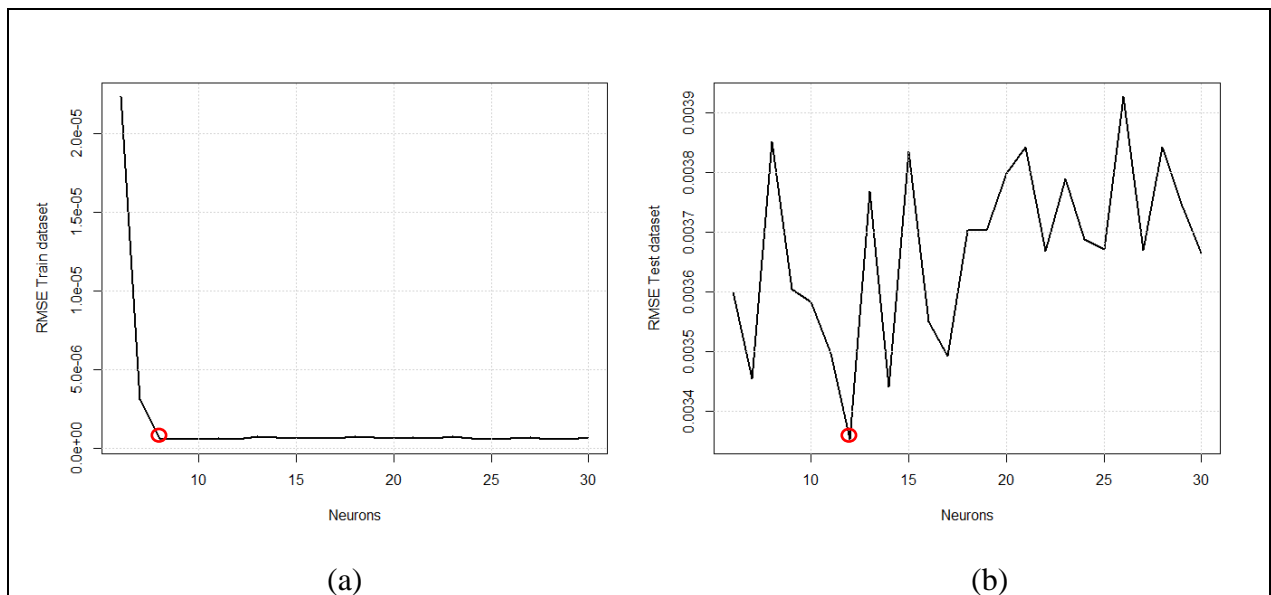
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.1618	0.3256	0.497	0.622073	
conflict_hits_lag1	-0.408	0.2119	-1.926	0.06164	.
debt_hits_lag1	0.3891	0.2054	1.894	0.065816	.
earnings_hits_lag1	-0.2629	0.1438	-1.828	0.075366	.
inflation_hits_lag1	0.5183	0.1426	3.635	0.000821	***
CPI_lag1	0.7714	0.3429	2.249	0.030371	*
INFPP_lag1	-0.2052	0.1475	-1.392	0.172112	
IPG_PP_lag1	-0.5086	0.1332	-3.819	0.000481	***
UNEMP_R_lag1	0.4317	0.2385	1.81	0.078215	.

26. lentelė 10-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-5.406e-18	0.00258	84.78	80.85
Testavimo imtis	-5.787e-05	0.00595	21.42	26.66



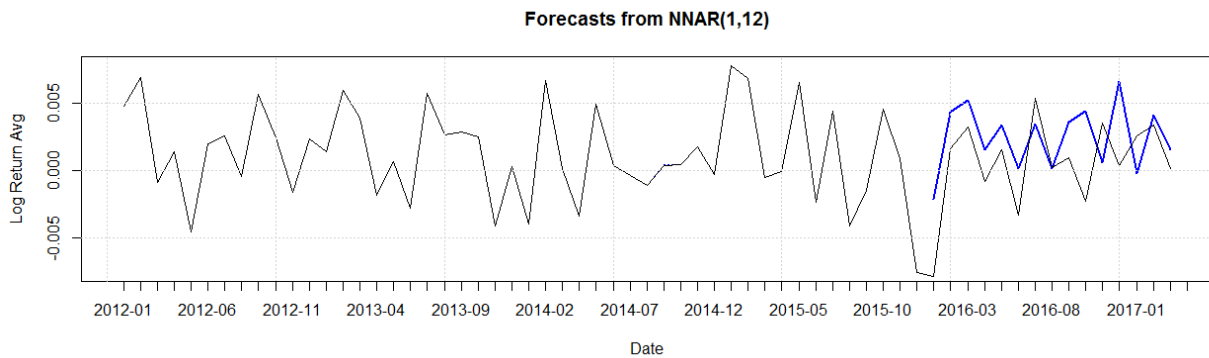
48. pav. 10-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištisinė – testavimo imtis).



49. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

27. lentelė 10-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	3.437e-08	6.916e-07	100.00	100.00
Testavimo imtis	-1.873e-03	3.464e-03	71.42	73.33



50. pav. 10-ojo klasterio tikrosios akcijos grąžos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

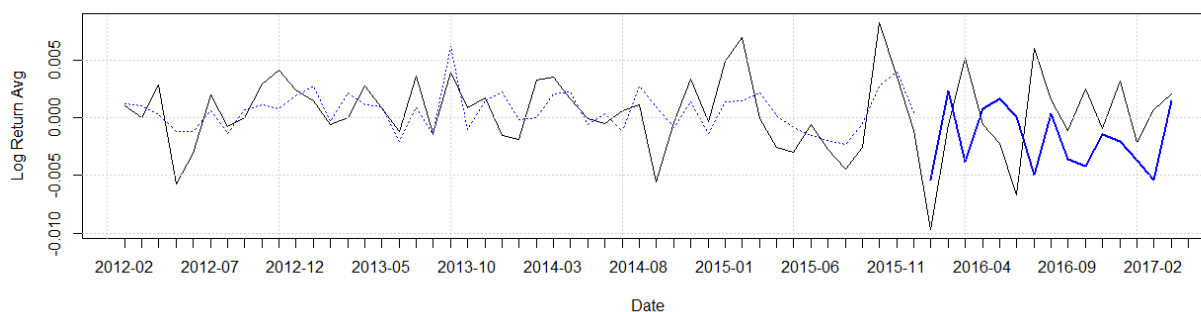
15 PRIEDAS. 11-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

28. lentelė 11-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

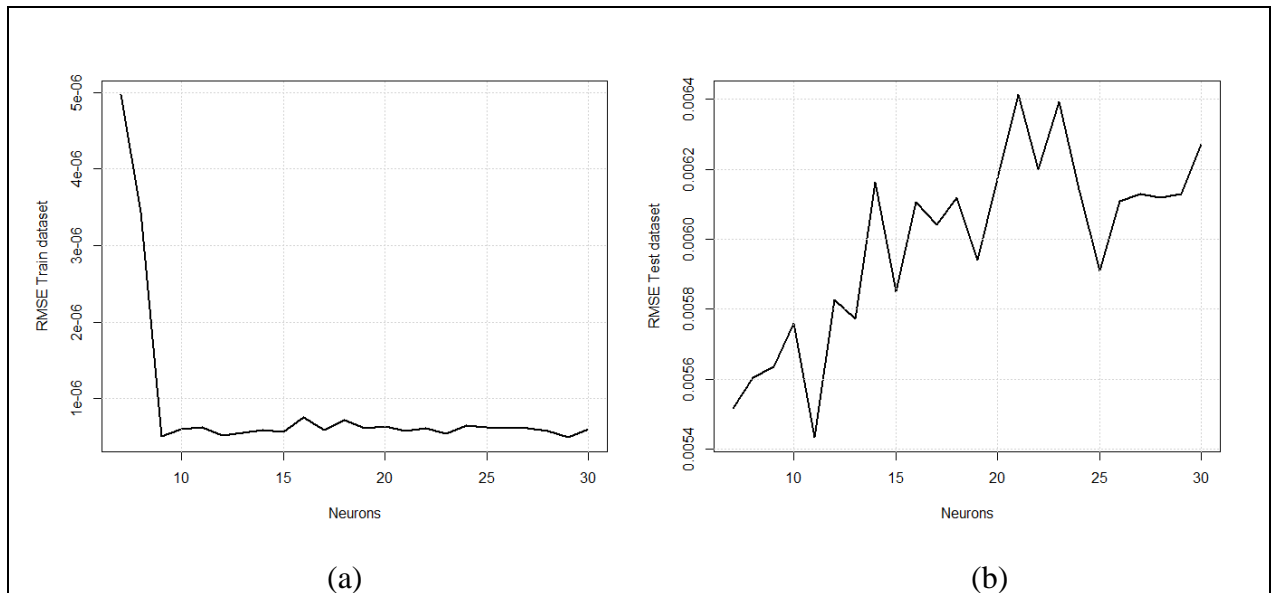
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.2626	0.1642	1.599	0.11783	
chance_hits_lag1	0.542	0.1983	2.732	0.0094	**
debt_hits_lag1	0.2778	0.1796	1.547	0.13002	
earnings_hits_lag1	-0.2035	0.1431	-1.422	0.16308	
INFPP_lag1	-0.3581	0.1456	-2.46	0.01842	*
IR3TIB_lag1	-0.487	0.2423	-2.01	0.05142	.
UNEMP_R_lag1	0.549	0.1929	2.846	0.00702	**
ReturnAvg_lag1	0.2048	0.1298	1.578	0.1227	

29. lentelė 11-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	3.610e-19	0.00236	69.56	80.85
Testavimo imtis	0.00167	0.00523	35.71	40.00



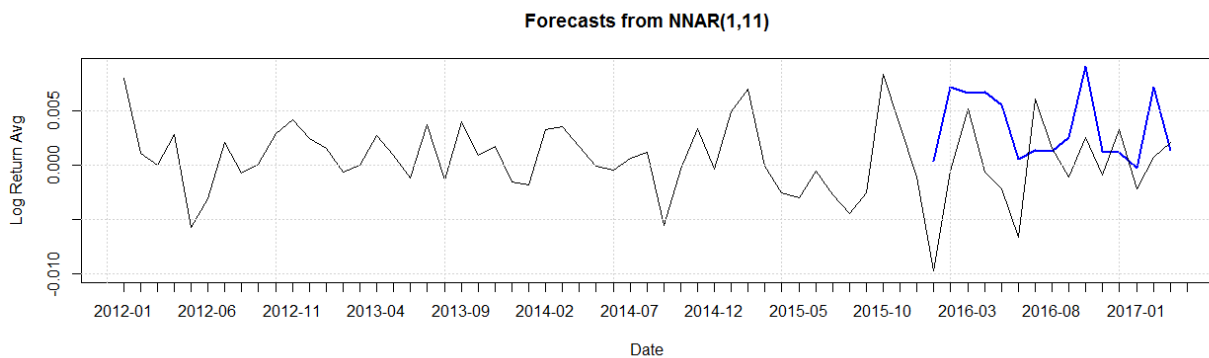
51. pav. 11-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



52. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

30. lentelė 11-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	1.221583e-08	5.476640e-07	100.00	100.00
Testavimo imtis	-4.021618e-03	5.819571e-03	64.28	46.67



53. pav. 11-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

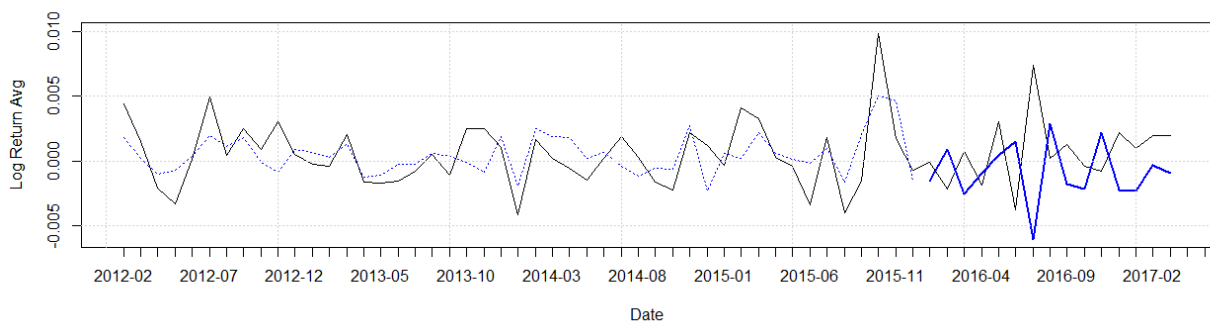
16 PRIEDAS. 12-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

31. lentelė 12-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

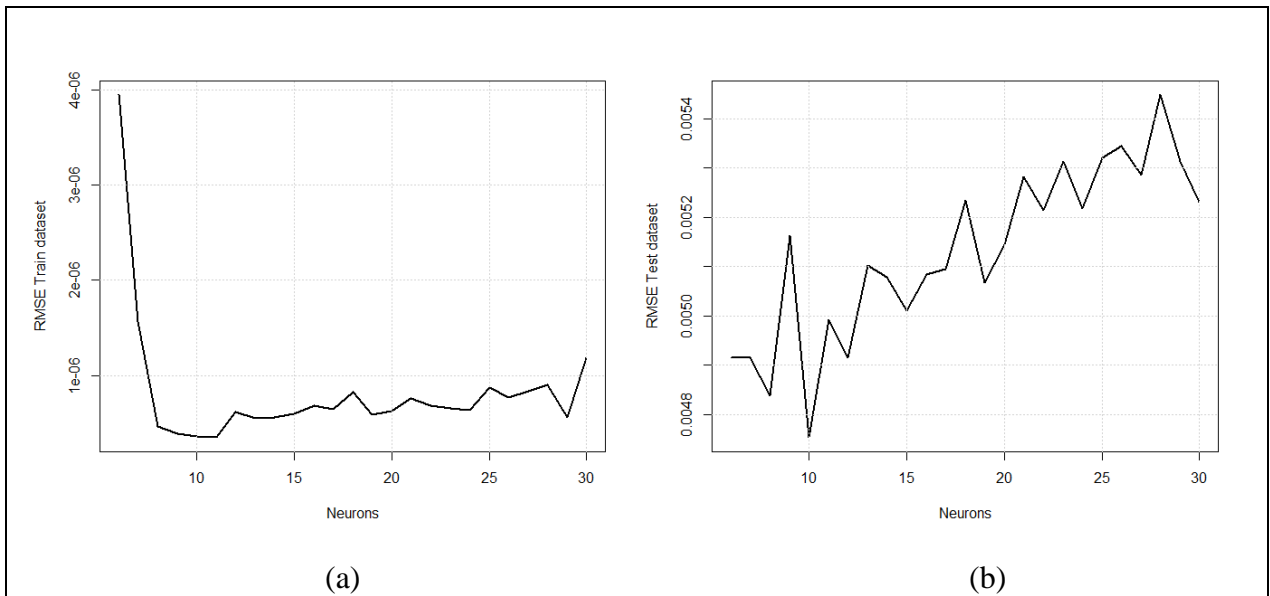
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	4.68E-01	0.1199	3.901	0.000367	***
hits_lag1	-1.97E-01	0.1388	-1.421	0.163272	
conflict_hits_lag1	-3.46E-01	0.1604	-2.154	0.037455	*
inflation_hits_lag1	0.3533	0.1175	3.007	0.004597	**
INFPP_lag1	-0.1874	0.122	-1.536	0.132676	
IPG_PP_lag1	-0.1551	0.1147	-1.353	0.183847	
IR3TIB_lag1	-6.90E-01	2.20E-01	-3.133	0.003278	**
UNEMP_R_lag1	6.02E-01	1.66E-01	3.638	0.000795	***

32. lentelė 12-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-3.852e-18	0.00199	73.91	68.08
Testavimo imtis	0.00155	0.00455	28.57	33.33



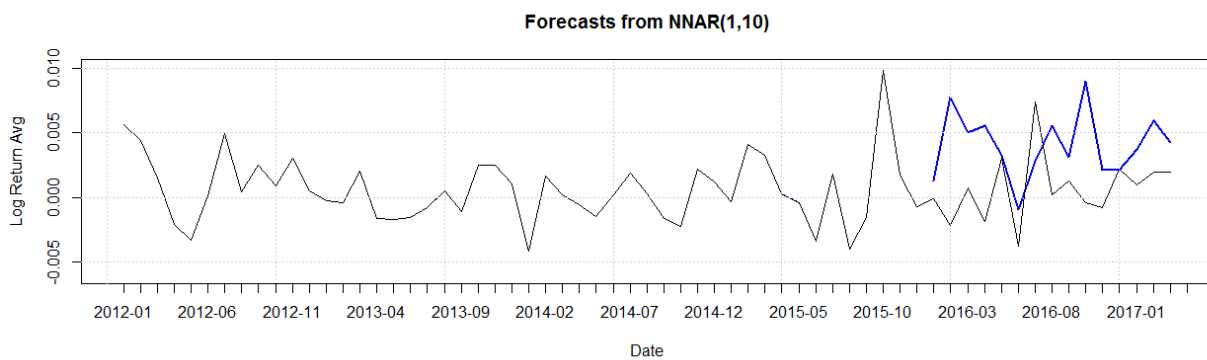
54. pav. 12-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokyimo imtis, mėlyna ištinė – testavimo imtis).



55. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

33. lentelė 12-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-4.422675e-08	4.466122e-07	100.00	100.00
Testavimo imtis	-3.321978e-03	4.891906e-03	35.71	66.66



56. pav. 12-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

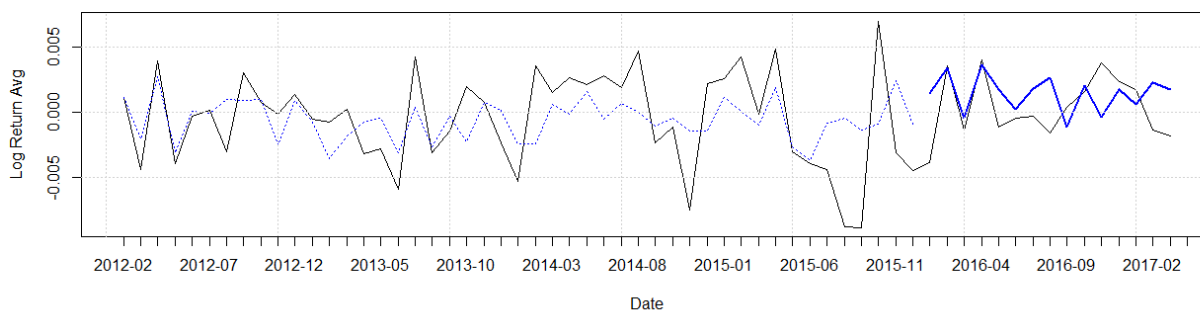
17 PRIEDAS. 13-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

34. lentelė 13-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

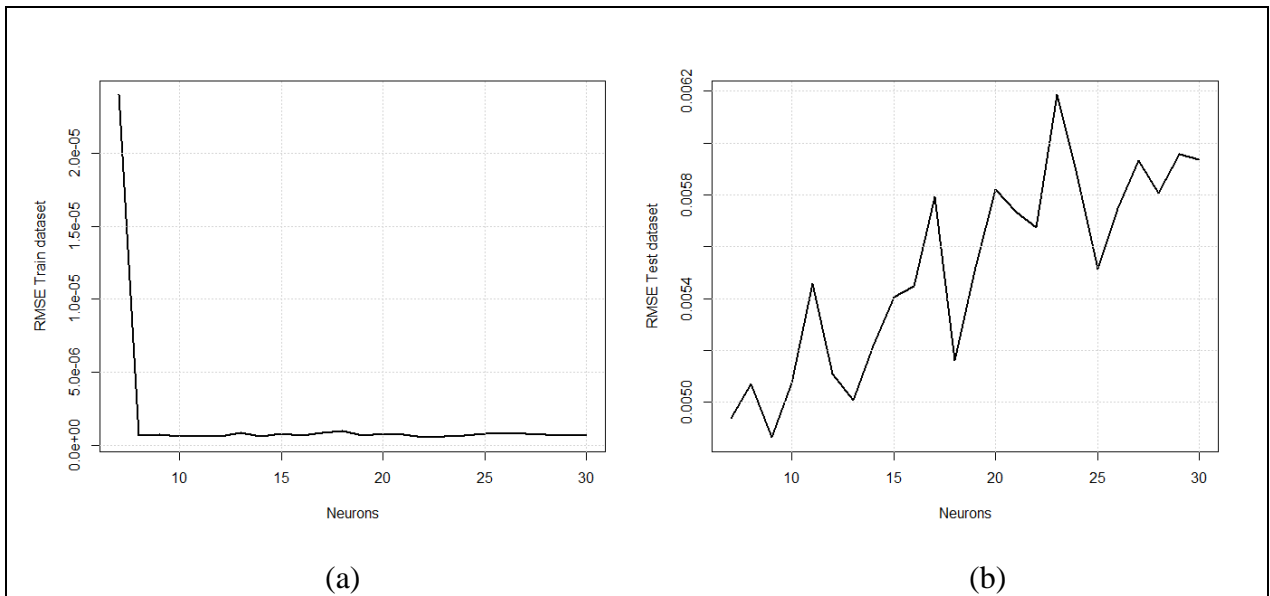
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.2781	0.2638	1.054	0.2979	
inflation_hits_lag1	0.224	0.1392	1.61	0.1152	
Nasdaq_hits_lag1	-0.3416	0.1904	-1.794	0.0802	.
CPI_lag1	0.5528	0.3244	1.704	0.096	.
INFPP_lag1	-0.3749	0.1761	-2.128	0.0394	*
IR3TIB_lag1	0.562	0.3211	1.75	0.0875	.

35. lentelė 13-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	1.858043e-18	0.00332545	63.04	74.46
Testavimo imtis	-0.001050078	0.002679979	57.14	40.00



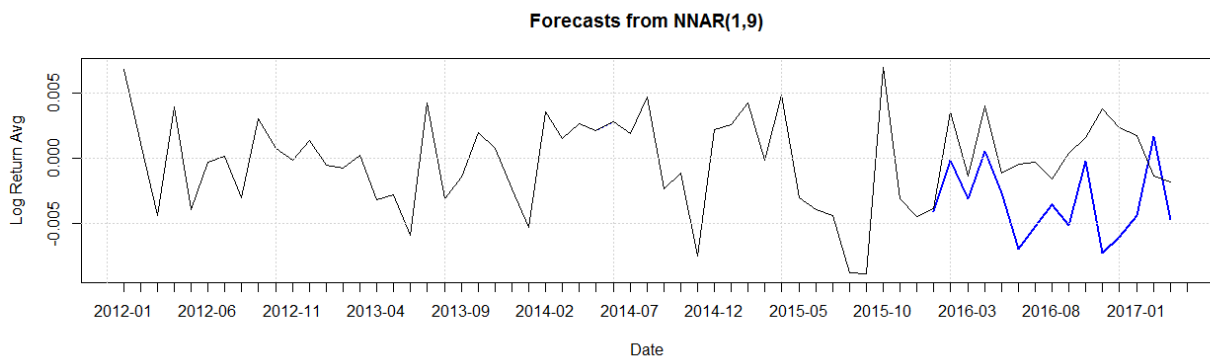
57. pav. 13-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna ištinė – testavimo imtis).



58. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

36. lentelė 13-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	1.054883e-08	7.185915e-07	100.00	100.00
Testavimo imtis	3.801895e-03	5.071386e-03	50	53.33



59. pav. 13-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);

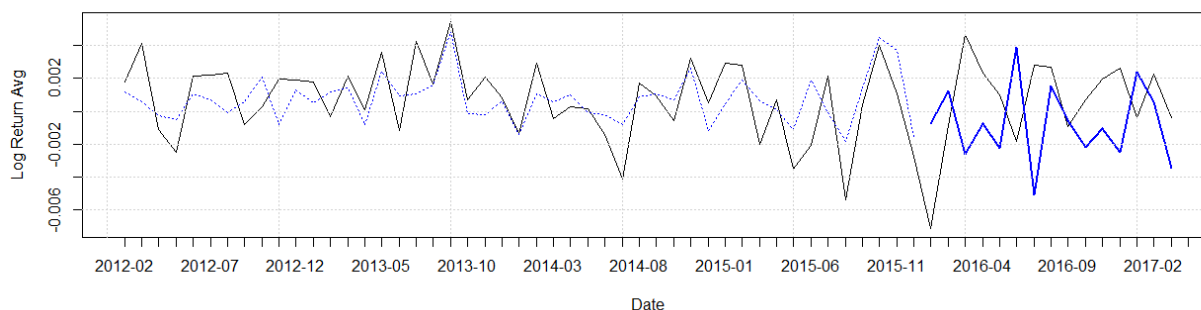
18 PRIEDAS. 14-OJO KLASTERIO AKCIJŲ PROGNOZĖS MODELIŲ TYRIMŲ REZULTATAI

37. lentelė 14-ojo klasterio reikšmingosios akcijos regresijos modelio rezultatai.

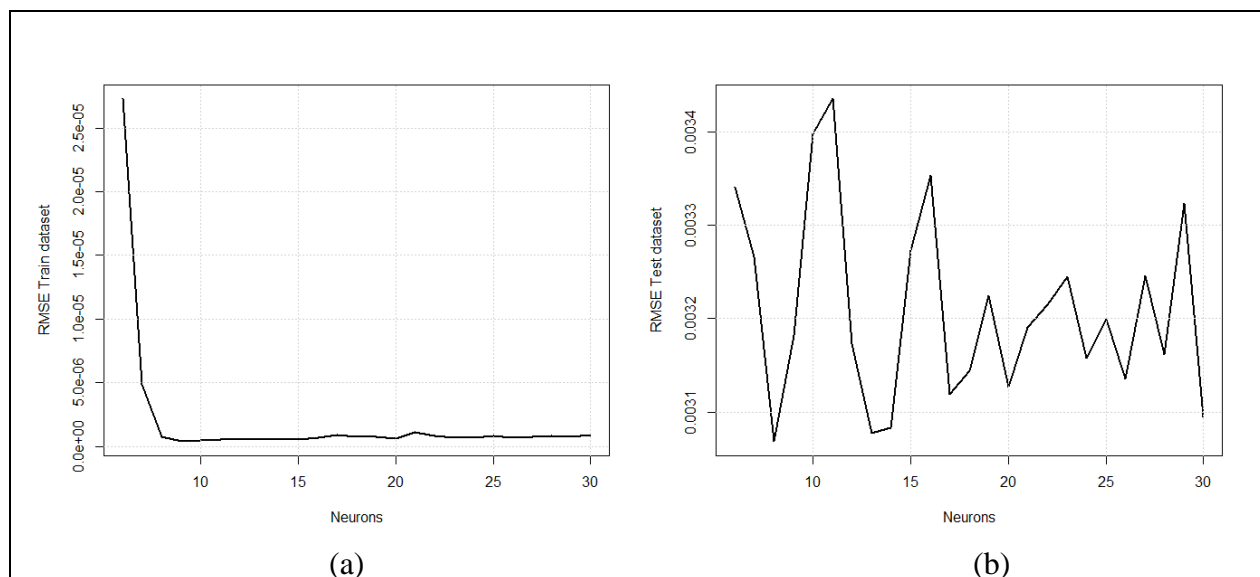
	Koeficientas	Standartinė paklaida	t reikšmė	P(> t)	Reikšmingumo simbolis
(Intercept)	0.4099	0.2192	1.87	0.06902	.
conflict_hits_lag1	-0.2207	0.1615	-1.366	0.17963	
debt_hits_lag1	0.3838	0.1705	2.251	0.03009	*
INFPP_lag1	-0.2462	0.1383	-1.78	0.08282	.
IPG_PP_lag1	-0.3519	0.1301	-2.706	0.01006	*
IR3TIB_lag1	-0.4546	0.2486	-1.829	0.07508	.
SHARE_PRICES_lag1	0.4409	0.2006	2.198	0.03396	*
UNEMP_R_lag1	0.6387	0.2006	3.184	0.00285	**

38. lentelė 14-ojo klasterio reikšmingosios akcijos regresijos modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-4.080212e-18	0.001814499	69.56	68.08
Testavimo imtis	0.001412017	0.004314481	35.71	33.33



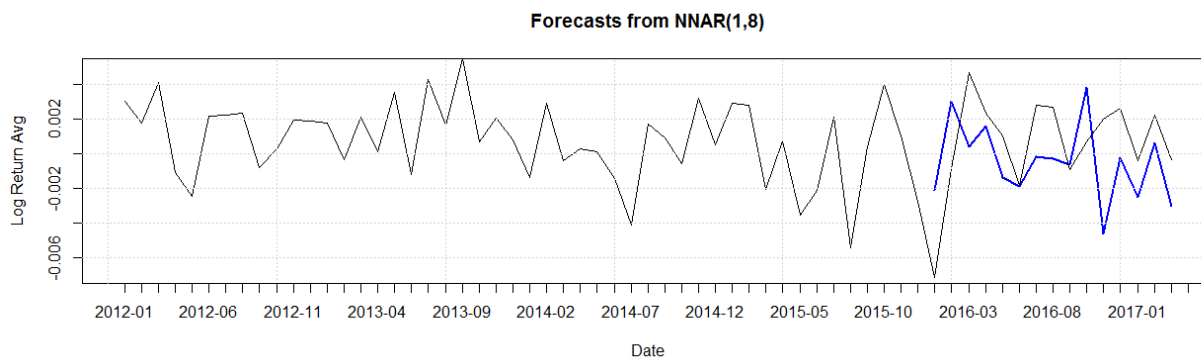
60. pav. 14-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos regresijos modeliu (mėlyna punktyrinė – apmokymo imtis, mėlyna išsistinė – testavimo imtis).



61. pav. *RMSE* paklaidos kitimas neuronų skaičiaus atžvilgiu; (a) dalyje – apmokymo imties; (b) dalyje – testavimo imties

39. lentelė 14-ojo klasterio reikšmingosios akcijos neuroninio tinklo modelio kokybės įvertinimo rezultatai.

	ME	RMSE	POCID, %	POS, %
Apmokymo imtis	-8.154507e-08	4.578426e-07	100.00	100.00
Testavimo imtis	1.127114e-03	3.250981e-03	78.57	60.00



62. pav. 14-ojo klasterio tikrosios akcijos gražos kitimas (juoda linija) ir prognozuotos NARX modeliu testavimo imties (mėlyna);