Kauno technologijos universitetas

Informatikos fakultetas

AKCIJŲ RINKOS DINAMIKOS TYRIMAS IR MODELIAVIMAS

Baigiamasis magistro studijų projektas

M. Tamašauskas

Projekto autorius

Doc. dr. V. Raudonis

Vadovas

Kaunas, 2019

Kauno technologijos universitetas

Informatikos fakultetas

AKCIJŲ RINKOS DINAMIKOS TYRIMAS IR MODELIAVIMAS

Baigiamasis magistro studijų projektas

Informatika (6211BX007)

M. Tamašauskas

Projekto autorius

Doc. dr. V. Raudonis

Vadovas

Doc. dr. A. Misevičius

Recenzentas

Kaunas, 2019

Kauno technologijos universitetas

Informatikos fakultetas

Modestas Tamašauskas

AKCIJŲ RINKOS DINAMIKOS TYRIMAS IR MODELIAVIMAS

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Modesto Tamašausko, baigiamasis projektas tema „Akcijų rinkos dinamikos tyrimas ir modeliavimas“ 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)

Modestas, Tamašauskas. Akcijų rinkos dinamikos tyrimas ir modeliavimas. Magistro baigiamasis projektas / vadovas doc. dr. Vidas Raudonis; Kauno technologijos universitetas, Informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Fiziniai mokslai, informatika.

Reikšminiai žodžiai: *akcijų birža, rinkos indekso prognozavimas, LSTM, statistiniai modeliai.*

Kaunas, 2019. 64 p.

Santrauka

Efektyvios rinkos hipotezė teigia, jog vertybinių popierių kaina beveik realiu laiku atitinka visą informaciją, šie duomenys apima ir istorinius kainų rodmenis. Vadinasi, akcijos kaina tuojau pat sureaguoja į naujai atsiradusią informaciją. Remiantis šia hipoteze, optimali ateities akcijos kainos prognozė yra dabarties kaina. Tačiau praktikoje pastebimos rinkos dinamikos, kurios leidžia dalį akcijos kainos pokyčio sklaidos modeliuoti pasinaudojant istoriniais duomenimis. Pasirinktas tyrimo objektas yra S&P 500 akcijų rinkos indeksas. Tyrimo metu yra modeliuojamas šis procesas siekiant prognozuoti jo pokytį ir įvertinti investavimo į rinką galimybes, pasinaudojant šio proceso prognozėmis. Prognozuojamos indekso dienos vidutinės reikšmė su horizontais nuo 1 iki 4. Tyrimui pasirinktas duomenų intervalas nuo 2017 iki 2019 metų. Darbe naudojami duomenų rinkiniai parsisiųsti iš *Yahoo finance* portalo ir FRED duomenų bazės. Pasirinktos akcijų laiko eilutės yra modeliuojamos 4 skirtingais modeliais: VAR, ARMA+GARCH, SARIMA, ir LSTM. Tyrimo metu nustatyta, jog modeliai skirtingais laiko intervalais geba vienas už kitą padaryti mažesnes paklaidas prognozuojant nuo 1 iki 4 dienų į priekį indekso vertes. Šią struktūrą nutarta modeliuoti, siekiant pagerinti prognozių tikslumą. Šiam tikslui pasiekti sukurti 3 skirtingos modelių ansamblių architektūros. Taip pat atlikta rinkos simuliacija ir įvertintos modelių prognozių panaudojimo galimybės investuojant į rinką, panaudojant tyrimo metu pasiūlytą strategiją.

Tamašauskas, Modestas. Modeling and Research of Stock Market Dynamics. Master ‘s thesis / supervisor assoc. prof. Vidas Raudonis. The Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Physical science, informatics.

Key words: *stock market, stock market index forecasting, LSTM, statistical models*

Kaunas, 2019. 64 p.

Summary

The efficient market hypothesis states that price of securities fully reflects, almost in real-time, all available information, including historical price data. Hence, the stock price immediately responds to new information. Based on this hypothesis, the optimal future stock price prediction is its current price. However, in practice, market dynamics are observed, which allow modeling part of the future price change using historical data. In order to correctly identify and consequently model market dynamics, the study of S&P 500 stock market index is carried. This process is modeled during the study in order to forecast its changes and to evaluate the possibilities of investing in the stock market. Estimations of daily averaged value of the index with horizons 1 to 4 are evaluated. Data range from 2017 to 2019 has been selected for research. Data sets used in the work are downloaded from Yahoo finance and FRED database. The selected time series are modeled using 4 different models: VAR, ARMA+GARCH, SARIMA, and LSTM. In process of this study, observation that distinct models at different time intervals were able to make smaller biases in predicting 1 to 4 day ahead index values. This structure was modeled in order to improve the accuracy of the forecasts. To achieve this, 3 different architectures of model ensembles were created. Market simulations have also been carried, and the use of different model predictions has been assessed by proposed strategy.

Turinys

Lentelių sąrašas

Paveikslų sąrašas

Santrumpų ir terminų sąrašas

Santrumpos:

DI – dirbtinis intelektas (angl. *artificial intelligence*);

RNN – grįžtamojo ryšio neuroniniai tinklai (angl. *recurrent neuron network*);

LSTM – trumpalaikės-ilgalaikės atminties neuronų tinklų modelis, paremtas grįžtamojo ryšio modelių architektūra (angl. *long short term memory*);

ROC – akcijos kainos kitimo dydis (angl. *rate of change*);

MACD – akcijos kainos judėjimo konvergencijos / disertacijos (angl. *moving average convergence / divergence*);

DNT – dirbtiniai neuroniniai tinklai;

IPO – pradinis viešas pasiūlymas (angl. *initial public offering*);

EMH – efektyvios rinkos hipotezė (angl. *efficient market hypothesis*);

VAR – vektorinės autoregresijos modelis;

GARCH – apibendrintas autoregresinio sąlyginio heteroskedastiškumo (angl. *general autoregressive conditional heteroskedastic*) modelis;

ARIMA – autoregresinis integruotas slenkančio vidurkio modelis (angl. *autoregressive integrated moving average*);

AIC – Akaikės informacinis kriterijus (angl. *Akaike information criterion*);

BIC – Bajeso informacinis kriterijus (angl. *Bayesian information criterion*);

MSE – vidutinė kvadratinė paklaida *(angl. mean squared error*);

MAE – vidutinė absoliutinė paklaida (angl. *mean absolute error*).

Įvadas

Projekto naujumas ir aktualumas

Vertybinių popierių birža yra viešoji rinka, kurios tikslas – suteikti priemones ir alpiną, kuriais naudojantis galima prekiauti įmonių akcijomis bei kitais vertybiniais popieriais iš anksto sutartomis kainomis. Veiksmai pasaulinėje biržoje apima didžiulius finansinius sandorius. 2018 metais šios rinkos dydis siekė 65,66 trilijonų JAV dolerių [1]. Akivaizdu, jog galima gauti didelį pelną šioje srityje, jei iš anksto būtų žinoma, kas atsitiks rinkoje ateityje. Efektyvios rinkos hipotezė (angl. *efficient market hypothesis*, EMH) teigia, jog akcijų kaina yra funkcijos reikšmė, kurios argumentai yra visa tuo metu egzistuojanti informacija, kuri apima ir istorinius akcijų kainų rodmenis. Vadinasi, akcijos kaina tiesiogiai susijusi ir todėl reaguoja tik į naujai atsiradusią informaciją. Remiantis šia hipoteze, optimali ateities akcijos kainos prognozė yra dabarties kaina. Šiame darbe yra apžvelgiami statistiniai bei DI (dirbtinio intelekto) metodai, leidžiantys modeliuoti akcijų kainų proceso savybes, kurių neaprėpia stipri EMH. Laiko eilučių analizės sferoje yra sukurti metodai, kurie geba modeliuoti tokias struktūras kaip procesų tarpusavio autokoreliacija, sąlyginė sklaida bei impulsų poveikis procesui. Tačiau empiriniai tyrimai rodo, jog finansinių duomenų kitimo procesas yra sunkiai nuspėjamas ir nauji rinkos statistiniai modeliai dažnai negeba pralenkti atsitiktinio klaidžiojimo metodo (angl. Random Walk, kurio optimali ateities prognozė yra dabartinė kaina) savo prognozėmis [2].

Egzistuoja keturi akcijų biržos prognozavimo metodai. Pirmasis yra vadinamas technine analize, kuri apima istorinių akcijos kainų kitimų istorijos analizę, siekiant identifikuoti pasikartotinus šablonus, kuriuos identifikavus galima nuspėti akcijos pobūdį ateityje. Šis metodas dažnai remiasi ekspertų duomenimis. Šios analizės populiarūs subjektai yra akcijos kainos kitimo dydis (ROC), kitimo vidurkis, konvergencijos / disertacijos (MACD) bei tendencingumas. Antrasis metodas – fundamentalioji analizė. Ji analizuoja ryšius tarp finansinių įmonės rodiklių bei kitų aplinkybių, tokių kaip gauto ketvirčio pelno paskelbimas įmonės viduje ir kitų faktorių. Trečiasis metodas – laiko eilučių analizė. Šioje tyrimų šakoje (finansinių procesų prognozavimo) akcijų kainos prognozavimas yra traktuojamas kaip vienas iš sudėtingiausių taikymo sričių. ARH, GARCH, ARMA, AR bei kiti tradiciniai metodai nesugeba pakankamai tiksliai prognozuoti šiuos procesus, siekiant finansinės naudos. XXI a. pradžioje pradėti taikyti sudėtingesnė nelinijinių metodų grupę (ketvirta metodų grupė). Į šią grupę įeina šie metodai: atraminių vektorių klasifikatorius, atraminių vektorių regresija (SVR) bei dirbtiniai neuroniniai tinklai (DNT). DNT šioje srityje naudojami, nes geba mokytis iš nelinijinių duomenų kitimo tendencijų, tai šioje srityje yra labai aktualu.

Šio darbo metu yra siekiama sujungti statistinių bei DNT modelių savybes, siekiant pagerinti bendrą modelių prognozių tikslumą bei stabilumą. Šiam tikslui pasiekti yra atliekama išsami literatūros analizė, apimanti akcijų kainų prognozavimą. Identifikuotos geriausios praktikos, algoritmai bei metodai. Pasirinktas tyrimo objektas S&P akcijų rinkos indeksas. Rinkos indeksas pasirinktas, nes jis atspindi visos rinkos dinamikas, o ne pavienės akcijos. Pasirinkti akcijų bei tiriamo rinkos indekso istoriniai duomenys iš *Yahoo finance* portalo [3] bei kiti tyrime naudojami duomenų paketai iš FRED [4] duomenų bazės. Šios finansinių duomenų laiko eilutės yra modeliuojamos keturiais skirtingais modeliais: VAR, ARMA+GARCH, SARIMA ir DNT modeliu LSTM (angl. *Long Short Term Memory*). Ištyrus minėtus modelius, pasiūlyti bei sukurti trys hibridinės dirbtinio intelekto bei statistinių modelių sistemos. Palygintos visų minėtų rinkos modelių ateities prognozavimo bei

prognozių panaudojimo investavimui į rinką savybės. Tam sukurta rinkos simuliacija bei strategija, kuri panaudodama sukurtų modelių prognozes atlieka akcijų pirkimus ir pardavimus.

Tikslas ir uždaviniai

Tyrimo tikslas – atlikti pelningas investicijas į akcijų rinką pasinaudojant DNT bei statistinių modelių rinkos ateities prognozėmis:

- nustatyti rinkos dinamikas ir joms projektuoti tinkamus metodus;
- sukurti statistinių bei DNT ansamblius, siekiant pagerinti sukurtų rinkos modelių prognozių charakteristikas;
- nustatyti, ar modeliai gali sukurti vienos dienos į priekį prognozę likus 1 minutei iki biržos uždarymo;
- įvertinti sukurtų modelių tiriamo proceso pokyčio krypties prognozavimo charakteristikas;
- įvertinti sukurtų prognozių tinkamumą atliekant investicijas į rinką;
- sukurti ir ištestuoti strategiją, kuri panaudodama sukurtų modelių rinkos prognozes atlieka pelningas investicijas į rinką;
- nustatyti, ar metrikos, kuriomis yra vertinamos modelių prognozės, atspindi simuliacijos rezultatus.

Darbo struktūra

Darbą sudaro trys dalys. Pirmoje dalyje apžvelgiama vertybinių popierių birža, bei jos dalyviai, jai prognozuoti taikytini metodai, aptariami ankstesni tyrimai bei jų pasiekimai. Antroje šio darbo dalyje aprašomas tyrimo procesas ir pasiūlytos modelių ansamblių architektūros. Šiame skyriuje taip pat pateikiamas investavimo strategijos aprašas, kuriame vertinamos sukurtų modelių prognozių panaudojimo investuojant į rinką galimybės. Trečioje šio darbo dalyje aptariami pasirinkti duomenys, pateikiamos sukurtų modelių realizacijos, jų suprojektuotų prognozių charakteristikos, aptariami atliktų investavimų į rinką simuliacijų tyrimo rezultatai.

1 AKCIJŲ MODELIAVIMO LITERATŪROS ANALIZĖ

Šiame skyriuje apžvelgiama akcijų rinka, jos analizavimo metodai bei naudingi rodikliai, indeksai, taikytini prognozuojant ateities akcijų kainas. Taip pat pateikiami statistiniai bei DI (dirbinio intelekto) metodai, tinkantys modeliuoti finansines laiko eilutes. Skyriaus pabaigoje yra aptariami šios srities kitų mokslininkų darbai bei pasiekti rezultatai.

1.1 Vertybinių popierių birža

Vertybinių popierių birža yra korporacija ar neutrali organizacija, kuri teikia „prekybos" aplinką brokeriams ir prekybininkams, taip sukurdama rinkos erdvę (virtualią ar fizikinę [5]). Įmonės, siekdamos sugeneruoti pelną, gali parduoti įmonės dalį kaip akcijas. Akcijų birža suteikia platformą minėtoms įmonėms. Rinkos dalyviai yra tiek juridiniai asmenys, tiek didelės kompanijos. Nors biržos suteikia aplinką, kurioje galima lengvai ir greitai atlikti finansinius sprendimus, jos trūkumas yra tai, jog ji nepasižymi skaidrumu. Sunku nustatyti, kokia yra geriausia kaina tam tikru laiko momentu bei kaip ji kis ateityje. Didžiausios bei populiariausios rinkos pateiktos 1.1 lent. Lentelėje pateiktuose duomenyse pastebima, jog didžiausia ir labiausiai aktyvi akcijų birža yra *New York Stock Exchange*, o antroje vietoje NASDAQ, abi rinkos apima prekyba akcijomis Amerikoje įkurtų įmonių.

1.1 lentelė *The World Federation of Exchanges* pateikia didžiausių pasaulio vertybinių popierių biržų penketuką [6].

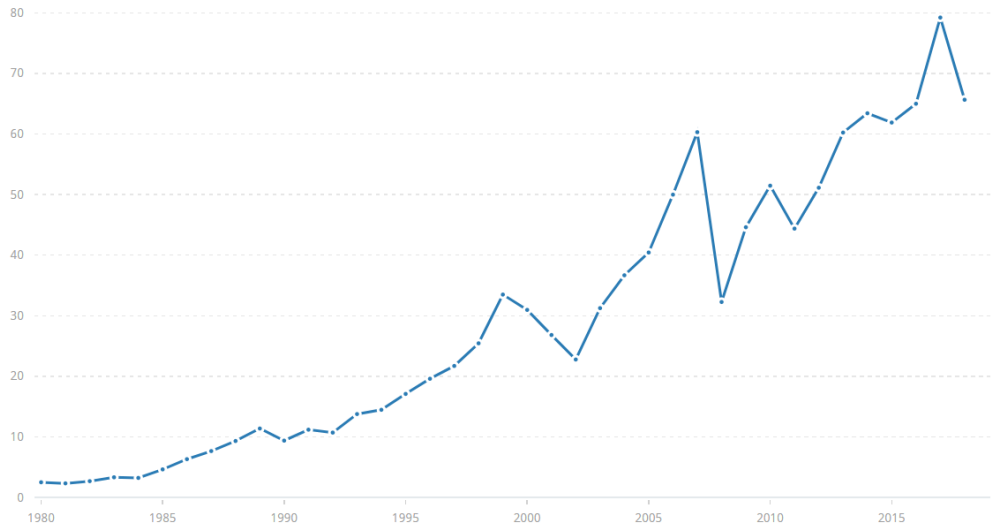
Pavadinimas	Rinka	Pagrindinis padalinys	Rinkos riba (USD milijardai)	Mėnesio prekybos vertė (USD milijardai)
New York Stock Exchange	JAV	Niujorkas	21,377	1,781
NASDAQ	JAV	Niujorkas	9,585	948
Japan Exchange Group	Japonija	Tokijas	5,974	536
Shanghai Stock Exchange	Kinija	Šanchajus	5,043	517
Euronext	Europos Sąjunga	Amsterdamas, Briuselis, Lisabona, Londonas, Paryžius	4,388	167

Yra trys rūšys vertybinių popierių, kuriomis dažniausiai prekiaujama, – akcijos, obligacijos ir investiciniai fondai. Akcijos yra dokumentai, kuriuos išdavė bendrovė, suteikiantys jų savininkui teisę būti vienu iš kompanijos savininkų. Šie dokumentai yra tiesiogiai išduodami bendrovės per pradinį viešą pasiūlymą (IPO) arba juos galima įsigyti iš akcijų biržos. Jų savininkas, turėdamas dalį kompanijos akcijų, gali uždirbti dalį bendrovės pelno, vadinamo dividendais. Be to, jų savininkas taip pat gali pasipelnyti iš vertybinių popierių juos pirkdamas ir parduodamas ir taip gaudamas pelną. Obligacijos tai skolos, vertybinis popierius, kuris įpareigoja jos turėtoją, juridinį asmenį, kompaniją ar net valstybę mokėti palūkanas jų savininkui. Palūkanos paprastai mokomos intervalais (kas pusmetį ar trumpesnį periodą). Investiciniai fondai – tai kompanijos, kurios siekia pelno investuodamos į kitus vertybinius popierius. Investicinius fondus galima skaidyti į dvi grupes: ETFs (angl. *exchange traded funds*) ir MF (angl. *mutual funds*). ETFs yra investicinės įmonės, išleidžiančios vertybinius popierius, kuriais yra prekiaujama viešoje rinkoje. Dauguma ETFs teisiškai struktūrizuoti kaip atviros

investicinės bendrovės, didelė dalis jų siekia vertybinių popierių verte „pririšti“ prie populiarių rinkos indekso reikšmių. Skirtingai nuo MF, kurie tik leidžia investuotojams įgyti ar parduoti jų akcijas prekybos dienos pabaigoje, ETF leidžia investuotojams prekiauti jų akcijomis visą prekybos dieną [7].

Pasyvios investicijos į investavimo fondus sparčiau didėja nei aktyvios strategijos, o ETFs yra šio trendo priekyje. Tarp rinkos dalyvių plinta požiūris, kad atskaičiavus sąnaudas labai mažai aktyvių strategijų nuosekliai pranoksta investicijas, paremtas rinkos indekso analize. 2017 m. atliktoje rinkos analizėje prognozuota, jog pasyvios investicijos turėtų pralėkti, rinkos kapitalizacija 2027m. aktyvės. Ši prognoze paremta 10 metų tendencija. [8]

Veiksmai pasaulinėje biržoje apima didžiulius finansinius sandorius. 2018 metais šios rinkos dydis siekė 65.66 trilijonų JAV dolerių [1]. Rinkos kapitalizacijos grafikas yra pateiktas 1.1 pav. Grafike matyti, kaip pasaulinės rinkos dydis kito nuo 1980 m. iki 2018 m. Pastebima akivaizdi didėjimo tendencija, didesni nuokrypiai nuo augimo tendencijos, žinomos kaip ekonominės krizės.



1.1 pav. Pasaulinės rinkos kapitalizacija [1]

Siekiant įvertinti akcijos kokybę, yra įvesta indekso koncepcija. Indeksas yra statistinė sudėtinė metrika, parodanti kainų judėjimą bendrojoje rinkoje arba pramonėje. Iš esmės indeksai leidžia matuoti įmonių grupės veiklos rezultatus tam tikru laiko tarpu. Įmonės yra suskirstytos į indeksą pagal du pagrindinius metodus ar svorius. Kainų judėjimas rinkoje ar skyriuje išryškėja iš kainų indekso, kuris yra vadinamas akcijų rinkos indeksu. Populiariausi ir daugiausiai naudojami indeksai yra *NASDAQ Composite*, *Dow Jones Industrial Average* (DJIA) ir S&P 500 [9].

NASDAQ Composite (ženklinimo simbolis ^IXIC) vertybinių popierių indeksas sudarytas iš prekiaujamų akcijų NASDAQ biržoje. *NASDAQ Composite* indekso vertė lemia informacinių technologijų įmones. Šis rinkos rodiklis buvo sukurtas 1971 m., kurio pradinė vertė – 100. Per šio indekso gyvavimo laikotarpį jis smarkiai padidėjo, nepaisant daugelio nuosmukio laikotarpių. Tam, kad kompanijų akcijų kaina būtų įtraukta į šį indeksą, JAV kompanijos vertybiniai popieriai turi būti

įtraukti tik į NASDAQ vertybinių popierių rinką, išskyrus atvejus, kai akcijos buvo įtrauktos į kitą rinką iki 2004 m. [10]

„Dow Jones Industrial Average“ (ženklinimo simbolis ^DJI), arba tiesiog DOW, yra akcijų rinkos indeksas, rodantis 30 didelių viešųjų kompanijų, įkurtų JAV, vertybinių popierių prekybą rinkoje. Šios 30 bendrovių taip pat yra įtrauktos į S&P 500 indeksą. DOW vertė nėra svorinis aritmetinis vidurkis bei neatitinka indekso dedamųjų kompanijų, o tiesiog yra dedamųjų kompanijų akcijų kainų suma. Ši suma yra koreguojama atsižvelgiant į 30 kompanijų veiksnių, tokių kaip akcijų dalijimą ar jų dividendus, siekiant stabilizuoti indekso reikšmę. Šis indeksas yra vienas iš seniausių, pirmą kartą jis buvo paskelbtas 1885 metais. Industrinė šio indekso pavadinimo dalis yra istorinis artefaktas ir nereprezentuoja ar labai silpnai reprezentuoja 30 kompanijų, sudarančių indeksą. Nors DJIA yra projektuojamas siekiant įvertinti pramonės sektoriaus veiklos rezultatus JAV ekonomikoje, indekso veiklą ir lemia ne tik įmonių ekonominiai rodikliai, bet ir vidaus bei užsienio politiniai įvykiai, pavyzdžiui, karas ir terorizmas, taip pat stichinės nelaimės, galinčios sukelti ekonominę žalą. Norėdami apskaičiuoti DJIA, visų 30 kompanijų akcijų kainų suma yra padalijama iš daliklio (*d*):

$$DJIA = \frac{\sum p}{d}; \tag{1}$$

čia p yra akcijos kaina, o d yra *Dow Divisor*. Šis daliklis yra koreguojamas, jei įvyksta akcijų dalijimasis ar kiti struktūriniai pokyčiai, siekiant užtikrinti, kad tokie įvykiai savaime nekeičia DJIA skaičiaus:

$$DJIA = \frac{\sum p_{sena}}{d_{sena}} = \frac{\sum p_{nauja}}{d_{nauja}}. \tag{2}$$

Pradinėse stadijose šis indeksas buvo lygus įmonių skaičiui, indeksas reiškė paprastą akcijų kainų sumų aritmetinį vidurkį, tačiau po daugelio korekcijų yra mažesnis nei 1. O tai reiškia, jog indeksas yra didesnis nei į indeksą įtrauktų akcijų kainų sumos. [11]

S&P 500 (ženklinimo simbolis ^GSPC), arba tiesiog S&P, yra Amerikos akcijų rinkos indeksas, pagrįstas 500 didelių bendrovių kapitalizacijos rinkų, kurių vertybiniais popieriais yra prekiaujama NYSE, NASDAQ arba Cboe BZX rinkose. S&P buvo sukurta ir yra prižiūrima įmonės S&P Global. Šis indeksas pirmą kartą buvo paskelbtas 1923 m., tačiau indekso kompozicija apėmė tik mažą kiekį akcijų ir tik po 34 metų, 1957 m., jis išsiplėtė iki dabartinio 500. S&P sudedamąsias dalis pasirenka komitetas. Indeksas panašus į DJIA, tačiau skiriasi tuo, jog indekso komponentės nėra parenkamos remiantis griežtomis taisyklėmis. Pagrindiniai kriterijai, pagal kuriuos atrenkamos įmonės, yra rinkos kapitalizacija, likvidumas, regionas, kur įmonė vykdo pagrindinę veiklą, viešoji apyvarta, sektoriaus klasifikacija, finansinis gyvybingumas ir viešai prekiaujama trukmė. Kiekvienas iš šių pirminių kriterijų turi konkrečius reikalavimus, kuriuos reikia įvykdyti. Indeksas apima ne vien JAV bendroves. Norint apskaičiuoti S&P 500 indeksą naudojama formulė:

$$S\&P\ 500 = \frac{\sum(p_i \cdot q_i)}{d}; \quad (3)$$

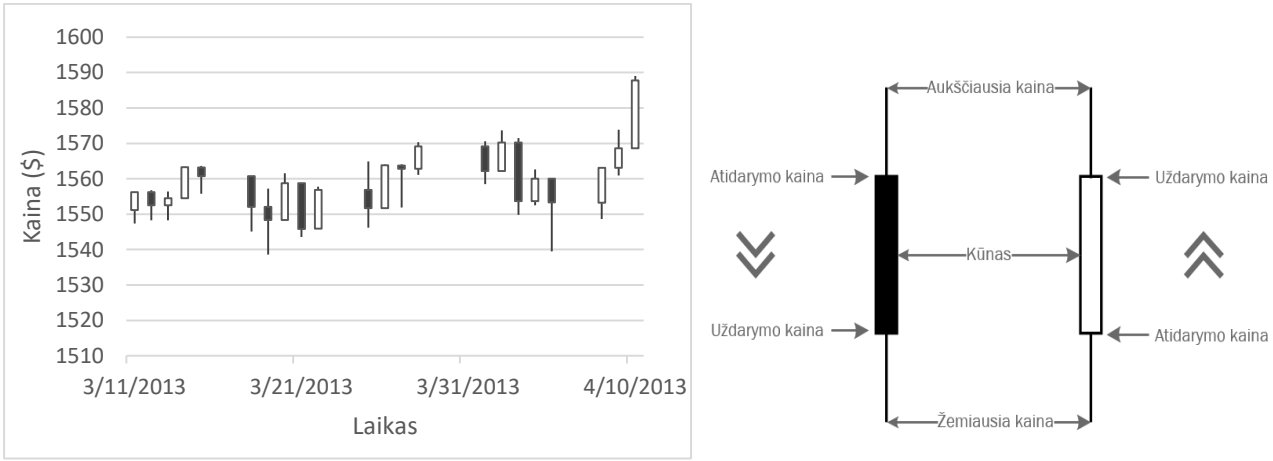
kur p_i yra i tojo indekso akcijų kaina, o q_i – kiekvienai akcijai viešai prieinamų akcijų skaičius, d – yra daliklis, kuris yra koreguojamas atsižvelgiant į struktūrinius pokyčius, atsiradusius dėl akcijų dalijimų ar kitų panašių įvykių. [12]

1.2 Klasikiniai akcijų prognozavimo metodai

Yra daugybė tradicinių metodų, kurie taikomi akcijų kainų kitimo bei akcijos kainos mainų dienos pabaigai prognozuoti. Šiam tikslui egzistuoja dvi labai svarbios teoremos – veiksmingos rinkos hipotezė (EMH) [13] ir atsitiktinio dreifo teorema [14].

E.F. Fama publikavo efektyvios rinkos hipotezę 1964 [13]. Grindžiant EMH hipotezėmis, būsimoji akcijų kaina ateityje yra nenuspėjama, atsižvelgiant į akcijų istorinius duomenis. Nesubalansuota akcija yra iš karto identifikuojama ir greitai atnaujinama teisingu kainos pokyčiu [9]. EMH egzistuoja trijomis formomis – silpnas EMH, pusiau stipri EMH ir stipri EMH. Silpnos EMH istoriniai duomenys yra naudojami prognozuojant akcijų kainą. Pusiau stiprioje EMH, be istorinių duomenų, prognozuojant akcijų kainą, taip pat naudojama visa dabartinė visuomenei prieinama informacija. Stipraus EMH atveju, visi duomenys, įskaitant istorinius, valstybinius ir privačius duomenis (pavyzdžiui, viešai neatskleista informacija) yra naudojama prognozuoti akcijų kainas. O atsitiktinio dreifo hipotezės teigia, jog akcijų kainos nepriklauso nuo ankstesnių kainų [14]. Vadinasi, nėra pasikartotinų šablonų, kurie gali būti panaudojami, nes istoriniai duomenys nenulemia dabartinės akcijos vertės.

Du klasikiniai vertybinių popierių prognozių modeliai yra techninė analizė ir fundamentalioji analizė [15]. Techninė analizė yra skaitinė laiko eilučių analizių metodika, naudojama prognozuoti akcijų rinką naudojantis istoriniais duomenų diagramas kaip pagrindinį įrankį. 1.2 pav. pateikiamas japonų žvakių diagramos pavyzdys bei dedamųjų elementų paaiškinimai.

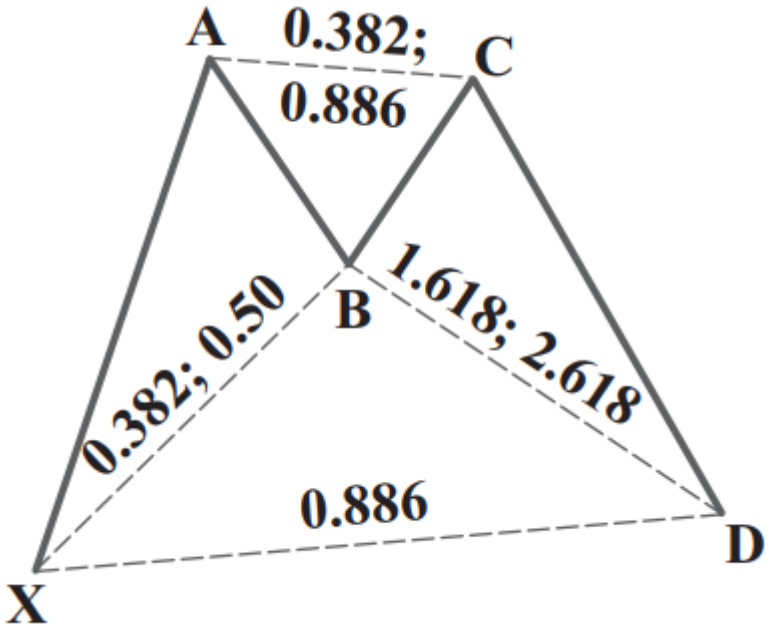


1.2 pav. Kairėje pusėje japonų žvakinė diagrama (angl. *Japanese Candlestick Chart*). Dešinėje pusėje pateikti diagramos simbolių paaiškinimai

Ši akcijų kainų vaizdavimo forma, kaip ir pavadinimas, buvo sukurta Japonijoje daugiau negu prieš 100 metų. Dažniausiai viena žvakė rodo vienos dienos akcijos kainos pobūdį, bet periodas gali būti ir

kitoks. Žvakės kūnas (angl. *real body*) rodo, kokia buvo akcijos kaina periodo pradžioje (angl. *open*) bei pabaigoje (angl. *closed*). Jeigu žvakės kūnas yra nuspalvintas, tai rodo, jog intervalo pradžioje akcijos kaina buvo didesnė nei pabaigoje (kartais naudojamos indikacinės spalvos, užpildymo spalva dažniausiai būna raudona, o jos priešingybė žalia). Ūsai, linijos, einančios iš žvakės kūno, parodo didžiausią bei mažiausią akcijos kainą mainų periode. [16]

Techninė analizė bando taikyti duomenų gavybos metodus siekiant išgauti informaciją iš istorinių duomenų, tam kad nustatyti pasikartotinus šablonus (vadinama laiko eilučių kasyba [15]), dar vadinamus harmonikomis. Harmoninės kainų konfigūracijos pavyzdys pateiktas 1.3 pav. Šiame paveiksliuke pateikta šikšnosparnio harmonikos modelio atmaina, kuri pirma karta aptikta 2001 m. Scott Cerney [17].



1.3 pav. *Bullish Bat* harmoninėmis kainų konfigūracijos pavyzdys [17]

Fundamentalioji analizė – mokslas, tiriantis priežastis, įvykius, kurie nulemia paklausos ir pasiūlos santykį [18]. Šio metodo pagrindinis įrankis yra informacijos rinkimas ir jos interpretacija, siekianti nustatyti akcijų kainas. Šios analizės rezultatas suteikia prekybos galimybę, identifikavus atotrūkį tarp įvykio ir rinkos reagavimo į įvykį skirtumo. Svarbūs duomenys, naudojami fundamentalioje analizėje, yra įmonių ekonominiai duomenys (pavyzdžiui, metiniai ir ketvirtiniai pelnai), audito ataskaitos, balansai ir pajamų paskelbimas. Naujienos taip pat lemia esminę analizę, nes naujienos atspindi dabartinę pasiūlos ir paklausos grandinę akcijų rinkoje. Šie tradiciniai metodai šių dienų rinkoje tampa nebenaudingi dėl padidėjusios kompiuterių skaičiavimo galios, kuri dabar gali tiksliau analizuoti didelius duomenų rinkinius per trumpesnį laiką. Tačiau šie metodai vis dar yra pagrindas kuriant naujus metodus, tokius kaip dirbtinis intelektas bei kiti.

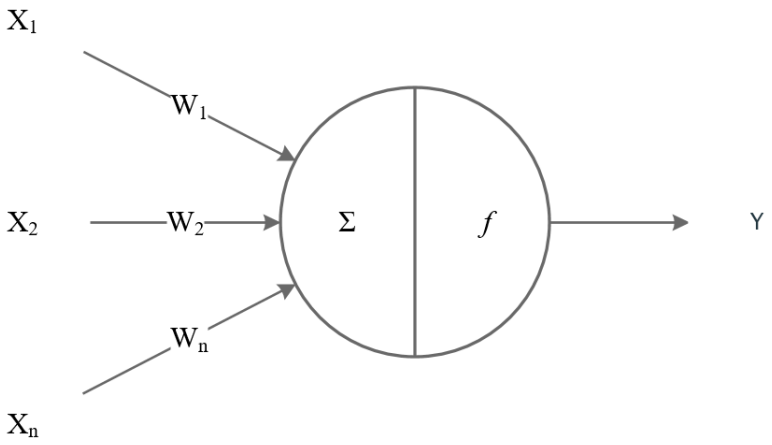
1.3 Dirbtinio intelekto metodai

Iš prigimties akcijų kainų kitimo procesas yra nepastovus ir netiesinis. Šio pobūdžio uždaviniams įvairiuose sektoriuose, kaip ir akcijų kainų prognozavime, yra taikomi dirbtiniai neuronai tinklai. Tačiau jie pradėti taikyti tik pastaraisiais dešimtmečiais, nes jie reikalauja daugiau skaičiuojamosios galios. Be to, ypač mokslinių tyrimų sektoriuje, į jų taikymą žiūrima skeptiškai, nes interpretuoti

sukurtą modelį yra sunku, dažnai neįmanoma (literatūroje dar vadinami juodosiomis dėžėmis angl. *Black Box*). Nepaisant to, jų savybės yra naudingos akcijų kainoms prognozuoti.

1.4 Dirbtiniai neuroniniai tinklai

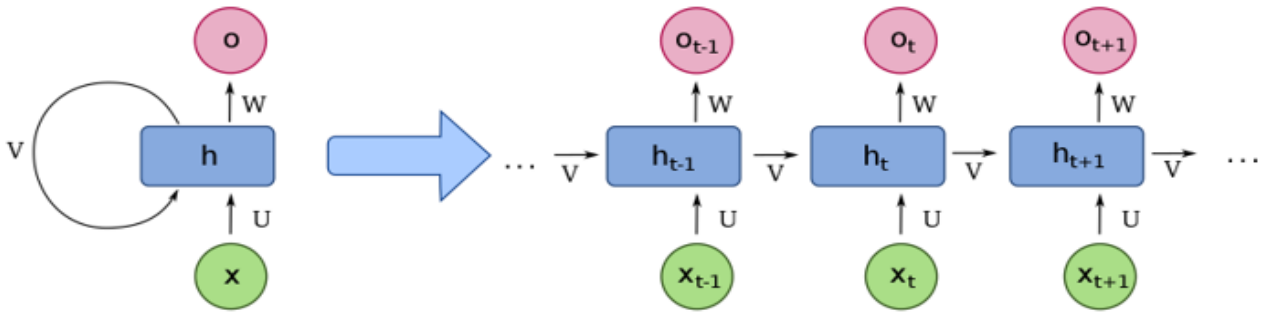
Dirbtiniai neuroniniai tinklai (angl. *artificial neural networks*, DNT) yra iš mazgų, neuronų sudarytos sistemos. Šios sistemos yra sumodeliuotos replikuojant biologinius neuroninius tinklus, matematinis biologinio neurono modelis yra vadinamas parceptronu, o jo schema pateikta 1.4 pav. Parceptono įvesties signalai pavaizduoti schemoje X_i, kur *i* yra įvesties signalo numeris. Įvestys yra agreguojamos, operacija žymima Σ ženklu, jas padauginus iš atitinkamų svorių W_i. Agreguotos reikšmė toliau perduodama aktyvacijos funkcijai (žymimai *f*). Jei agreguota įvesties reikšmė viršija kritinę ribą (priklausomai nuo aktyvacijos funkcijos), modelis siunčia signalą, žymimą *Y*, kuris atitinkamai yra siunčiamas kitiems parceptronams ar DNT išvesties gavėjams.



1.4 pav. Parceptronas

1.5 Grįžtamojo ryšio neuroniniai tinklai

Grįžtamojo ryšio neuroninių tinklų (angl. *Recurrent Neural Networks*, RNN) architektūros pagrindinė savybė išnaudoti įvesties sekos savybes. Klasikinėje neuroninių tinklų architektūroje visi kintamieji, įvestys ir išvestys yra traktuojami kaip vienas nuo kito nepriklausomi. RNN tinklų architektūroje kiekvienai įvesčiai atliekama ta pati seka operacijų, tačiau rezultatai priklauso nuo praeities rezultatų. Šią priklausomybę nuo praeities reikšmių dar galima interpretuoti kaip atmintį. RNN architektūros tipinė architektūros schema pateikta 1.5 pav.

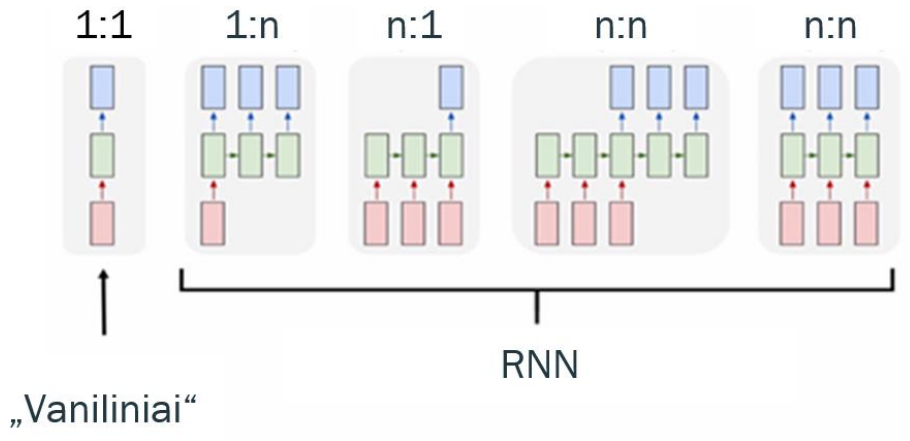


1.5 pav. RNN išskleista architektūra [19]

Aukščiau pateiktame paveiksle pateikta RNN „išskleista“ schema, joje tiesiog vaizduojama eilė sujungtų neuronų kiekvienam įvesties sekos elementui. Jei duomenų seka susidarytų iš 3 elementų, tai grįžtamojo ryšio neuroninį tikslą galėtume pavaizduoti kaip 3 sluoksnių neuroninių tinklo sistemą. Aukščiau pateikta architektūra sudaryta iš tokių elementų [20]:

- x_t – įvesties duomenys laiko momentu *t;*
- h_t – paslėpta ląstelės būsena laiko momentu *t* literatūroje dar vadinama paslėpta būsena (angl. *hidden state*). Šią būseną galima traktuoti kaip ląstelės atmintį. Celės būsena yra apskaičiuojama atsižvelgiant į ankstesnę ląstelės būseną ir naują įvesties informaciją $h_t = f(U \cdot x_t + V \cdot h_{t-1})$. Funkcija f dažniausiai yra netiesinė duomenų transformavimo funkcija, o pradinė ląstelės būsena kai *t* = -1, yra parenkama modelio sukūrimo metu (dažnai įvesties vektorius su 0 reikšmėmis).
- o_t yra DNT išvesties rezultatas laiko momentų *t*.

Šio dirbtino intelekto modelio panaudojimo atvejai labai skiriasi nuo tiesioginio sklidimo tinklų, kurie dar vadinami „Vaniliniais“ neuroniniais tinklais (1.6 pav.). RNN nėra apriboti įvesties ir išvesties santykiu – įvesties ir išvesties santykis neprivalo būti vienas su vienu. Informacijos šaltinis gali būti vaizdo įrašo failas (laiko eilutės), kur kiekvienas kadras gali būti paduodamas atskirai, o rezultatas gali būti aprašas to, kas atskleidžiama vaizdo įrašo kadruose, arba bitinė reikšmė, pavyzdžiui, faktas, ar žmogus bėga.



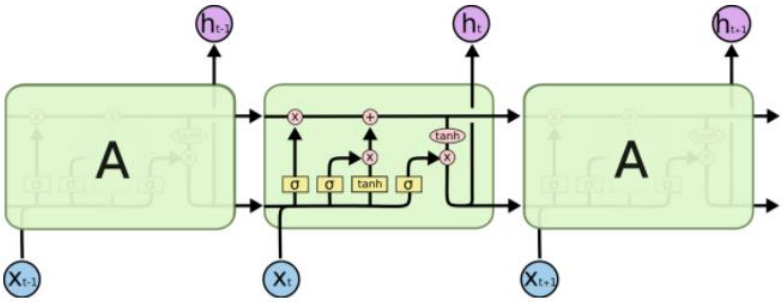
1.6 pav. Neuroninių tinklų tipai

1.5.1 Trumpalaikės-ilgalaikės atmintimi grįstas modelis

Trumpalaikės-ilgalaikės atminties neurotinis tinklas, dažnai dar vadinamas LSTM (angl. *Long Short Term Memory*), yra išskirtinis grįžtamojo ryšio neuroninis tinklas, gebantis mokyti ilgalaikes priklausomybes. Šis modelis buvo pristatytas Hochreiterio ir Schmidhuberio 1997 metais [21], po publikacijos šis modelis buvo toliau tobulinamas ir populiarinamas kitų mokslininkų bei sričių specialistų. Šie tinklai geba išspręsti daug problemų ir yra plačiai naudojami šiuolaikiniuose projektuose.

LSTM buvo sukurti siekiant išspręsti ilgalaikių priklausomybių problemą. Gebėjimas prisiminti įvykius ir nustatyti jų priklausomybes yra šio modelio stipriausia savybė. Šios savybės dauguma dirbtinio intelekto modelių stokoja. Visi grįžtamojo ryšio neuroninių tinklų modeliai turi grandis pasikartojančių neuroninių tinklų modulių. Standartiniuose RNN tinkluose šis modelis turės labai paprastą struktūrą, sudarytą iš vieno neuroninio tinklo lygmens. LSTM vietoj vienintelio neuronų

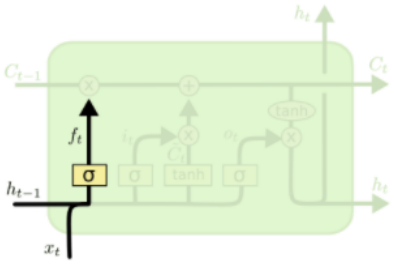
tinklo lygmens turi keturis. 1.7 pateiktoje diagramoje kiekviena linija keliauja duomenų vektorius nuo išvesties mazgo į kito modulio įvestį. Rožiniai apskritimai reprezentuoja transformacijas, atliekamas su vektoriais – sudėties, daugybos ir t.t. Geltoni stačiakampiai simbolizuoja apmokytus neuroninius tinklus. Linijose, kurios yra sujungiamos, žymi sąveika, o linija, kurios išsišakoja, reiškia, kad jos turinys yra kopijuojamas ir kopijos pasiskirsto į skirtingas vietas. Svarbi LSTM savybė yra celės būsena, tai horizontali linija, esanti diagramos viršuje. Celė eina horizontaliai per visą grandinę su keliomis mažomis linijinėmis sąveikomis. Dėl šios priežasties informacija gali keliauti grandimis nepakitusi.



1.7 pav. Pasikartojantys LSTM tinklo moduliai, kurie sudaryti iš keturių neuroninių tinklų [22]

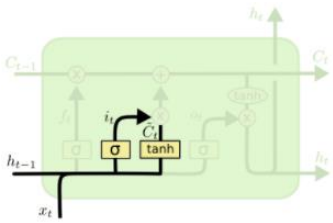
LSTM turi galimybę pridėti arba pašalinti informaciją iš ląstelės, kuri yra reguliuojama ląstelės slenksčiais. Slenksčiai yra būdas optimaliai praleisti informaciją. Jie susideda iš sigmoidinio neuroninio tinklo sluoksnio ir kryptinės daugybos operacijos. Sigmoidinis lygmuo rodo skaitmenis intervale nuo nulio ir vieno, o tai apibūdina, kiek specifinės informacijos turi būti praleista. Nulio vertė reiškia, jog informacija bus visiškai užslopinta, o vienetas reiškia, jog informacija nėra koreguojama.

Pirmasis žingsnis LSTM modelyje yra sprendimas, lemsiantis, kurios ląstelės informacijos reikia atsisakyti. Šis sprendimas yra atliekamas sigmoidiniame lygmeniu (diagramoje pažymėtas σ), vadinamu užmiršties slenksčio lygmeniu (angl. *forget gate layer*). Šis lygmuo, atsižvelgdamas į ankstesnę ląstelės išvestį (*h*) laiko momentu *t*-1 ir naują įvestį *x*, sudaro slopinimo reikšmę tarp 0 ir 1 kiekvienai celei (1.8 pav.).



1.8 pav. Pirmasis LSTM veikimo žingsnis **[22]**

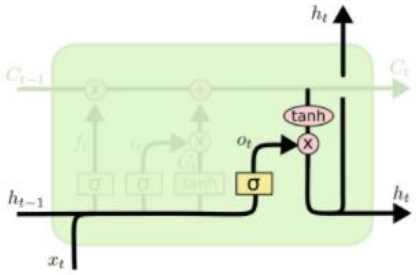
Kitame žingsnyje (1.9 pav.) yra nusprendžiama, kurią naują ląstelės informaciją reikia išsaugoti. Šis žingsnis susideda iš dviejų dalių. Pirmiausia sigmoidinis lygmuo, pavadintas įvesties slenksčio lygmeniu (angl. *input gate layer*), nusprendžia, kurią reikšmę reikia atnaujinti. Po to tanh lygmuo sukuria vektorių su naujomis kandidatų reikšmėmis Ct, kurios gali būti pridėtos prie ląstelės būsenos. Kitame žingsnyje šios reikšmės bus sujungiamos sukuriant ląstelės būsenos atnaujinimo vektorių.



1.9 pav. Antrasis LSTM veikimo žingsnis [22]

Atlikus aukščiau išvardintus žingsnius, galima atnaujinti seną ląstelės būseną. Šiam tikslui pasiekti seną ląstelės būseną padauginame iš slopinimo reikšmės (žiūrėti pirmo žingsnio aprašymą) tam, kad nereikšminga informacija būtų pašalinama arba prislopinta. Po to prie atnaujintos celės būsenos yra pridedama antrame žingsnyje sumodeliuota ląstelės būsena.

Paskutiniame žingsnyje (1.10 pav.) priimamas sprendimas, kokias reikšmes reikia išvesti. Išvestis yra sudaroma iš transformuotos ląstelės būsenos. Pirmiausia, sigmoidinis lygmuo „nusprendžia“, kuri ląstelės būsenos dalis yra reikšminga. Tuomet celės reikšmė yra normalizuojama, panaudojant tanh operaciją (tam, kad reikšmės patektų į -1 ir 1 intervalą) gautas rezultatas yra padauginamas iš sigmoidinio slenksčio lygmens rezultato tam, kad būtų išvestos tik reikšmingos reikšmes.

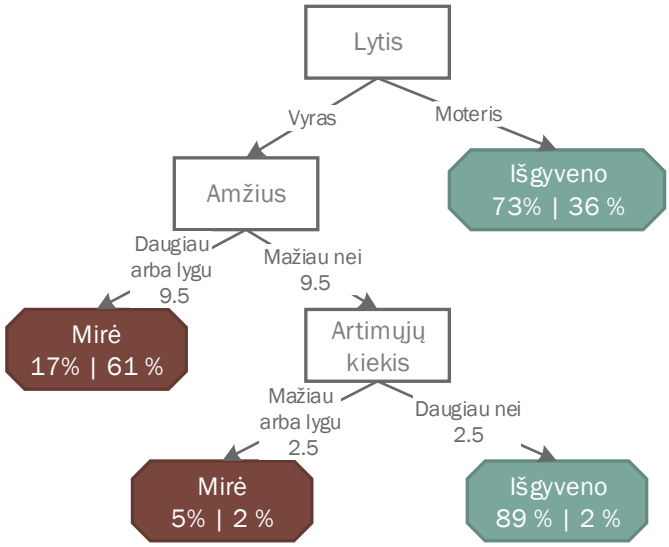


1.10 pav. Informacijos išvedimo žingsnis [22]

1.6 Sprendimų medis

Sprendimų medis – remiantis objekto charakteristikomis sukurtas modelis, kurį galima pavaizduoti kaip medį. Medžio briaunos sujungia objektą charakterizuojančius faktus su išvadomis, lapais, kurie parodo objektui priskirtą klasę. Klasifikavimo ir regresiniai medžiai yra DI metodai, skirti sukurti duomenų prognozavimo modelius. Šie modeliai yra sukurti rekursyviai dalijant objektą aprašančių duomenų aibę ir apmokant primityvų prognozavimo modelį kiekvienai išskirtai charakteristikų imčiai. Šią duomenų poaibių hierarchiją galima nesunkiai pavaizduoti grafiškai. Klasifikavimo medžiai yra sumodeliuoti katekoriniams kintamiesiems, šiuo atveju modelio paklaida yra vertinama blogai klasifikuotu klasių kiekiu. Regresijos medžiai modeliuoja tolydžius priklausomus kintamuosius, šiuo atveju modelio prognozių paklaida yra apskaičiuojama vidutiniu kvadratiniu nuokrypiu nuo prognozuojamos reikšmės. Norint sprendimų medžiuose naudoti tolydžius kintamuosius reikia parinkti reikšmių skėlimo tašką (angl. *split-point*), pagal kurį kintamojo reikšmės bus skaidomos (paprastai) į dvi dalis. Tam gali būti naudojami parasti metodai, kai skaidymui naudojama mediena (tuomet imtis padalijama į dvi lygias dalis), bei sudėtingi metodai, kai papildomai išbandomi įvairūs skaidymo variantai ir siekiama parinkti optimalų kriterijų. [23]

Žemiau pateiktas sprendimų medžio pavyzdys (1.11 pav.). Diagramoje galima matyti sprendimų medį, kuris sudarytas remiantis Titaniko katastrofos statistiniais duomenimis. Šio medžio lapai yra klasės, nusakančios, ar Titaniko keleivis išgyvens ar mirs. Šakos parodo atributus, pagal kuriuos skaidomi duomenys: lytis, amžius ir artimųjų kiekis. Lapuose antrasis skaičius parodo, kuri duomenų dalis yra klasifikuojama atitinkama sprendimų seka, o pirmasis skaičius parodo, kuri dalis duomenų yra teisingai klasifikuojama (abi reikšmės išreikštos procentais). [24]



1.11 pav. Titaniko katastrofą išgyvenimo tikimybe vaizduojantis sprendimų medis

Medžiams sukurta daug įvairių algoritmų, populiariausi iš jų: ID3, C4.3, CERT, CHAID. ID3 yra vienas iš pirmųjų algoritmų, skirtų šiems medžiams kurti. Algoritmas sukurtas J. R. Quinalo [25], jo pagrindinės koncepcijos yra naudojamos ir modernesniuose algoritmuose. Šio algoritmo principas yra kiekviename medžio lygyje pasirinkti tokį skirstymo kriterijų, jog būtų prarandama kuo mažiau informacijos. ID3 algoritme medžio šakos skaidymui parenkamas kintamasis pagal didžiausią informacijos išlošimą (angl. *information gain*). Pirmame žingsnyje duomenų rinkiniui yra apskaičiuojama entropija pagal formulę:

$$E(S) = -\sum_{x \in X} p(x)\log_2 p(x)\,; \tag{4}$$

čia E – duomenų informacija (taip pat vadinama entropija, matuojama bitais), S – duomenų rinkinys, kuriam apskaičiojama entropija, X – rinkinio S klasės, p(x) – klasės x proporcijos rinkinyje S kiekis. Tuomet duomenų imtis yra skaidoma pagal kiekvieną atributą, ir apskaičiuojamas informacijos gavimas pasinaudojant formulę:

$$E(S,A) = H(s) - \sum_{t \in T} p(t)E(t) = H(S) - H(S|A)\,; \tag{5}$$

čia T imtys sukurtos imtį S padalijus pagal atributą A. Pasirenkamas atributas, pagal kurį perskelto duomenų rinkinio entropija yra mažiausia. Tada sukuriamas išsišakojimas ir šioms naujoms šakoms taikomas pirmas žingsnis, kol skaidymas į naujas šakas nebesumažina naujų imčių entropijos.

Sprendimų medžiai yra linkę persimokyti. Norint to išvengti reikia medžius genėti (angl. *prune*), atsikratant šakų esančių arti medžio lapų. Medžio genėjimo metodai skaidomi į du tipus:

- išankstinis genėjimas (angl. *prepruning*) — šiuo atveju medelio sudarymo metu yra įvedama šakojimosi stabdymo sąlyga, tai kriterijus, kuriuo remiantis skaidymas į šakas stabdomas anksčiau;
- genėjimas po medžio sudarymo (angl. postpruning) — medis sudaromas iki galo, o vėliau naudojamas koks nors algoritmas (kriterijus) šakoms genėti.

1.6.1 Atsitiktinis miškas

Sprendimų medžių modeliai yra dažnai naudojami kaip vaizdinė priemonė, tačiau jie nėra dažnai naudojami kaip klasifikatoriai. To priežastis – sprendimų medis yra linkęs persimokyti ir neturi didelio klasifikavimo tikslumo. Taip pat šiems modeliams sunku parinkti optimalų genėjimą. Šiuos trūkumus bandoma išplėsti panaudojant atsitiktinius miškus (angl. *random forest*), kurie yra priskiriami metodams ansambliams. Atsitiktiniai miškai yra dažniau naudojami nei sprendimų medžiai, nes jie nelinkę persimokyti, yra daug tikslesni ir gali klasifikuoti daugiau klasių. Naudojant šį metodą galima sukurti labai galingus klasifikatorius, kurie gali būti panaudojami žaidimų, automatizavimo ar net ligų prognozavimo srityje. Šio modelio principas vienai apmokymo imčiai sudaromas ne vienas medis, o daug sprendimų medžių (pvz. 100 ar 1000). Tam, kad galėtume sudaryti medžius, kurie nebūtų identiški, medžių kūrimo metu naudojamas atsitiktinumas (angl. *randomization*) [26]:

- medžiui sudaryti naudojama ne visa, o atsitiktinė imties dalis;
- kintamasis imties skaidymui į medžio šakas parenkamas ne optimaliai iš visų kintamųjų, o iš atsitiktinės dalies kintamųjų.

Turint sprendimų medžių mišką, kiekvienas medis atlieka klasifikavimo užduotį, o galutinė klasė parenkama atliekant „balsavimą“ pagal daugumos principą. Kartu tai leidžia nustatyti ir tikimybę, su kuria taškas yra priskiriamas tam tikrai klasei. Taigi, atsitiktiniai miškai nėra linkę persimokyti, net jei juos kuriant naudojami į persimokymą linkę algoritmai.

Atsitiktiniai miškai turi daug savybių, kurių parinkimas gali nulemti klasifikatoriaus kokybę:

- Medžių kiekio parinkimas**.** Ši savybė nulemia, kaip tiksliai atsitiktinis miškas geba klasifikuoti objektus.
- Triukšmas ir klasių kiekis**.** Atsitiktinis miškas taip pat išsiskiria tuo, jog jo tikslumui mažai įtakos turi klasių kiekis bei triukšmas. Šie modeliai yra labai stabilūs ir, nors pavieniai medžiai linkę persimokyti, miškas šio trūkumo neturi.
- Miško gylis. Šis parametras lemia algoritmo persimokymo lygį. Parinkus didelį gylį modelis persimoko ir pradeda identifikuoti klasės triukšmą, kai tuo tarpu mažas gylis nulemia, jog duomenys bus nepakankamai gerai atskiriami.

Atsitiktinis miškas yra plačiau naudojamas analitikoje, objektų klasifikavime bei komerciniuose produktuose, tokio produkto vienas iš pavyzdžių Kinect [27]. Sukurtame produkte, šis sprendimų miškų architektūrą panaudota siekiant greitai ir tiksliai nuspėti 3d aplinkoje kūno dalių padėtį. Šiame projekte mokslininkai naudojo didelius duomenų kiekius, kuriuose asmenys buvo fiksuojami įvairiose padėtyse ir aplinkose. Sukurtas produktas geba veikti 200 kadrų per sekundę greičiu.

1.7 Statistiniai metodai

Dauguma finansinių tyrimų tiesiogiai netiria proceso, o orientuojasi į proceso pokytį (šiuo atveju akcijos kainos pokytį laike). Campbellas ir kt. (1997) [28] pateikia dvi pagrindines priežastis. Pirma, investicijos grąža neturi priklausomybės nuo mastelio. Antra, grąžos proceso laiko eilutės yra lengviau modeliuojamos nei kainų proceso, nes pirmosios turi daugiau savybių, naudingų statistinei analizei. Egzistuoja keletas investavimo grąžos apibrėžimų, tarkime, P_t yra tiriamo objekto kaina laiko momentu t. Tegul už įsigytą objektą ar jo dalį negaunama dividendų.

Vieno laikotarpio paprastoji grąža. Turto nusipirkimas laiko momentu *t*-1 ir jo pardavimas laiko momentu t atneš bendrą pelną R_t:

$$1 + R_t = \frac{P_t}{P_{t-1}}, \text{arba } P_t = P_{t-1}(1 + R_t). \tag{6}$$

Išreiškus R_t gauname vieno periodo paprastąją grąžą:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{7}$$

Atitinkamai turto laikymas *k* periodu sukurs *k* periodų paprastąją grąžą:

$$\begin{aligned} 1 + R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_1}{P_{t-2}} \cdot \ldots \cdot \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1}) \ldots (1 + R_{t-k+1}) \\ &= \prod_{j=0}^{k-1}(1 + R_{t-j}). \end{aligned} \tag{8}$$

1.7.1 Eksponentinio glodinimo metodai

Eksponentinio glodinimo metodai glodina laiko eilutes vidurkinant stebėjimus. Glodinimo procesas leidžia išlyginti laiko eilučių duomenis, naudojant eksponentinio lango funkciją. Šią metodų grupę paprasta interpretuoti bei taikyti remiantis prielaidomis apie tiriamą procesą, kaip sezoniškumo egzistavimą ar liekanų autokoreliaciją. Šis metodas taip pat leidžia išskaidyti laiko eilutę į jos dedamąsias, tokias kaip sezoniškumą ar trendą. Eksponentinis glodinimas taip pat yra naudojamas finansiniams duomenims prognozuoti, kai optimali ateities prognozė yra proceso praeities reikšmių svorinis vidurkis. Šiame skyriuje naudojami įprasti eksponentinio glodinimo žymėjimai: y_t žymės stebėtos laiko eilutės reikšmę, o $\hat{y}_{T+h|T}$ žymės y_{T+h} prognozės reikšmę, kai stebėtos reikšmės iki $y_1, \ldots, y_T$. Toliau pateiktas trigubas arba *Hot-Winters* sezoninis modelis:

$$\hat{y}_{T+h|T} = \mathrm{l_t} + \mathrm{hb_t} + \mathrm{s_{t+m+((h-1)mod(m))+1}}; \quad (9)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1}+b_{t-1}); \quad (10)$$

$$b_t = \beta^*(l_t+l_{t-1}) + (1-\beta^*)b_{t-1}; \quad (11)$$

$$s_t = \gamma^*(y_t-l_{t-1}-b_{t-1}) + (1-\gamma^*)s_{t-m}; \quad (12)$$

čia (9) – prognozės lygtis, (10) vidurio glodinimas, (11) trendo glodinimas, (12) sezoniškumo glodinimo lygtis, β^* – trendo glodinimo parametras, γ^* – sezoniškumo glodinimo parametras, t – laiko momentas, m – sezoniškumo parametras. Šis modelis susideda iš adityvaus vidurkio ir sezoniškumo trendų. Praktikoje naudojami ir kiti trendų tipai, kurie yra žymimi taip: N — nenaudojamas trendas, A — adityvus, Ad — adityvus nuslopintas, M — multiplikavus (eksponentiniai), Md — multiplikavus nuslopintas. Trumpai šie modeliai žymimi ETS(1, 2, 3) ir skliaustuose nurodomi trendų tipai, 1 - sis parametras nurodo vidurkio glodinimo tipą ir gali būti N, A, Ad, M arba Md, 2-sis parametras nurodo sezoniškumo tipą N, M, A, o trečiasis, kuris taip pat gali būti N, M arba A, liekanų. ETS(A, A, N) ir atitinka lygtyje aprašytą modelį. Tokių modelių yra 30. [29]

1.7.2 ARIMA

ARIMA – autoregresinis (angl. *autoregressive*, AR), integruotas (angl. *integrated*, I) judančio vidurkio (angl. *moving average*, MA) modelis. ARIM modelis yra naudojamas tiek duomenų analizei, tiek atsitiktinio proceso, žymimo ξ_t, reikšmių, prognozavimui. Šis modelis yra naudojamas, kai tiriamas procesas yra nestacionarus ar pasižymi nestacionaraus proceso savybėmis. AR – šio modelio dalis modeliuoja tiriamo proceso dabartinių (laiko momentu *t*) reikšmių priklausomybę nuo praeities reikšmėms (*t-k,* kur k poslinkis laike). MA modelio dalis apibūdina modelio paklaidų priklausomybę nuo praeities paklaidų (paklaida tai tiesiškai išreikšta praeities ir dabarties paklaidų kombinacija). I – reiškia, jog duomenys, naudojami modelyje, yra integruoti. ξ_t vadinamas d eilės integruotu procesu (žymima $\xi_t \in \mathrm{I}(d)$), jei jo d eilės pokyčiai yra stacionarus procesas, o d-1 eilės pokyčiai nėra stacionarūs. [29]. Šis modelis žymimas ARIMA(p, d, q), kur p, d ir q atitinkami modelio parametrai, o jo matematinė išraiška:

$$P(L)(1-L)^d\xi_t \ = Q(L)\varepsilon_t; \quad (13)$$

čia $P(z) = z - a_1z - \cdots - a_pz^p$, o $Q(z) = z + a_1z + \cdots + a_pz^q$ yra atitinkami AR ir MA modelio dalies polinomai, o p ir q atitinkami jų laipsniai ar modelio eilės, ε_t – balto triukšmo procesas. L^k yra k postūmių atgal operatorius:

$$L^k = \xi_{t-k}. \quad (14)$$

1.7.3 GARCH

Praktikoje dažna situacija, kai modelio liekanos, nors ir yra nekoreliuotos, tačiau jų dispersija nėra pastovi. Su tokia situacija dažnai susiduriama regresiniuose modeliuose. Tiesinėje regresijoje tokiu atveju taikomas apibendrintas mažiausių kvadratų metodas. Laiko eilučių analizėje nepastovios liekanų dispersijos atvejis yra taip pat labai dažnas. Pavyzdžiui, jei analizuotume akcijų grąžas, tai

pastebėtume, kad jos linkusios įgauti didesnes išsibarstymo reikšmes tam tikram laikotarpiui, o po to grąžos nusistovi ir jų kintamumas vėl įgauna ankstesnes nedideles reikšmes. Atsitiktinis procesas tenkina ξ_t apibendrintą autoregresinį sąlyginio heteroskedastiškumo GARCH(*p, q*) (angl. *general autoregressive conditional heteroskedastic*) modelį [29], jei:

$$\xi_t = \sigma_t \varepsilon_t; \tag{15}$$

čia ε_t – baltas triukšmas, o σ_t – sąlyginis heteroskedastiškumo (sklaida), tenkina lygtį:

$$\sigma_t{}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \, \xi_{t-1}{}^2 + \sum_{i=1}^{p} \beta_i \, \sigma_{t-1}{}^2; \tag{16}$$

čia α_i ir β_i, parenkami parametrai.

1.8 Ankstesni akcijų prognozavimo bandymai

Kadangi ši sritis yra labai populiari, dėl techninių iššūkių ir finansinės naudos yra atlikta daugybė bandymų ir sukurta daug modelių. ANN su atgaliniu ryšiu yra bandyta taikyti įvairiose rinkose. R. Aghababaeyanas (2011m.) [30] atliko tyrimą dėl prognozės Teherano vertybinių popierių biržoje, kur sukūrė įrankį, pagrįstą ANN, kurio tikslumas siekia 97%. Khanas ir kiti (2011) [31] atliko Bangladešo vertybinių popierių rinkos tyrimą, kuriame sukūrė įrankį, kurio vidinė paklaida yra 3,7% ir 1,5% dviem simuliacijoms. Nė vienas iš tų įrankių nėra viešai skelbiamas, nebuvo sukurtas komerciškai ar tikslingai atitinkamiems vertybinių popierių makleriams.

Šiose srityje taip pat labai gerus rezultatus parodė hibridiniai metodai, naudojantys kelis prognozavimo metodus iš karto. G. I. Sheras sukūrė ir ištyrė kainos prognozavimo sistemą, kuri yra paremta neuroniniais tinklais, tačiau atsižvelgia ir į tai, ar susidariusi geometrinė konfigūracija. Buvo iškelta ir patvirtinta hipotezė: atsižvelgiant į geometrines konfigūracijas ir naudojant dirbtinį neuroninį tinklą galima pasiekti didesnio pelningumo nei vien tik naudojant nėrininį tinklą. Buvo pasiektas 52% pelnas per 15 dienų laikotarpį su 1:50 svertu [32].

2 AKCIJŲ RINKOS MODELIAVIMO PROJEKTAS

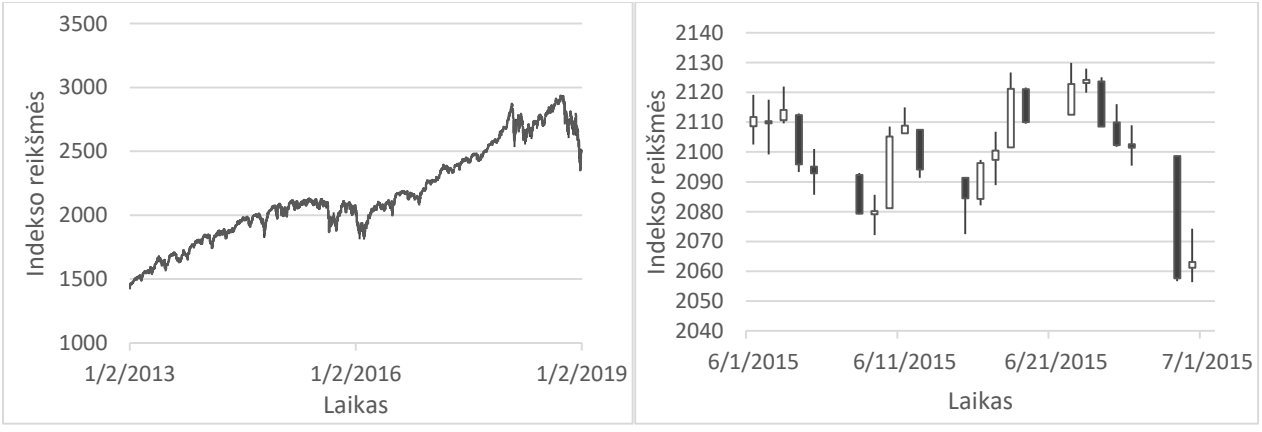
Šiame skyriuje aptariama tyrimo eiga, kuriami modeliai bei aptariami kriterijai, kuriais remiantis yra tikrinami sukurti prognozavimo akcijų kainų metodai. Įvardijami pastebėjimai bei rezultatai, leidę suprojektuoti modelių asamblėjas. Skyriaus pabaigoje aprašoma sukurta sistema, kurioje atlikti bandymai ir aptartos technologijos bei programinės įrangos priemonės, naudotos šio tyrimo metu.

2.1 Tyrimo eiga

Šio projekto tikslas yra tirti akcijų kainas, siekiant prognozuoti jų pokyčius ateityje, norint prognozes panaudoti atliekant finansinius sprendimus. Tam, kad tinkamai identifikuotume akcijų kainų charakteristikas, leidžiančias nuspėti proceso pokytį ateityje, pirmiausia pasirinkti tyrimui duomenys. Istoriniams duomenims yra keliami tokie reikalavimai:

- kompanija, kurios akcijų kainų reikšmės bus naudojamos tyrime, turi būti gerai žinoma;
- akcijos beta reikšmė pasirinktame intervale turi būti didesnė arba lygi 0.8, bet mažesnė už 1.4;
- pasirinktos kompanijos, kurios vertybinių popierių kainų pokyčiai yra analizuojami, turi patekti į 500 didžiausių pasaulio kompanijų sąrašą pasirinktu analizės laikotarpiu;
- pasirinkta akcija turi būti prekiaujama bent 5 metus;
- pasirinkti duomenys turi būti viešai prieinami.

Nuspręsta prognozuoti ne pasirinktų akcijų paketų duomenis, o prognozuoti akcijų rinkos indeksą. Šis sprendimas priimtas analizuojant rinkos tendencijas. Akcijų rinkos dalyviai vis dažniau renkasi ne aktyvias investavimo strategijas, o investavimą į fondus, ar jų akcijas (EDFs, žiūrėti skyrelį 1.1). Rinkos indeksas yra akregacija jo komponenčių (dažnai akcijų kainų), tiesiškai transformuota atsižvelgiant į pokyčius apimančius jos komponentus, akcijas. Todėl rinkos indeksas yra visą rinką reprezentuojantis, procesas iš dalies normalizuotas – jo rodmenys nekinta jai įvyksta jo komponenčių restruktūrizacija (akcijų skaidymas, padidėja akcijų kiekis it t.t.). Tyrimui pasirinktas S&P 500 (ženklinimo simbolis ^*GSPC*, ^ – pažymi jog tai indeksas) procesas, kurio komponentės yra 505 didžiausių Amerikos kompanijų akcijos ir apima 80% Amerikos rinkos kapitalizacijos, beta tyrimo metu svyruoja nuo 0.96 – 1 (periodas 1 metai). 2.1 pav. kairėje pusėje pateikti istoriniai ^*GSPC* indekso rodmenys pasirinktame tyrimo intervale nuo 2013 iki 2019 metų, o dešinėje pusėje šių duomenų vieno mėnesio segmentas.



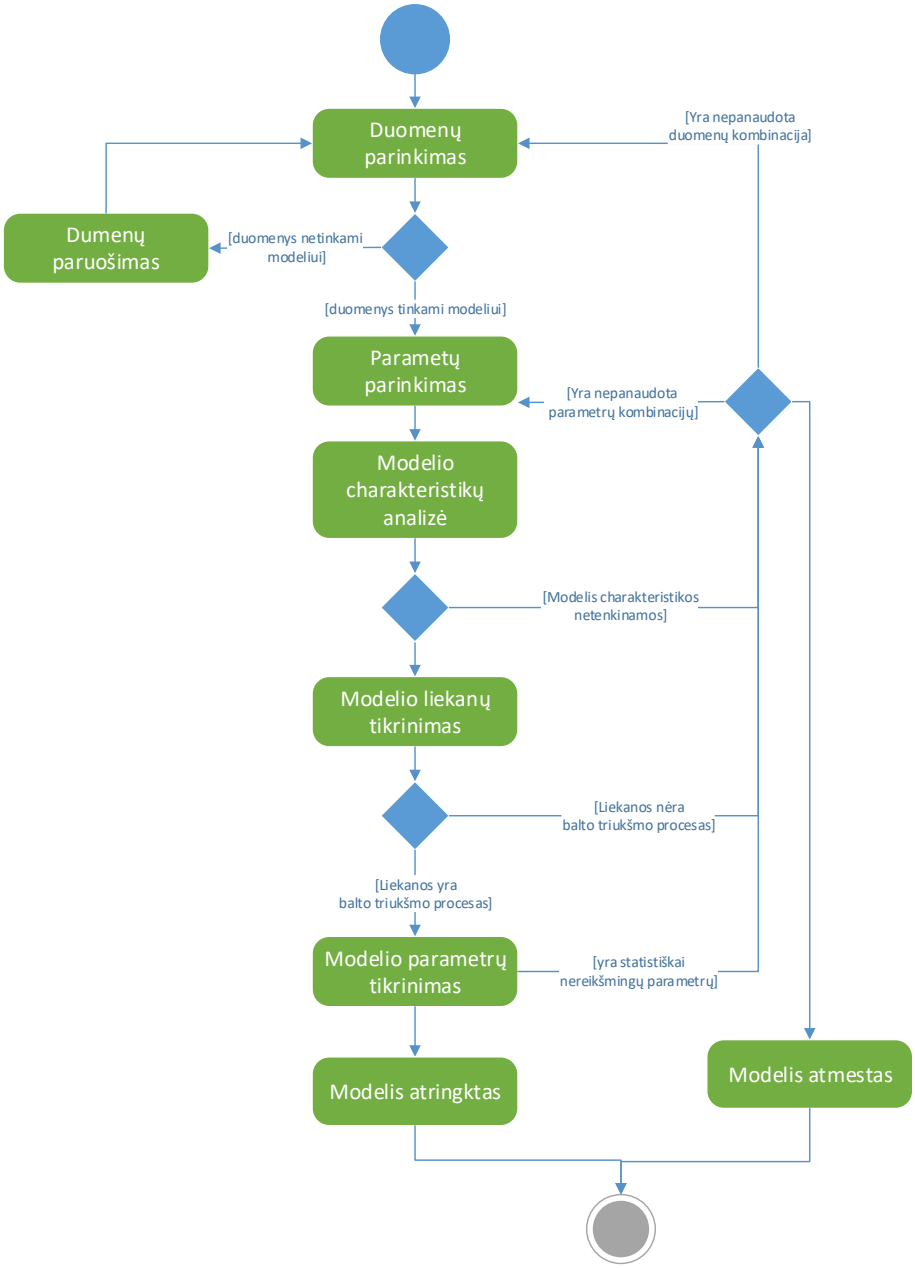
2.1 pav. ^*GSPC* indekso istoriniai duomenys, naudojami tyrime

Akcijos kainos pasižymi savybe – priklausomai nuo populiarumo, kaip dažnai yra prekiaujama įmonės vertybiniais popieriais, impulsai, paveikiantys rinką, tokie kaip stichijos, resursų kainų šuoliai ar kiti panašūs įvykiai, atsispindi kainų pokyčiuose skirtingais tarpais nuo impulso atsiradimo. Todėl tyrimo pradžioje yra identifikuotos bendrovės, kurios tikėtina, jog pastebimai ir reliatyviai greitai reaguos į naujai atsiradusią informaciją, bei kiti procesai, kurie gali būti panaudojami kaip regresoriai, leidžiantys nuspėti rinkos pokytį. Šiam tikslui pasiekti pasitelktos sklaidos diagramos, padedančios nustatyti procesus, turinčius tarpusavio ryšius. Taip pat tyrimo pradžioje atliktas duomenų apdorojimo žingsnis leidžia apskaičiuoti pirmos eilės integruotus procesus visoms naudojamoms laiko eilutėms pasinaudojant (6) formule. Integruotas akcijos kainos procesas yra akcijos grąžos procesas, parodantis, kaip kaina kito per prekybos dienas.

Literatūros analizės metu buvo identifikuoti statistiniai modeliai, gebantys modeliuoti laiko eilutes. Parinkti modeliai plačiai naudojami įvairiose mokslo ir verslo srityse, kur siekiama nustatyti proceso savybes ar prognozuoti jų reikšmes ateityje. Jie taip pat taikyti ir akcijų kainų prognozavime. Tačiau ankstesniuose tyrimuose sukurtų modelių negalima priskirti tiriamajam procesui, nes skirtingi finansiniai procesai pasižymi skirtingomis savybėmis ir dėl to kiekvienam procesui reikia sukurti naujų modelių. Šio tyrimo metu parinkti modeliai tiriamoms laiko eilutėms atsižvelgiant į statistinius modelių rodmenis ir analizuojant jų liekanų savybes. Statistinių modelių atrinkimo procesas apibendrintai pavaizduotas 2.2 pav. pateiktoje veiklos diagramoje.

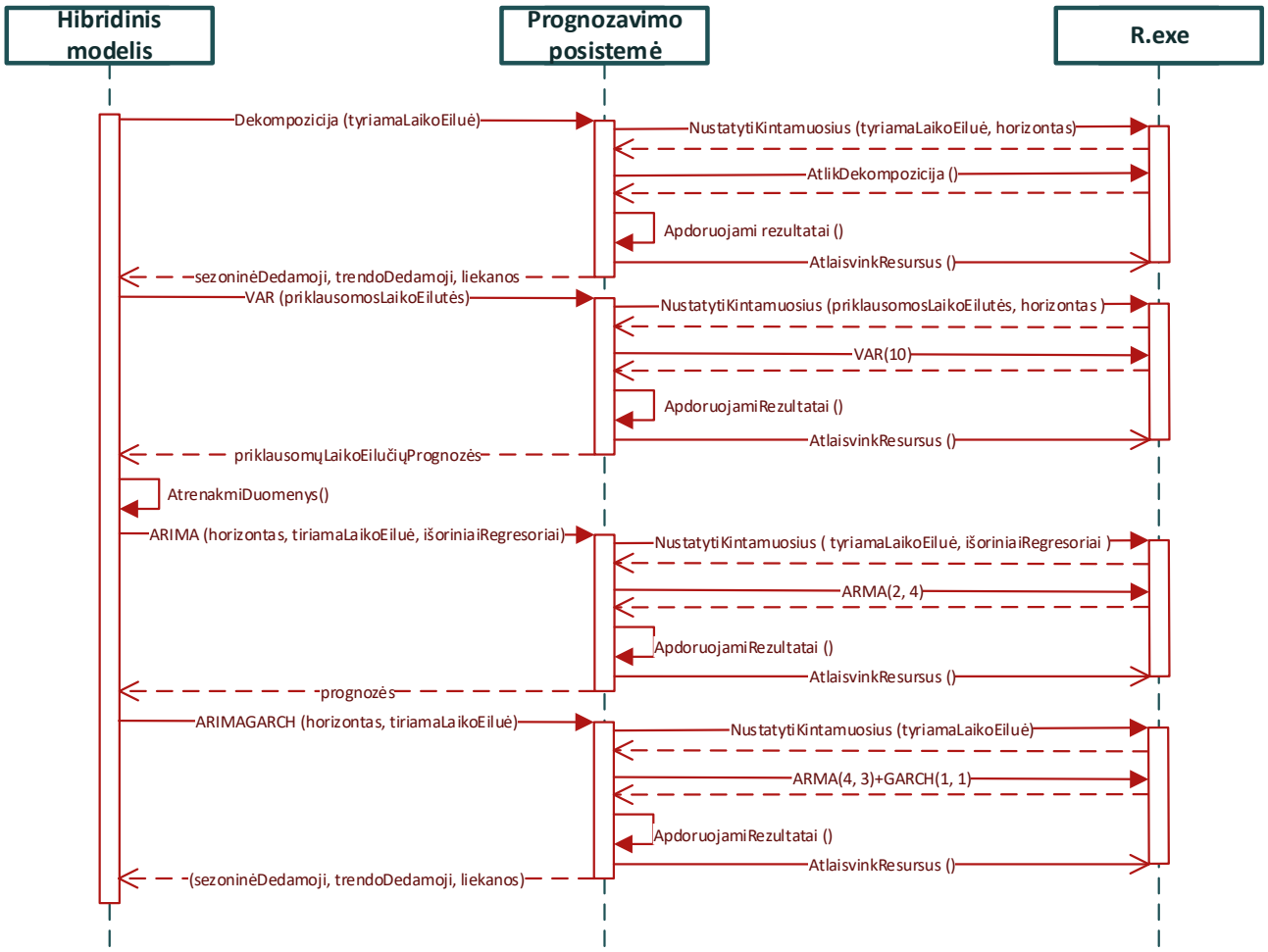
Pirmiausia, yra atrenkami duomenys, tinkantys modeliui. VAR modeliui parenkami duomenys, kurie turi tiesioginę priklausomybę tiriamam procesui, o AIMA atveju parenkami išoriniai regresoriai. Atsižvelgiant į reikalavimus, keliamus kiekvieno modelio duomenims, taip pat reikia parošti duomenis, pagal kuriuos atliekami laiko eilučių integravimo sezoninės komponentės šalinimo ar duomenų norminimo procesai.

Kiti du žingsniai tarpusavyje yra susiję. Dažnai parametrai yra parenkami iteraciniu procesu naudojantis AIC, BIC informaciniais kriterijais ar MSE, MAE (vidutinė kvadratinė ir absoliutinė paklaidos atitinkamai) juos minimizuojant. Norint pasiekti geriausių rezultatų reikia išmėginti visas įmanomas kuriamų modelių variacijas, tačiau to padaryti neįmanoma ir todėl taikomi įvairūs klasikiniai automatizuoti euristiniai modelių atrinkimo metodai. Tačiau praktikoje, duomenų analitikos srityje, automatinis parametrų parinkimas yra naudojamas tik kaip pagalbinis instrumentas. Parametrai dažniausiai yra parenkami analizuojant statistinio modelio parametrus bei metrikas. Šio darbo metu parametrų pirminiam parinkimui yra taikomi 2 iteraciniai metodai. Pirmuoju metodu išbandomos visos galimos parametrų kombinacijos nurodytame intervale. Intervalas parenkamas atliekant pradinių duomenų analizę. Iteracijų metu sukuriama statistinių įvarčių matrica, analizuojama, kaip įvarčių kitimas priklauso nuo parametrų kombinacijų. Analizuojant rezultatus galima nuspręsti, ar reikia ištirti daugiau kombinacijų. Antrasis metodas vadinamas parinkimas į priekį (angl. *forward selection*) [33]. Šis metodas naudojamas regresoriams atrinkti. Metodas kiekvienos iteracijos metu parenka vieną iš galimų regresorių, kurio įdėjimas į modelį yra naudingiausias (dažnai remiamasi BIC arba AIC statistika), procesas kartojamas, kol pasirinkta statistika nebedidėja arba nelieka regresorių. Iteracinio proceso metu parenkami kandidatai, šie kandidatai toliau kitame žingsnyje renkami pasitelkiant medalių liekanų analize, tikrinama, ar paklaidų procesas yra balto triukšmo procesas, bei tiriamas koeficientų statistinis reikšmingumas. Remiantis šiais duomenimis identifikuojami ir kuriami kiti modelio variantai, kol randami tinkami modelio parametrai, arba nustatoma, jog modelis netinka tiriamajam procesui.



2.2 pav. Modelių kūrimo procesas

Tyrimo pradžioje sukurti ir ištirti 4 (neskaičiuojant tarpinių variacijų) modeliai. 2.3 pav. pateikta prognozių generavimo segmento sekų diagrama. Šis segmentas pakartotinai naudojamas tolimesniame darbe. Šioje, kaip ir kitose, diagramose visi pavadinimai yra sulietuvinti ir todėl tiesiogiai neatitinka funkcijų ar duomenų struktūrų pavadinimų, pateiktų kodo segmentuose esančiuose prieduose. Pateiktoje diagramoje matyti, jog pirmiausia yra atliekama proceso dekompozicija panaudojant eksponentinį glodinimą. Šio proceso metu sukurtos tiriamo proceso dedamosios yra panaudojamos kuriant kitus modelius. LSTM modelis šioje sekų diagramoje nenurodytas, siekiant struktūrizuoti tolimesnes diagramas.



2.3 pav. Prognozių generavimo segmentas

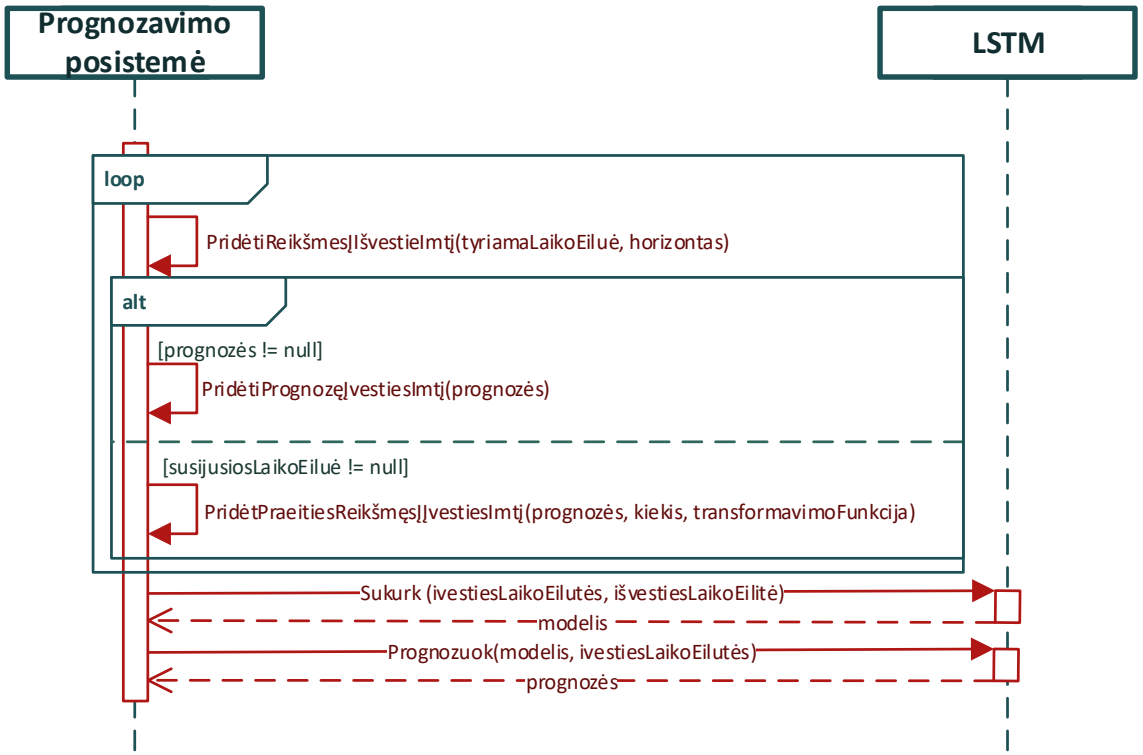
Pirmasis parinktas modelis yra vektorinis autoregresijos modelis. Šis modelis pasirinktas, nes jis geba sumodeliuoti daug tarpusavyje susijusių procesų ir surasti jų priklausomybes. Be to, identifikuotos tiesinės priklausomybės tarp modelių suteikia papildomos informacijos apie įmonių tarpusavio sąveiką ir leidžia vienu metu prognozuoti visus į modelį įtrauktus procesus. Parinktas modelis VAR(10) yra su šešiais į modelį įtraukiamais procesais (žiūrėti skyrių 3.1.4).

Antrasis panaudotas modelis sezoninis ARIMA. Šis modelis parinktas, nes tiriant *^GSPC* indekso kitimo laike savybes panaudojus ACF bei PAC identifikuoti praeities reikšmių poveikiai dabarties reikšmei bei sezoninės struktūros egzistavimas. Tačiau tolimesni tyrimai parodė, jog sezoninės komponentės įtraukimas į modelį nėra naudingas (Žiūrėti skyrių 3.1.2). Be to, nuspręsta, jog norimas prognozių rezultatas yra ne kaina, o kainos pokyčiai, todėl parinktas modelis ARMA (2, 4), kuris apmokomas jau integruotais duomenimis. Siekiant toliau pagerinti modelio savybes dar papildomai į modelį įtrauktas neapdorotos naftos kainos kitimo procesas kaip išorinis regresorius. Šio proceso modeliavimas yra už šio darbo apimties ribų. Darbe jo reikalingos n (pasirinkto horizonto) dienų į priekį prognozės yra suprojektuojamos VAR modelio.

Trečiasis pasirinktas modelis yra ARMA+GARCH. Šis modelis pasirinktas dėl gebėjimo modeliuoti sąlyginį heteroskedastiškumą. Sąlyginis finansinių duomenų nepastovumas yra gana gerai žinomas ir tyrimais patvirtintas procesas [34]. Šis modelis dažnai geba parodyti geresnius rezultatus nei ARMA,

tačiau šiame darbe bus naudojamos ne tik šio modelio tiriamo proceso prognozavimo savybės, o ir ^GSPC indekso sąlyginės sklaidos prognozės kitame žingsnyje.

Paskutinis pasirinktas modelis LSTM. Tai yra trumpalaikės, ilgalaikės atminties DI modelis. Šis modelis yra dažnai pasirenkamas, modeliuojant procesus kurie vystosi laike. Norint sukurti modelį pirmiausia yra apdorojamos duomenų imtys (2.4 pav.): į įvesties imtį yra įtraukiami praeities faktoriai, kurie leidžia prognozuoti išvesties aibę (tiriamos laiko eilutės reikšmės per n (horizontas) dienų ateityje). Kuriant įvesties aibę algoritme, vartotojui leidžiama nurodyti, kiek praeities reikšmių įtraukti (kiekis), ar jas reikia transformuoti (transformavimo funkcija). Transformavimo funkcija nurodama iškviečiant prognozavimo posistemės LSTM funkciją, taip galima įtraukti praeities reikšmes, jų vidurkį ar sumą į įvesties imtį. Tyrimo metu nustatyta, jog ši DNT architektūra nesugeba tinkamai suprojektuoti akcijų kitimo proceso savybių (žiūrėti skyrių 3.1.5), tačiau sugeba konkuruoti su kitais modeliais.



2.4 pav. LSTM prognozių generavimo segmentas

Parinkus optimalius modelių parametrus tiriamajam procesui, buvo atlikta jų ateities prognozių tyrimas. Tyrimo metu duomenys dalijami į dvi grupes: mokymo imtis sudaro 90% ir prognozavimo imtis – 10%. Atliktas kryžminis modelių tikrinimas, kai, apmokius tiriamą modelį ir panaudojus apmokymo imtį, atliekama n dienų į priekį prognozė. Pirma prognozuota reikšmė įtraukiama į apmokymo imtį, ir šis procesas kartojamas, kol prognozavimo horizontas yra mažesnis už testavimo imtį [35]. Surinkus kryžminio patikrinimo rezultatus, jie yra toliau analizuojami: kiekvieną dieną buvo surasta geriausia prognozė remiantis mažiausios absoliutinio nuokrypio nuo prognozuojamos kainos pokyčio metrika. Grafiškai šio tyrimo rezultatai rodo, jog modelių prognozių pasiskirstymas analizuojamame laiko intervale nėra atsitiktinis. Parinkti modeliai skirtingais laiko periodais geba atskleidžia tikslesnes prognozes. Ankstesniuose tyrimuose mokslininkai, siekdami pagerinti finansinių

proceso prognozavimo tikslumą, tyrė įvairių modelių sistemų kombinacijas, kur kelių modelių prognozės yra tiesiogiai transformuojamos (dažnai paprastuoju aritmetiniu vidurkiu, žiūrėti skyrių 1.8), siekiant sumažinti vidutinę prognozės paklaidą. Šio darbo metu, remiantis nustatyta modelių prognozavimo charakteristika, pasiūlytos trys architektūros, kuriomis siekiama tirti šią struktūrą.

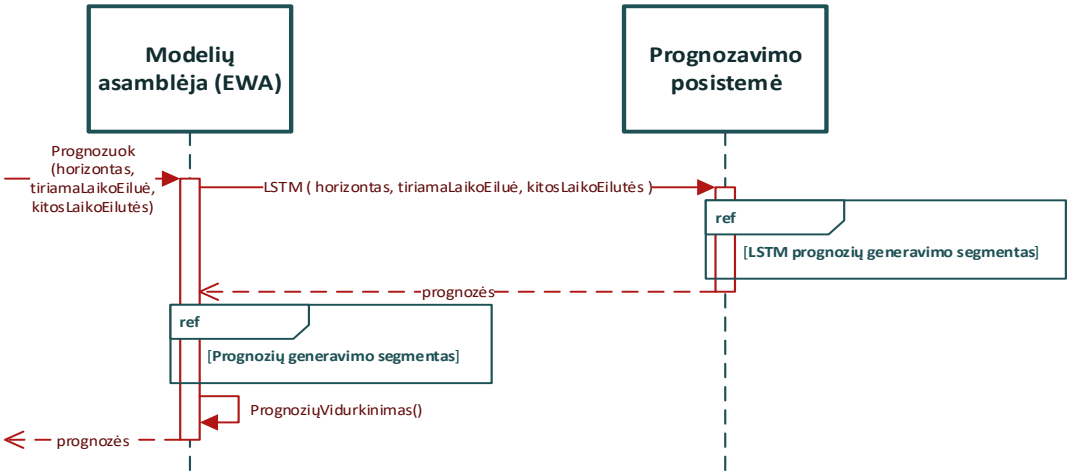
Paskutiniame tyrimo etape visi sukurti modeliai palyginami, analizuojant jų prognozių charakteristikas. Šiame etape, kaip ir ankstesniame, panaudotas kryžminis modelių tikrinimas, tačiau panaudota mažesnė mokymo imtis ir didesnė prognozavimo (33%). Norint objektyviau ištirti skirtingų modelių savybes, sukurta investavimo į rinką simuliacija panaudojant istorinius duomenis bei modelių prognozes.

2.2 Modelių asamblėjos

Remiantis gautais tarpiniais rezultatais, tiriant rinkos indeksą, panaudojant statistinius bei DNT modelius, pasiūlytos trys hibridinių modelių architektūros, skirtos suprojektuoti trumpalaikėms (vienos dienos į priekį) rinkos prognozėms. Atliekant tiriamų modelių analizę, nustatyta, jog nors ir visi modeliai buvo optimizuojami, nuspėtos to paties proceso reikšmės, bet jie turi savybių, suteikiančių pranašumų skirtinguose laiko perioduose. Šios charakteristikos buvo tikėtasi perrenkant modelius. Analizuojant rinkos bei finansinių procesų charakteristikas, nesunku pastebėti, jog tai dinamiški procesai. Tų pačių įvykių poveikis procesui vystantis irgi kinta. Priežastys gali būti įvairios: įmonė perkelia savo padalinius į kitas šalis, tokiu atveju geopolitinių veiksnių poveikis gali padidėti ar sumažėti; didėjant įmonės populiarumui, kituose procesuose atsiranda impulsų (pavyzdžiui, konkurencingos kompanijos naujo produkto išleidimas), kurių poveikis tiriamajam procesui tipiniu atveju pasireikšdavęs po *n* dienų pasireikš anksčiau po *n - delta* dienų ir, atvirkščiai, mažėjant įmonės populiarumui autoregresiniai poveikiai pasireikš po *n + delta* dienų. Šis pastebėjimas leistų daryti prielaidą, jog tiesioginiai modeliai šiam uždaviniui netinka. Tačiau, kaip teigia G. Boxas [36], „iš esmės visi modeliai yra klaidingi, tačiau kai kurie yra naudingi“.

2.2.1 Modelių asamblėja (EWA)

Primoji pasiūlyta architektūra atitinka sukurtų modelių prognozių vidurkius. Kadangi visi statistiniai modeliai yra neteisingi, o yra aproksimacijos tiriamų tiek jų parametrų, tiek pačių modelių, parinkimo procese lieka neapibrėžtumas. Į šia elementarią modelių savybę būtina atsižvelgti darant bet kokias prielaidas ar prognozes. Laiko eilučių analizėje apibrėžtumo mažinimas gali būti mažinamas kuriant modelių asamblėjas, sujungiant naudojamų modelių prognozes. Praktikoje dažnai yra naudojamas vidurkis. Vidurkio svoriams parinkti yra sukurta daug algoritmų: vienodų svorių aritmetinis vidurkis (angl. *equal weights averaging*, EWA), Bateso-Grangerio vidurkis (BGA), AIC ir BIC, pagrįstas AICA ir BIC atitinkamai vidurkiniams, ir kiti [37]. Šio tyrimo metu yra naudojamas aritmetinis sukurtų modelių vidurkis (algoritmo sekų diagrama pateikta 2.5 pav.).

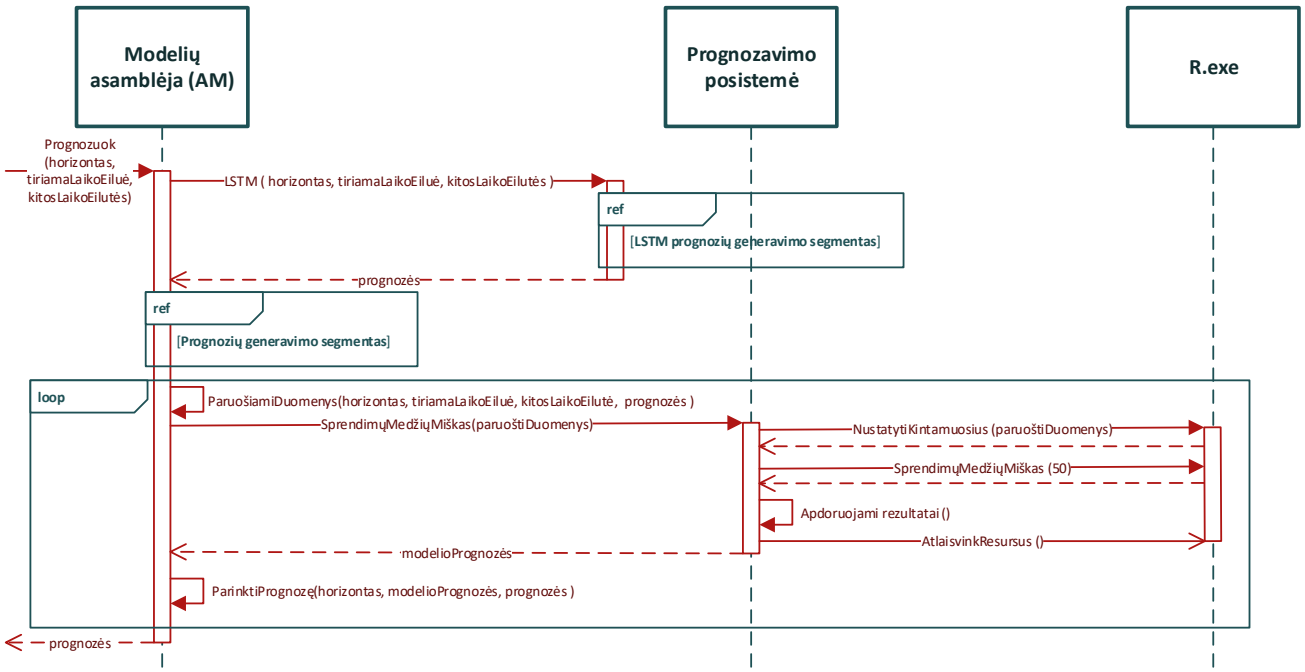


2.5 pav. Modelių asamblėjos (VSV) sekų diagrama

2.2.2 Modelių asamblėja (AM)

Remiantis prielaida, kad tiriamų modelių tikslumas yra sąlyginis, uždavinį galima traktuoti ne kaip laiko eilučių analizę, o klasifikavimo uždavinį. Sugeneravus skirtingų modelių prognozes, kiekvienai tiriamo proceso ateities reikšmei galima priskirti klasę (modelio pavadinimą), kuris laiko momentu *t* - 1 prognozavo akcijos reikšmę *t* momentu su mažiausia absoliutine paklaida. Tokiu būdu iš 4 (modelių skaičius) kiekybinių tolydžiųjų kintamųjų sudaromas vienas kategorinis kintamasis. 2.6 pav. patiekta šios architektūros sekų diagrama. Pirmiausia, sukuriamos ateities prognozių imtys pasinaudojant 50 % mokymo imtimi kiekvienam tiriamam horizontui. Šios imtys toliau apdorojamos, kiekvienam horizontui sukuriant kategorinį kintamąjį (apibūdintas anksčiau). Panaudojant sukurtą kintamąjį bei apmokymui skirtų duomenų imtį sukuriama nauja mokymo imtis, kurios dydis atitinkamai yra lygus 50 % imties dydžio.

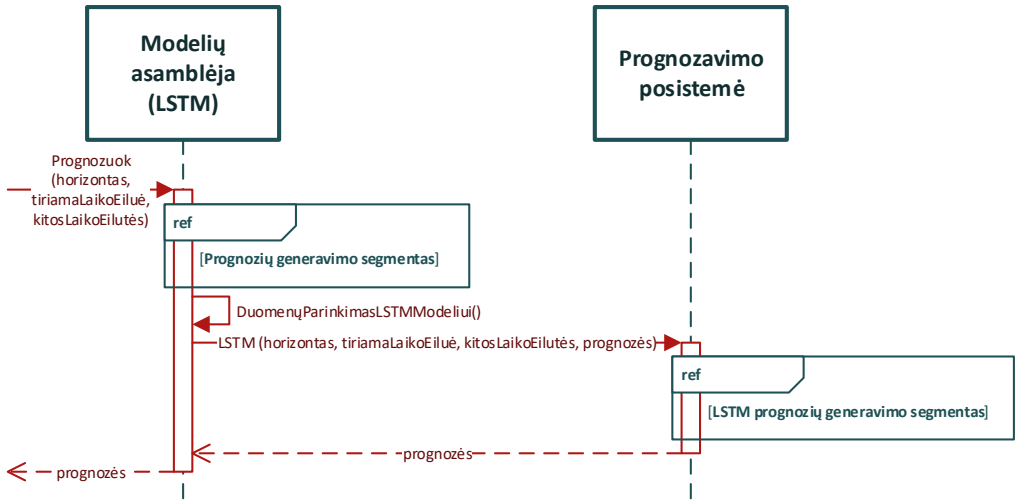
Atsitiktinis miškas pasirinktas šiems duomenims klasifikuoti, kadangi atrinktų faktorių (priklausomų procesų, aprašytų skyriaus pradžioje) ir sukurtų procesų (sąlyginės sklaidos, sezoninės komponentės) kiekis yra didelis. Be to, atlikti bandymai tiriant sukurtą duomenų imtį (panaudojant sprendimo medžius) neparodė gerų rezultatų. Panaudojant mokymui skirtą duomenų imtį buvo parinktas miško dydis 30 su atsitiktinai parenkamų charakteristikų kiekiu 3 kuriant kiekvienam medžiui. Pasiektas aukštas 92% tikslumas nuspėjant tinkamą modelio mokymo imtį.



2.6 pav. Modelių asamblėjos (AM) sekų diagrama

2.2.3 Modelių asamblėja (LSTM)

Kaip ir minėta 1.5.1 paragrafe, šis modelis itin tinka laiko eilutėms prognozuoti, nes pasižymi savybėmis, leidžiančiomis jam identifikuoti harmonikas, pasikartotinus duomenų šablonus. Atlikus tyrimus su visa duomenų imtimi nustatyta, jog šis modelio tikslumas tiriant ateities kainas yra mažiausias lyginant su statistiniais modeliais, tačiau analizuojant prognozių tikslumą taip pat pastebėta, jog jis skirtinguose laiko intervaluose sugeba konkuruoti su kitais modeliais, nors ir vidutiniškai pralaimi. Dėl anksčiau minėtų savybių modelį nuspręsta naudoti kaip apibendrinantį statistinių modelių prognozes. Ši architektūra pavaizduota 2.7 sekų diagramoje.



2.7 pav. Modelių asamblėjos (LSTM) sekų diagrama

Šios architektūros prototipo kūrimo metu taip pat atliktas modelio parametrų ir įvesties duomenų optimizavimas. Šio modelio parametrams nustatyti nenaudotas iteracinis procesas, nes šio modelio kryžminis patikrinimas trunka nuo 14 min. iki 6 h. priklausomai nuo parinktų parametrų. Todėl modelis yra kuriamas remiantis ankstesniais tyrimais, atliktais šio darbo metu, bei tarpiniais rezultatais. Optimizuojant modelio parametrus buvo minimizuotos įvertintos prognozių paklaidos ir tinkamo ženklo atpažinimo metrika. Ženklas – pokyčio kryptis (kainos ar indekso pakilimas arba kritimas). Ši metrika pasirinkta analizuojant investavimo į rinką simuliacijos rezultatus (žiūrėti skyrių 3.3.3). Į mokymui skirtų savybių imtį įtraukti duomenys, analizės metu išskirtos procesų savybės, ateities prognozės bei ^GSPC proceso praeities reikšmės. Norint pagerinti harmonikų atpažinimą taip pat atliktos 2 duomenų transformacijos operacijos. Pirmoji praeities reikšmių sumavimo:

$$suma(x,i) = \sum_{t=1}^{i} x_{t-i}. \tag{17}$$

Kita ženklų sumos operacija:

$$\check{z}enklas(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}; \tag{18}$$

$$\check{z}enklų\ suma\ (x,i) = \sum_{t=1}^{i} \check{z}enklas(x_{t-i}); \tag{19}$$

čia *x* – transformuojamo proceso reikšmės, *t* – laiko momentas, o *i* – reikšmių kiekio parametras. Kaip ir ankstesniam modeliui, taip ir šiam, apsimokymo imtis dalijama į dvi dalis panaudojant pirmą dalį sukuriamos prognozės visoms antros dalies reikšmėms. Tuomet pasinaudojant šiomis prognozėmis, bei kitomis laiko eilutėmis suformuojama nauja apsimokymo imtis LSTM modeliui. Pilnas įvesties duomenų aprašymas:

- transformuotos procesų reikšmės panaudojant (17) formulę, kur $i \in \mathbb{N} : p \in [1, 15]$, o *x* proceso reikšmės;
    - integruotas ^*GSPC* indekso vidutinės dienos vertės pokyčio procesas;
    - integruotas ^*GSPC* indekso aukščiausios dienos vertės pokyčio procesas;
    - integruotas ^*GSPC* indekso žemiausios dienos vertės pokyčio procesas;
- transformuotos procesų reikšmės panaudojant (19) formulę, kur $i \in \mathbb{N} : p \in [1, 15]$, o *x* proceso reikšmės;
    - integruotas ^*GSPC* indekso vidutinės dienos vertės pokyčio procesas;
    - integruotas ^*GSPC* indekso aukščiausios dienos vertės pokyčio procesas;
    - integruotas ^*GSPC* indekso žemiausios dienos vertės pokyčio procesas;
- keturių dienų į priekį prognozės (kiekvienai dienai skiriama prognozė) ;
    - ARMA+GARCH ^*GSPC* proceso sąlyginio heteroskedastiškumo prognozės;
    - ARMA+GARCH ^*GSPC* proceso ateities prognozės;
    - ARMA ^*GSPC* proceso ateities prognozės;
    - VAR ^*GSPC* proceso ateities prognozės.

2.3 Modelių palyginimo metodai

Egzistuoja 2 pagrindiniai būdai, kuriais pasinaudojant galima gauti pelną atliekant prekybą akcijomis. Pirmasis būdas yra pelno gavimas iš kainų augimo. Šio metodo metu akcijos įsigyjamos už žemą kainą ir parduodamos už aukštesnę. Antrasis būdas yra pelno gavimas iš kainų kritimo. Šio proceso metu akcijos yra pasiskolinamos iš akcijų savininko, jos yra parduodamos už didelę kainą. Po sutarto laiko žmogus turi pasiskolintą akcijų kiekį grąžinti jų savininkui, nupirkus jį už mažesnę nei pardavimo kainą gaunamas pelnas. Remiantis šiomis sąlygomis modeliai yra lyginami atsižvelgiant į tai, ar suprojektuotas kainos pokytis yra naudingas vienam ar kitam metodui. Kadangi yra prognozuojama vidutinė dienos indekso S&P 500 reikšmė, jos prognozė turi patekti į kitos dienos indekso kitimo rėžius: indekso didžiausia dienos reikšmė ≥ prognozuota reikšmė ≥ mažiausia indekso dienos reikšmė. Šios tinkamos prognozės toliau skaidomos į tikslingai atpažintus kainų pakilimus ir kritimus. Be šio kriterijaus, taip pat yra vertinama ir absoliutinė paklaida.

2.3.1 Prekybos rinkoje simuliacija

Norint dar objektyviau įvertinti pasiūlytų modelių tinkamumą duomenims bei panaudos atvejus, taip pat sukurta prekybos akcijomis simuliacija bei strategija (pavadinta *PirkIrParduok*), kuri panaudojant prognozių rezultatus atlieka finansinius sprendimus. Kadangi tiriamas procesas reprezentuoja rinkos veiklą, investavimą į šį indeksą galima traktuoti kaip pirkimą visų į indeksą įeinančių akcijų. Žemiau pateiktas simuliacijos pseudokodas. Pradinėje simuliacijos stadijoje akcijos yra nenupirktos ir pelnas yra lygus vienam (2 – 3 eilutės). 4 eilutėje pradedamas pagrindinis simuliacijos ciklas, pradedama nuo 2 reikšmės, nes, norint įeiti į rinką (nusipirkti akcijų), reikia žinoti praeities prognozę, kuri neegzistuoja *pc* imtyje *i*-tąja diena, kai *i* lygu 1. 5 eilutėje pradedama sprendimų seka, jei akcijos *i* – tuoju laiko momentu nėra nupirktos. Pirkimas įvyksta tada, jei akcijų kainos (indekso reikšmė) šios dienos ribose yra didesnė ar lygi ir ne mažesnė už vakarykštės dienos reikšmės prognozę. Tokiu būdu užsakoma nupirkti akcijas už prognozuotą kainą, žinoma, akcijos nebus nupirktos, jei jų reikšmės (atitinka indekso reikšmę) nekis dienos laikotarpyje nurodytose ribose (eilutė 7). Jeigu akcijos yra nupirktos, tai jos yra parduodamos, jeigu prognozė buvo teisinga ir šios dienos laikotarpyje akcijų kainos pasiekia šią prognozę, tai didžiausia dienos reikšmė ≥ prognozuota reikšmė ≥ mažiausia indekso dienos reikšmė (eilutės 12-16). Jeigu akcijos buvo neparduotos už suprognozuotą kainą, jos yra parduodama už rinkos kainą dienos pabaigoje, jei rytojaus kainų prognozė (kurią galima suprojektuoti dienos pabaigoje, kadangi jau žinomi visi duomenys, reikalingi atlikti naujai prognozei) yra mažesnė. Darbo dienos pabaigoje atsiranda vėl galimybė įsigyti akcijas, jei jos nebuvo nupirktos dienos laikotarpyje pasinaudojant vakarykšte prognoze; arba akcijos buvo parduota pasinaudojant vakarykšte ar rytojaus prognoze. Akcijos yra nuperkamos, kai rytojaus kainos prognozė yra didesnė už dabartinę indekso reikšmę, rinkos uždarymo reikšmę (eilutės 24-31).

```
1.  funkcija Prekiauti akcijomis(ap, lp, hp, cp, pp)
        Įvesties reikšmės
            ap – vidutinių indekso dienos reikšmių rinkinys
            lp – žemiausių indekso dienos reikšmių rinkinys
            hp – aukščiausių indekso dienos reikšmių rinkinys
            cp – indekso reikšmės biržos uždarymo metu rinkinys
```

pc – rytojaus indekso reikšmių prognozių pokyčių rinkinys

Rezultatų reikšmės

inv – pelnas

Lokalūs kintamieji

b – ar akcijos nupirktos

bp – akcijų nupirkimo kaina

pp – prognozuojama kaina

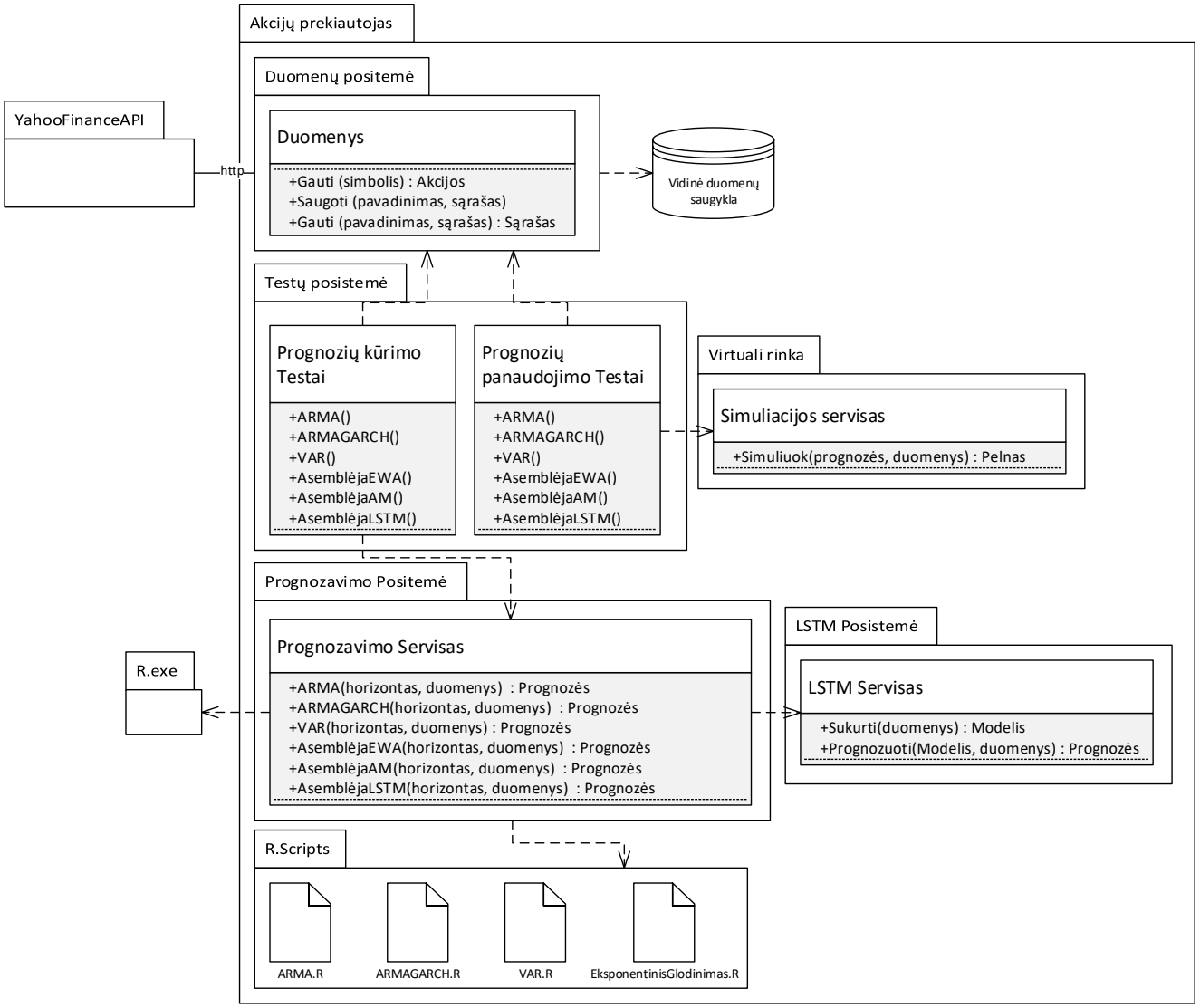
2. *b* = **NE**
3. *inv* = 1
4. **kiekvienai** *i* = 2 iki Ilgis(*ap*) **pradėti**
5. **jai** *b* = **NE** ir *pc*[*i* - 1] ≥ 0 **tada**
6. *pp* = *ap* · (1 + *pc*[i-1])
7. **jai** *pp* >= *lp*[i] ir *pp* ≥ *lp*[i] **tada**
8. *b* = **TAIP**
9. *bp* = *pp*
10. **baigti**
11. kitu atveju
12. *pp* = *ap*[*i* - 1] · (1 + *pc*[*i* - 1])
13. **jai** *pp* ≥ *lp*[*i*] ir *pp* ≤ *lp*[*i*] **tada**
14. $inv = inv \cdot \left(1 + \frac{pp - bp}{bp}\right)$
15. *b* = **NE**
16. kitu atveju
17. *pp* = *ap*[*i*] · (1 + *pc*[i])
18. **jai** *pp* < *cp*[*i*] **tada**
19. $inv = inv \cdot \left(1 + \frac{cp[i] - bp}{bp}\right)$
20. *b* = **NE**
21. **baigti**
22. **baigti**
23. **baigti**
24. **jai** *b* = **NE** ir *pc*[*i*] ≥ 0 **tada**
25. *pp* = *ap* · (1 + *pc*[i])
26. **jai** *pp* > *cp*[i] **tada**
27. *b* = **TAIP**
28. *bp* = *cp*[i]
29. **baigti**
31. **baigti**
32. **baigti**

```
33.     grąžinti inv
34. baigti
```

2.4 Tyrimo aplinka bei naudojamos technologijos

Šio tyrimo rezultatas yra pasiūlytos modelių asamblėjų architektūros bei statistinių modelių pritaikymas (parametrų parinkimas) ^GSPC indeksui prognozuoti. Šiam tikslui pasiekti kuriama sistema, kurios pagrindinė užduotis – sutelkti aplinką ir funkcijas, kurias panaudojant galima būtų kurti ir palyginti skirtingus rinkos dinamikų modelius. Šiam uždaviniui spręsti reikia įvairių analitinių resursų, DNT realizacijų bei programavimo aplinkos. Šiuo metu tyrimo realizacijos aplinkai teikia daug produktų. Yra sukurta aukšto lygio programavimo kalbos, tokios kaip python ar R, kurios specializuojamos duomenų analitikoje. Tačiau aukštas kalbų lygis turi ir didelę kainą – greitaveiką, be to, netinka kuriant daugiafunkcines sudėtingas sistemas. Todėl pasirinktas .NET karkasas ir C# programavimo kalba siūlomiems algoritmams realizuoti. Tačiau ne visi analitiniai resursai, reikalingi šiam projektui realizuoti, yra sukurti pasinaudojant .NET karkasu. Dėl to nuspręsta naudoti R karkasą iškviečiant jo funkcijas tiesiogiai iš kuriamos sistemos, panaudojant R.NET paketą, jei reikalingi analitiniai resursai. Pasiūlyta sistema „Akcijų prekiautojas“, šios sistemos paketų diagrama pateikta 2.8 pav. Sistema sudaryta iš tokių dalių:

- Duomenų posistemė – posistemė, skirta duomenims saugoti. Ji atlieka dvi pagrindines funkcijas. Jei kitai posistemei reikia istorinių duomenų apie akcijas ar kitus procesus, ji pirmiausia ieško duomenų, kurie yra lokalioje duomenų saugykloje. Jei jų neranda, kreipiasi į *Yahoo Finance* portalo [3]. Antra funkcija yra rezultatų saugojimas tolimesnei analizei. Tam, kad duomenis būtų lengva importuoti į kitas programas, pasirinktas primityvus duomenų saugojimo formatas – duomenys yra sudaryti iš kiekybinių reikšmių sąrašų, kurie saugomi csv formatu.
- Testų posistemė – sąsaja su kitomis posistemėmis. Testais atliekami visi reikalingi darbai, reikalingi darbui realizuoti. Šioje posistemėje parenkami modelių parametrai ir pasinaudojant kitomis posistemėmis atliekami tyrimai.
- Prognozavimo posistemė – šioje posistemėje atliktos pasiūlytų modelių realizacijos. Posistemė atitinkamai nukreipiama į R karkasą, jei norima sukurti statistinius modelius bei jų prognozes. R scenarijų rinkmenose aprašytos procedūros atitinkamiems modeliams sukurti. Modelių realizacijos aprašai sukurti remiantis eksperimentiniais tyrimais, detalesnės proceso specifikacijos pateiktos kitame skyriuje.
- Virtuali rinka – posistemė, skirta testuoti prognozėms. Detalus simuliacijos aprašas pateiktas 360 skyriuje.
- Posistemėje LSTM realizuotas DNT modelis, jo mokymas bei prognozių generavimas. LSTM modelis sukurtas pasinaudojant CNTK [38] karkasą. Šios posistemės naudojimas pateiktas 2.1 skyriuje.



2.8 pav. Sistemos „Akcijų prekiautojas“ paketų diagrama

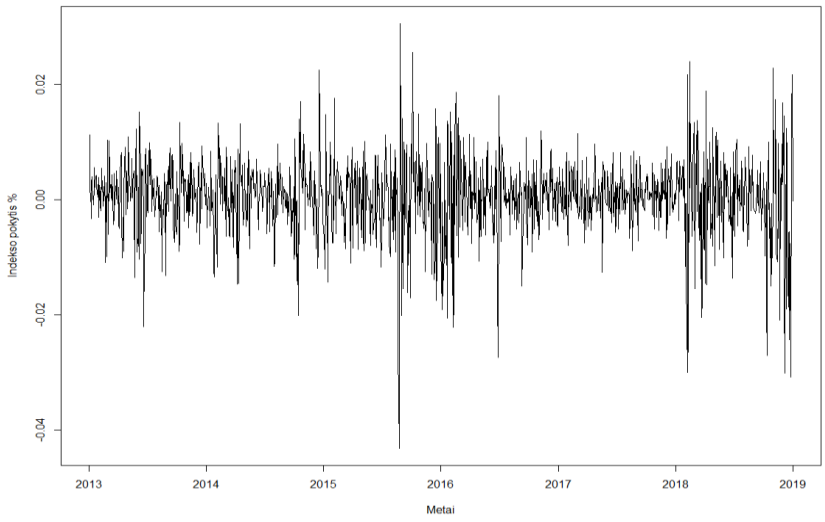
3 AKCIJŲ RINKOS MODELIAVIMO TYRIMAS

Tiriamojo darbo metu pasirinktas objektas istoriniai *^GSPC* indekso duomenys, kuriems parinkti optimalūs statistiniai modeliai bei neuroninių tinklų modeliai. Remiantis gautais rezultatais sukurtos modelių asamblėjos, siekiant išnaudoti statistinių modelių bei DNT savybes. Šio skyriaus pradžioje aptariami tyrimui pasirinkti duomenų paketai bei jų savybės. Toliau aprašomi sukurti modeliai bei jų charakteristikos. Palyginamos visų sukurtų modelių bei jų ansamblių, aprašytų ankstesniame skyriuje prognozių metrikos. Skyriaus pabaigoje pateikiami sukurtų modelių prognozių panaudojimo rezultatai investuojant į rinką sukurtoje simuliacijoje. Rezultatai taip pat lyginami su *pirk ir laikyk* investavimo strategija.

3.1 ^GSPC proceso dinamikų modeliavimo analizė

Kaip ir minėta 2 skyriaus pradžioje, tyrimui pasirinkti S&P 500 (*^GSPC*) indekso istoriniai duomenys. Tyrimo metu pasirinktos 4 į šį indeksą įeinančios kompanijos: United Continental *Holdings Inc.* (ženklinimo simbolis *UAL*), *Southwest Airlines Co.* (ženklinimo simbolis *LUV*) bei *American Airlines Group Inc.* (ženklinimo simbolis *AAL*), bei *Delta Air Lines* (ženklinimo simbolis *DAL*). Istorikai išvardintų kompanijų akcijų ir tiriamo proceso duomenys atsisiųsti iš *Yahoo finance* portalo [3]. Visos šios įmonės yra oro linijų bendrovės, šios įmonės pasirinktos, nes oro linijų bendrovių akcijos greitai reaguoja į rinką veikiančius procesus, kas gali leisti nuspėti visos rinkos elgseną artimoje ateityje. Be to, pasirinktas kaip galimai tiriamajai laiko eilutei darantis poveikį procesas – neapdorotos naftos kainos istoriniai duomenys (ženklinimo simbolis *DCOILWTICO*). Šis duomenų rinkinys parsisiųstas iš ir FRED [4] duomenų bazės. Jis pasirinktas, nes kuro kainos gali turėti poveikį visai rinkai. Trūkstamos reikšmės, kurios sudarė ~ 3,6 % duomenų imties tiriamajam laikotarpiui, sugeneruotos naudojant interpoliacijos metodą.

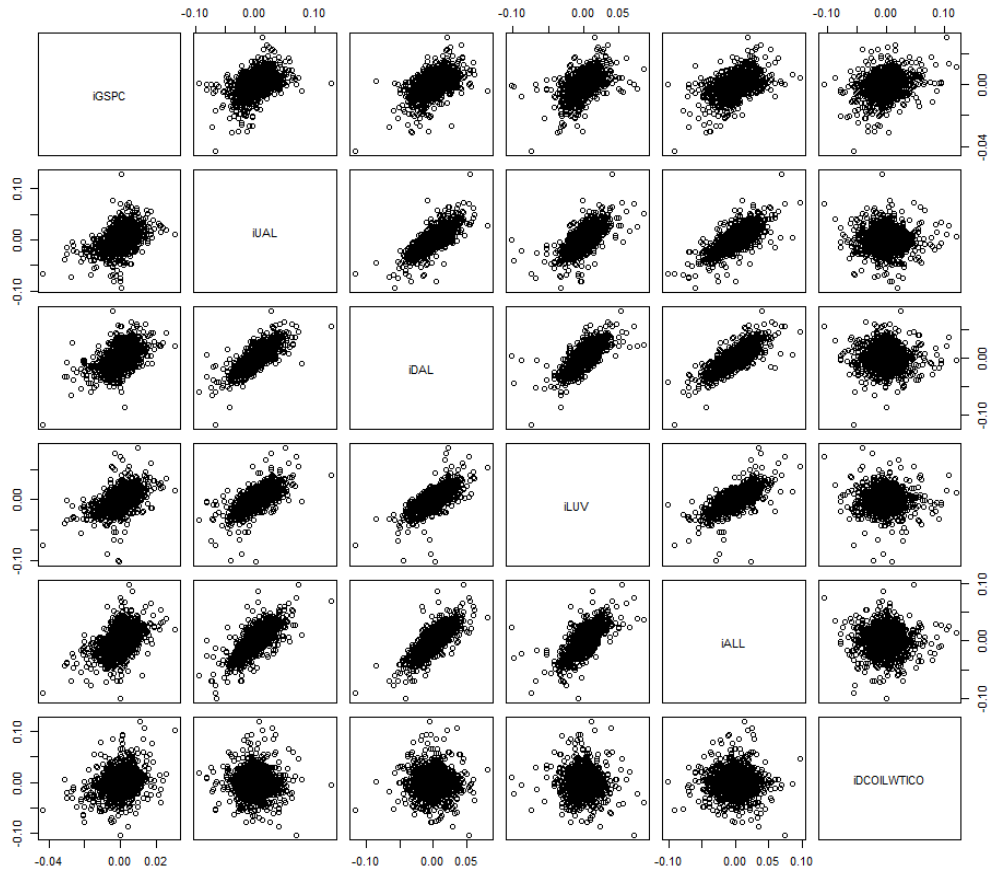
Pradiniame duomenų analizės žingsnyje reikia apskaičiuoti pirmos eilės integruotus procesus pasinaudojant (6) formule. Šis žingsnis naudingas, nes sukurti procesai turi savybių, naudingų statistinei duomenų analizei (dažnai turi pastovų vidurkį), be to, pašalinamas proceso mastelis, kas leidžia lengviau identifikuoti procesų sąveikas. 3.1 pav. pateiktas transformuotų *^GSPC* duomenų grafikas, kiti minėti duomenų rinkiniai taip pat transformuoti.



3.1 pav. *^GSPC* 1 eilės integruotas procesas, toliau žymimas *iGSPC*

Aukščiau pateiktame grafike matyti, jog dispersija laikui bėgant kinta. Matoma, jog egzistuoja periodai, atskleidžiantys didelius nuokrypius nuo vidurkio. Pavyzdžiui, nuo 2015 m. vidurio iki 2016 m. pradžios periodiškai išryškėja pasikartojančios dispersijos padidėjimas bei sumažėjimas apytiksliai kas mėnesį. Grindžiant šia prielaida galima naudoti ARMA+GARCH modelį, kuris projektuoja impulsų bei praeities reikšmių poveikį tiriamajam procesui bei geba tirti nuo laiko priklausančią liekanų dispersiją.

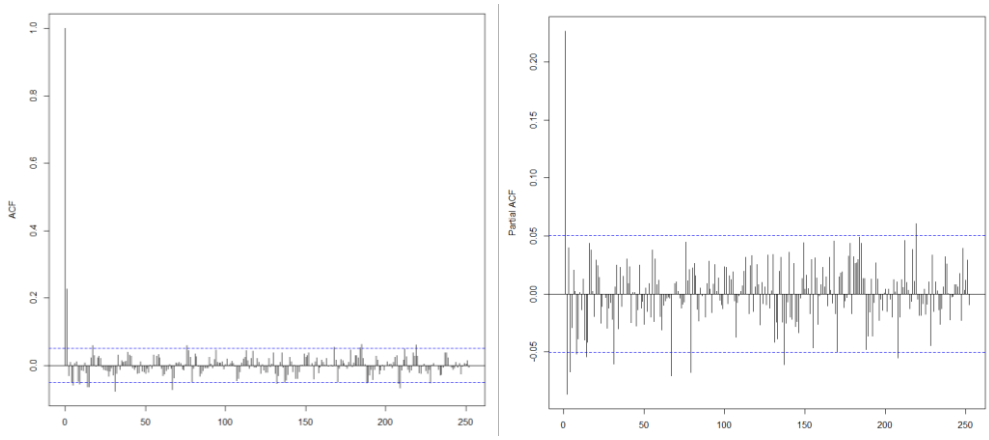
Norint pagerinti prognozių tikslumą ieškoma galimų išorinių regresorių. Pasirinkti kandidatai bei jų ženklinimo simboliai aprašyti šio skyriaus pradžioje. Nustatymui, ar egzistuoja tarpusavio ryšiai tarp identifikuotų procesų, panaudota sklaidos diagramų matrica. 3.2 pav. pateiktuose grafikuose pastebima, jog egzistuoja tiesinės priklausomybės tarp visų pasirinktų kompanijų akcijų ir S&P 500 indekso reikšmių silpna, bet statistiškai reikšminga nustatyta tiesioginės regresijos priklausomybė tarp tiriamo proceso bei *iDCOILWTICO* proceso (i žymi integruotą procesą). Šiuos sąryšius pasirinkta tirti VAR modeliu.



3.2 pav. Matrica su tiriamo proceso ir jam poveikį darančių procesų sklaidos diagrama

Norint nustatyti slenkančio vidurkio bei autoregresijos eilę, yra apskaičiuojama *iGSPC* proceso autokoreliacijos (angl. *autocorrelation function, ACF*) bei dalinės autokoreliacijos funkcijos (angl. *partial autocorrelation function,* PACF). Skaičiavimų rezultatai pateikti 3.3 pav. Gautuose grafikuose pastebima, jog yra statistiškai svarbių reikšmių, esančių už 5% intervalo tiek ACF, tiek PACF grafikuose. Naudojantis ACF grafiku pastebima, jog 2 vėlavimų reikšmė yra keliolika kartų didesnė už kitų vėlavimų reikšmes, taigi MA(1) modelis gali būti naudojamas kaip pradinis. Be to, pastebimi ir tolimesnės eilės reikšmingi įverčiai, tačiau vizuali grafikų inspekcija neleidžia nustatyti jų įtraukimo

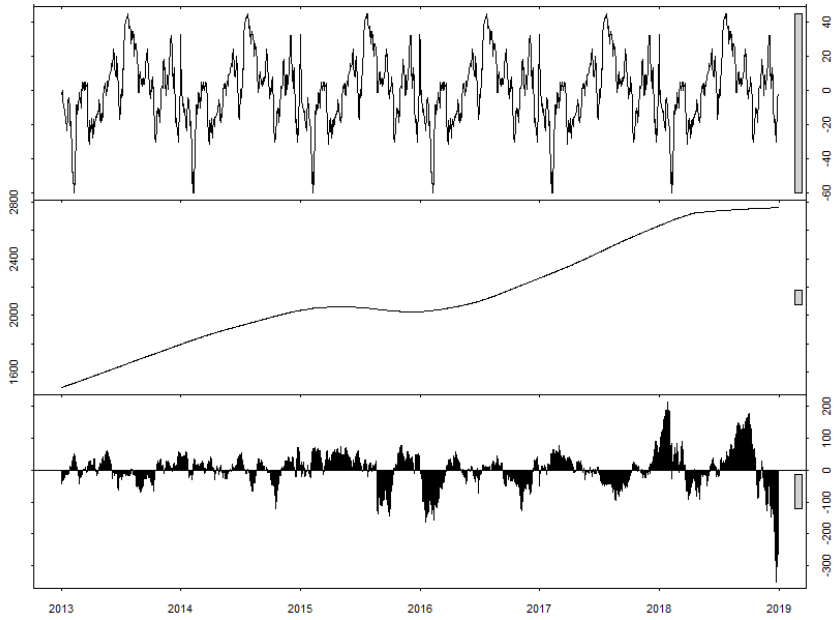
į modelį naudos. Be to, pastebima, jog jie turi periodiškumą kas 5 dienas, o tai parodo periodinės komponentės egzistavimą, tačiau iš grafiko sunku pasakyti, ar ji yra statistiškai reikšminga.



3.3 pav. Autokoreliacijos funkcijos (dešinėje) bei dalinės autokoreliacijos funkcijos (kairėje) grafikai

3.1.1 Sezoninės komponentės išskyrimas

Remiantis aukščiau pateiktais pastebėjimais, nuspręsta išskirti sezoniškumo komponentę. Praktikoje sezoniškumas yra šalinamas iš pradinių duomenų, tačiau nuspręsta jį palikti kuriant statistinius modelius. Sezoninės komponentės šalinimas nebuvo atliktas, nes sezoniškumą galima naudoti kaip išorinį regresorių, be to, ne visiems modeliams ši transformacija suteikia statistiškai reikšmingo pranašumo, nes modeliai patys gali priklausyti nuo sezoniškumo. Išskirta dedamoji taip pat bus naudojama kuriant hibridinę sistemą. Norint išskirti *^GSPC* indekso kitimo sezoninę komponentę, buvo atlikta proceso dekompozicija panaudojant slenkančio vidurkio metodą. Žemiau pateiktame grafike (3.4 pav.**)** pavaizduota *^GSPC* proceso dekompozicija. Viršutinėje dalyje pavaizduota sezoninė komponentė, vidurinėje trendas, o apatinėje liekanos, likusios po dekompozicijos.



3.4 pav. *^GSPC* kainų laiko eilutės dekompozicija

3.1.2 ARMA modelio sukūrimas

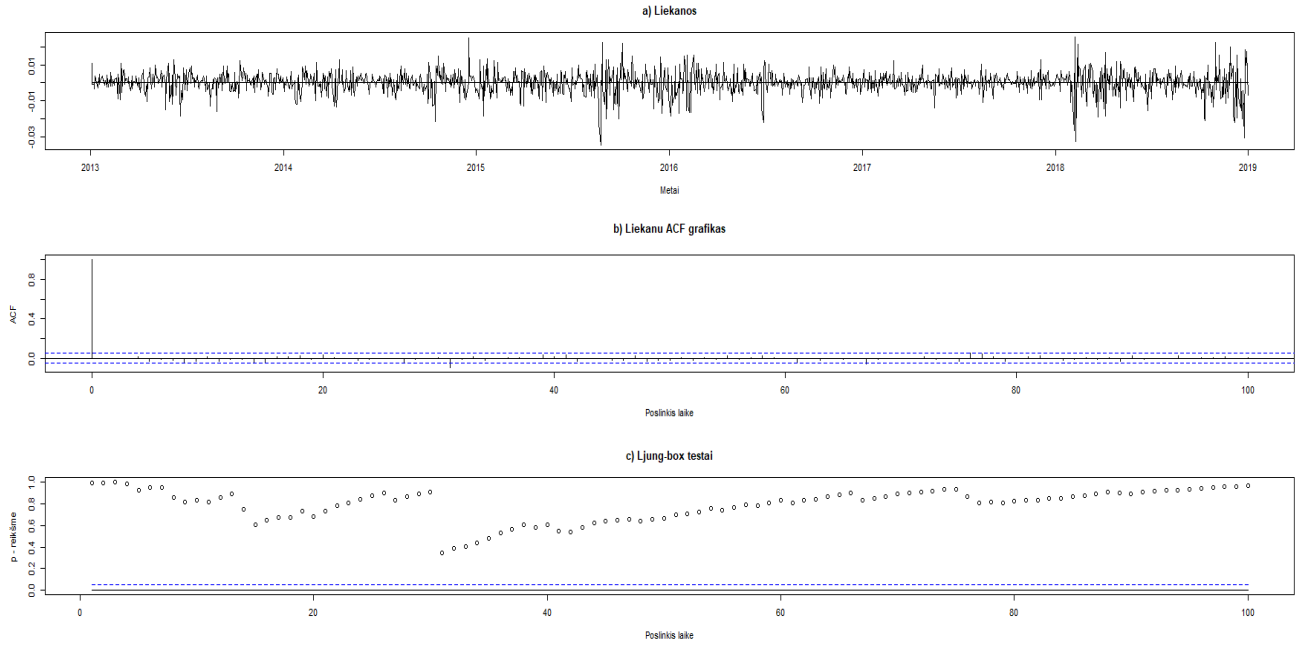
Kadangi vizuali integruoto *^GSPC* proceso analizė nepadėjo identifikuoti autoregresijos bei judančio vidurkio eilės, tik leido padaryti prielaidą, jog MA eilė 1 ar daugiau, modelio parinkimas atliekamas iteraciniu procesu. Šio proceso metu kombinuojamos modelio p ir q (žiūrėti skyrių 1.7.2) bei periodo ir sezoninės komponentės $\tilde{p}$ ir $\tilde{q}$ eilės. Tyrimo metu analizuojama, kaip keičiasi modelio statistiniai rodmenys, siekiama sumažinti AIC ir BIC statistiką bei vidutinę absoliučią (angl. *means absolute error*, MAE) bei kvadratinę paklaidas (angl. *mean squared error*, MSE). Tyrimo rezultatai pateikti 3.1 lent. Tyrimo metu sukurti 132 modeliai, lentelėje pateikti tik svarbesni modeliai, lėmę galutinio modelio parinkimą. Pirmiausia optimizuoti p ir q parametrai. Modeliuotos kombinacijos, kai $p \epsilon \mathbb{N} : p \epsilon [1, 10]$, o $p \epsilon \mathbb{N} : p \epsilon [1, 7]$. Lentelėje modeliai, kurių numeriai nuo 0 iki 87. Šio tyrimo metu atrinkti modeliai ARIMA(0,0,1), ARIMA(2, 0, 4). Pasirinktas modelinis ARIMA(2, 0, 4) tolimesniam tyrimui remiantis BIC, MAE ir MSE kriterijais. Sezoniškumo nustatymo tyrimo rezultatai lentelėje numeruojami nuo 88 iki 129. Pastebima, jog sezoniškumo periodo nustatymo testuose išsiskiria 2 periodai – 15 dienų ir 31 dienos. Kadangi prekyba akcijų rinkoje vyksta tik darbo dienomis (tipinė savaitė susideda iš 5 darbo dienų), tai 15 dienų periodą galima suprasti ir interpretuoti kaip elgseną, kuri priklauso nuo savaitės dienos, tačiau 31 dienų intervalą sunku interpretuoti ir daryti prielaidą, jog regresorių įtraukimas į modelį bus naudingas panaudojant kitokią duomenų imtį (tačiau verta paminėti, jog ši dinamika yra naudinga modeliui žiūrėti (bandymas nr. 133). Pasirinktas 15 dienų intervalas su $\tilde{p}$ eile 0 ir $\tilde{q}$ eile 1. Kadangi kuriamas vektorinis autoregresinis modelis šiame darbe, tai buvo bandomi tik 2 išoriniai regresoriai, sezoninė komponentė, kurios aprašas pateiktas 3.1.1 skyriuje, ir neapdorotos naftos kainos integruotas procesas. Ši laiko eilutė yra įtraukiama į modelį, remiantis visomis 3 statistikomis.

3.1 lentelė. Sezoninio ARIMA modelio su išoriniais regresoriais parinkimo rezultatai

Nr.	Modelis	Išorinis regresorius	MAE	MSE	AIC	BIC
0	SARIMA(0, 0, 0) x (0, 0, 0) 0		0.004558	4.19E-05	-10923.7	-10913
1	SARIMA(0 0 1) x (0 0 0) 0		0.004429	3.94E-05	-11013.7	-10997.8
20	SARIMA(2 0 4) x (0 0 0) 0		0.004424	3.90E-05	-11021.8	-10979.3
87	SARIMA(10 0 7) x (0 0 0) 0		0.004437	3.87E-05	-11009.2	-10908.1
88	SARIMA(2 0 4) x (1 0 0) 0		0.004425	3.90E-05	-11019.8	-10972
103	SARIMA(2 0 4) x (1 0 0) 15		0.004419	3.89E-05	-11023.9	-10976
119	SARIMA(2 0 4) x (1 0 0) 31		0.004423	3.88E-05	-11028	-10980.1
120	SARIMA(2 0 4) x (1 0 0) 32		0.004424	3.89E-05	-11020.7	-10972.8
121	SARIMA(2 0 4) x (1 0 0) 33		0.004417	3.89E-05	-11022.5	-10974.7
122	SARIMA(2 0 4) x (0 0 1) 15		0.004419	3.89E-05	-11024.1	-10976.2
124	SARIMA(2 0 4) x (1 0 0) 15		0.004419	3.89E-05	-11023.9	-10976
129	SARIMA(2 0 4) x (2 0 2) 15		0.004422	3.88E-05	-11019.1	-10955.3
130	SARIMA(2 0 4) x (0 0 1) 15	Sezoninė komponentė	0.004419	3.89E-05	-11024.1	-10976.2
131	SARIMA(2 0 4) x (0 0 1) 15	iDCOILWTICO	0.004321	3.63E-05	-11125.6	-11072.4

132	SARIMA(2 0 4) x (0 0 0) 0	iDCOILWTICO	0.004325	3.64E-05	-11124.6	-11076.7
133	SARIMA(2 0 4) x (0 0 1) 31	iDCOILWTICO	0.004317	3.60E-05	-11138.2	-11085

Modelio tinkamumas duomenims nustatytas analizuojant modelio liekanų savybes, tikrinama, ar liekanos yra balto triukšmo procesas, parinktų koeficientų statistinis reikšmingumas. Atlikus parinkto modelio SARIMA(2, 0, 4) x (0, 0, 1) 15 su išoriniu regresoriumi koeficientų standartizuotų paklaidų analizę, nustatyta, jog visi koeficientai yra statistiškai reikšmingi (su 95% tikimybe atmetamos nulinės hipotezės (H_o), jog koeficientai lygūs 0, panaudojant Z testą). Tačiau sezoninės komponentės z testo reikšmė apytiksliai lygi 0.08 (artima kritinei ribai 0.05). Taigi šis regresorius yra mažai reikšmingas ir yra šalinamas iš galutinio modelio, dėl to sumažėja ACF ir BIC statistikos, bet nežymiai padidėja modelio liekanų sklaida. Liekanų tyrimo rezultatai pateikti 3.5 pav. Liekanų vidurkis, lygus $1.02 \cdot 10^{-6}$, yra artimas nuliui. Pastebima, jog modelis tinka duomenims, kadangi negalima atmesti 95% garantijos, jog H_o liekanos yra nekoreliuotos, remiantis *Ljung-box* testų rezultatais. Visiems laiko intervalams negalime testų reikšmės nekerta kritinės ribos (mėlyna punktyrinė linija). ACF grafike taip pat nematyti reikšmių, didesnių nei 0.05, tai galima daryti prielaidą, jog liekanų procesas yra baltas triukšmas. Atlikę koeficientų Z testus, kur H_o hipotezės koeficientas lygus nuliui, nustatyta, jog visi modelio koeficientai prie regresorių yra statistiškai reikšmingi. Kadangi nuspręsta projektuoti kainos pokytį, kuris jau yra integruotas, ir nenaudoti modelio sezoninės dalies modeliui apibūdinti, gali naudoti trumpesnį pavadinimą ARMA(2, 4).



3.5 pav. Sukurto modelio ARMA(2, 4) su išoriniu regresoriumi liekanų analizės rezultatai

3.1.3 ARMA+GARCH modelis

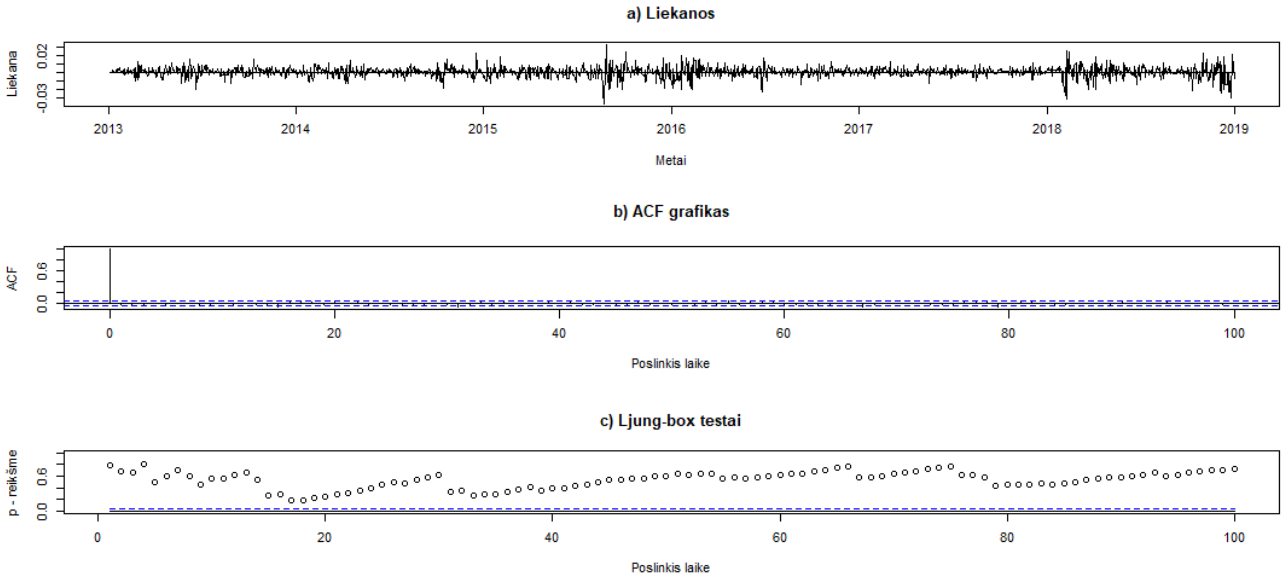
Kaip buvo minėta, *iGSPC* procese matomas dispersijos priklausomumas nuo laiko. Nors ši struktūra nėra tokia reikšminga, jog galima būtų teigti, kad *iGSPC* kitimo procesas yra stacionarus plačiąja prasme, tačiau galima panaudoti šią struktūrą gerinant proceso modeliavimo tikslumą. Tam pasitelktas autoregresinio sąlyginio heteroskedastiškumo, modelis Sąlyginė sklaida akcijų grąžų procesuose yra dažnai matomas reiškinys, nes, akcijos kainai labiau pakitus ar sumažėjus, pastebimas žmogiškojo

faktoriaus poveikis tiek akcijų pirkimo kiekiuose, tiek jos kainoje. Šiuo laikotarpiu akcijos kaina labai svyruoja, bet, praėjus tam tikram laikui, vėl nusistovi ir kinta tolygiai. Žinoma, ši dinamika pasireiškia ir išvestiniuose procesuose, šiuo atveju rinkos indekse. Šio modelio GARCH dalis taip pat yra naudojama akcijos rizikai vertinti šiame darbe kaip įvestis LSTM modeliui. Analogiškai SARIMA modeliui atliktas modelio parametrų nustatymas iteraciniu procesu. Lent. 3.2 pateikti modelio parinkimo rezultatai. Sukurti 431 modeliai. Remiantis BIC ir AIC statistika atrinkti 4 modeliai (NR.: 64, 65, 70, 71). Modeliai 65, 70 ir 71 nenaudojami, nes tolimesniuose tyrimuose bus naudojama mažesnė duomenų imtis modeliams apmokyti, dėl to aukštesnio laipsnio modeliai tampa nebestabilūs – koeficientai prie aukštesnių poslinkių gali įgauti labai dideles ar mažas reikšmes. Išsamesnė 64 modelio analizė parodė, jog jo liekanos nėra balto triukšmo procesas, todėl parinktas ARMA(4, 3) + GARCH(1, 1) modelis.

3.2 lentelė, ARMA+GARCH modelio parinkimo rezultatai

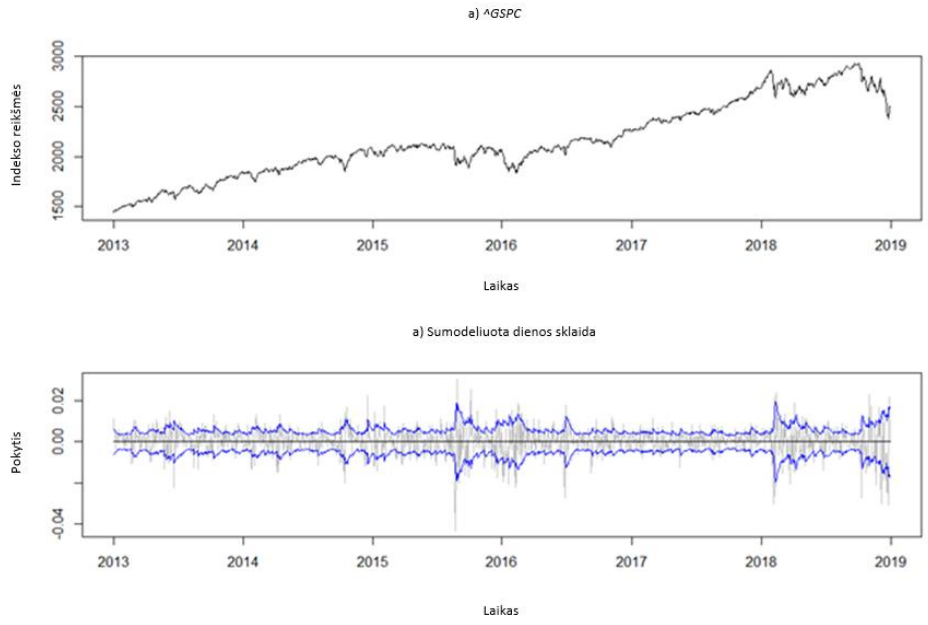
Nr.	Modelis	MAE	MSE	AIC	BIC
0	ARMA(0, 0) + GARCH(1, 0)	0.004597	4.21E-05	-7.54677	-7.52915
1	ARMA(0, 1) + GARCH(1, 0)	0.004442	3.95E-05	-7.6028	-7.58165
63	ARMA(4 3) + GARCH(1 1)	0.004414	3.92E-05	-7.60228	-7.55998
64	ARMA(4, 4) + GARCH(1, 1)	0.004371	3.89E-05	-7.62915	-7.58333
65	ARMA(4, 5) + GARCH(1, 1)	0.004368	3.89E-05	-7.62622	-7.57687
70	ARMA(5, 4) + GARCH(1, 1)	0.004362	3.87E-05	-7.63324	-7.58389
71	ARMA(5, 5) + GARCH(1, 1)	0.004364	3.87E-05	-7.6327	-7.57983
430	ARMA(5, 4) + GARCH(3, 3)	0.004363	3.87E-05	-7.6332	-7.5768
431	ARMA(5, 5) + GARCH(3, 3)	0.004364	3.87E-05	-7.63321	-7.57329

Atlikta parinkto modelio ARMA(4, 3) + GARCH(1, 1) liekanų bei regresorių analizė siekiant įsitikinti, jog modelis tinka duomenims. Analizuojant visi modelio regresoriai statistiškai skiriasi nuo nulio, remiantis t testo rezultatais, kur nulinė hipotezė, jog koeficientas prie regresoriaus lygus nuliui, o atrankai panaudojant 5%, pasikliautinas intervalas. Modelio tinkamumas duomenims nustatytas analizuojant modelio liekanų savybes (3.6 pav.), tikrinama, ar liekanos yra balto triukšmo procesas. Liekanų vidurkis, lygus -0.00026, ir yra artimas nuliui. Tiek ACF, tiek *Ljung-Box* testai 1-100 poslinkiams laike nurodo autokoreliacijos požymius ir atmetamos nulinės hipotezės visiems tiriamiems poslinkiams, jog liekanos yra priklausomos nuo praeities reikšmių.



3.6 pav. Sukurto modelio ARMA(4, 3) + GARCH(1, 1) liekanų analizės rezultatai

Lyginant šio modelio gautus rezultatus, modeliuojant rinkos dinamikas ARMA modeliu pastebima, jog liekanų savybės yra panašios. Tačiau verta atsižvelgti, jog šiame skyriuje yra tik parenkami modelių parametrai ir tiriamas jų tinkamumas duomenims, modelių prognozių savybės yra aptariamos kitame skyriuje. Šis modelio pranašumas lyginant su ARMA modeliu yra tas, jog jis leidžia taip pat modeliuoti ir prognozuoti proceso sąlyginį heteroskedastiškumą. 3.7 pav. matyti to rezultatai: a) paveikslo dalyje matyti vidutinės ^*GSPC* dienos reikšmės, o b) matyti indekso pokyčių procesas bei (mėlynai) sumodeliuota sklaida. Įvertinti paklaidas prognozuojant sąlyginę dispersiją yra sunku, empiriniuose tyrimuose dažnai yra pasitelkiami didesnio dažnio duomenys (prognozuojant dienos dispersiją, analizuojami valandiniai ar sekundiniai duomenys [34]). Šiame tyrime empirinė kainos nestabilumo prognozė nėra aktuali. Atlikta tik vizuali rezultatų analizė, o sklaidos prognozės bus naudojamos kaip įvesties parametrai LSTM modeliui.



3.7 pav. Sąlyginio heteroskedastiškumo grafikas (mėlyna spalva) bei tiriamas procesas (pilka)

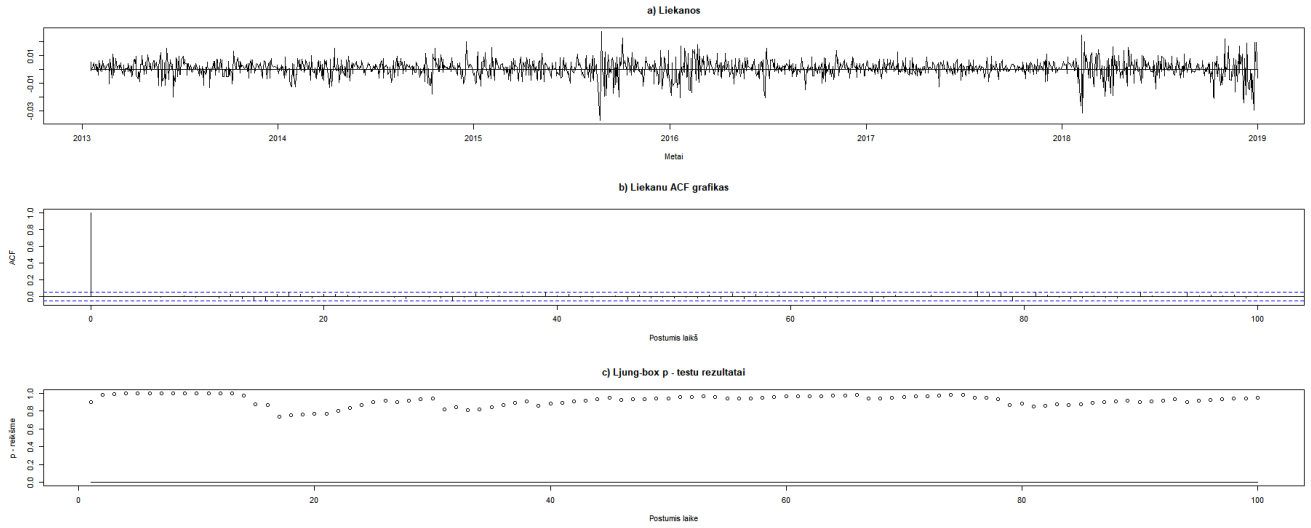
3.1.4 VAR modelis

VAR – vektorinės autoregresijos modelis. Šis modelis naudojamas tiriant konkurencingų kompanijų bei kitų faktorių poveikį vienas kitam. Kadangi šiam modeliui reikia parinkti ne tik autoregresijos laipsnį, bet ir endogeninių kintamųjų kombinaciją, paprastas iteracinis procesas nebėra naudotinas dėl didelių laiko kaštų. Šio modeliui parinkti panaudotas iteracinis procesas, kai į modelį įdedamas regresorius, kuris atneša daugiausiai naudos iš parenkamų regresorių aibės panaudojant AIC statistiką. Šio proceso parinktas geriausias modelis žemiau pateiktoje lentelėje numeris 18. Tačiau atliekant liekanų analizę pastebėta, jog jos nėra balto triukšmo procesas. Todėl šis modelis atmestas kaip netinkamas ir parinktas modelis 207.

3.3 lentelė. VAR modelio parinkimo rezultatai

Nr.	Modelis	Endogeniniai kintamieji	MAE	MSE	AIC	BIC
0	VAR(1)	*iGSPC, iUAL*	0.004464	3.97E-05	-18961.4	-18940.1
18	VAR(1)	*iGSPC, iAAL, iLUV, iDAL, iUAL iDCOILWTICO*	0.004467	3.97E-05	-54957.7	-54766.2
83	VAR(4)	*iGSPC, iAAL, iLUV, iDAL, iUAL, iDCOILWTICO, sezoninė komponentė*	0.004427	3.84E-05	-54235.3	-53193.2
124	VAR(6)	*iGSPC, iAAL, iLUV, iDAL, iUAL, sezoninė komponentė*	0.004439	3.85E-05	-46730.1	-45582
125	VAR(6)	*iDAL, iAAL, iLUV, iDAL, iUAL, iDCOILWTICO, sezoninė komponentė*	0.004433	3.82E-05	-54077.1	-52514.4
207	VAR(10)	*iGSPC, iAAL, iLUV, iDAL, iUAL, iDCOILWTICO*	0.004457	3.80E-05	-54378	-52465.5
208	VAR(10)	*iGSPC, iAAL, iLUV, iDAL, iUAL, Sesionality*	0.004459	3.80E-05	-46468.1	-44555.6
209	VAR(10)	*iDAL, iAAL, iLUV, iGSPC, iUAL iDCOILWTICO, sezoninė komponentė*	0.004452	3.77E-05	-53752.7	-51149.5

Kadangi svarbu gerai sumodeliuoti tik *iGSPC*, atliekama tik šio proceso modeliavimo liekanų analizė. Tačiau prognozuojant ateities reikšmes su ARMA(2, 4) modeliu taip pat bus panaudotos ir šio modelio *iDCOILWTICO* prognozės, nes šis procesas ARMA modelyje yra išorinis regresorius, todėl jo ateities reikšmės turi būti prognozuojamos. Parinkto VAR modelio liekanų analizės rezultatai patiekti 3.8 pav. Kaip ir ankstesnių modelių, šio modelio liekanos yra balto triukšmo procesas. Pastebima, jog VAR(10) su 6 endogeniniais kintamaisiais modelio tiek ACF, tiek ir *Ljong-Box* testų rezultatai rodo daug silpnesnę dabarties reikšmių priklausomybę nuo praeities reikšmių.



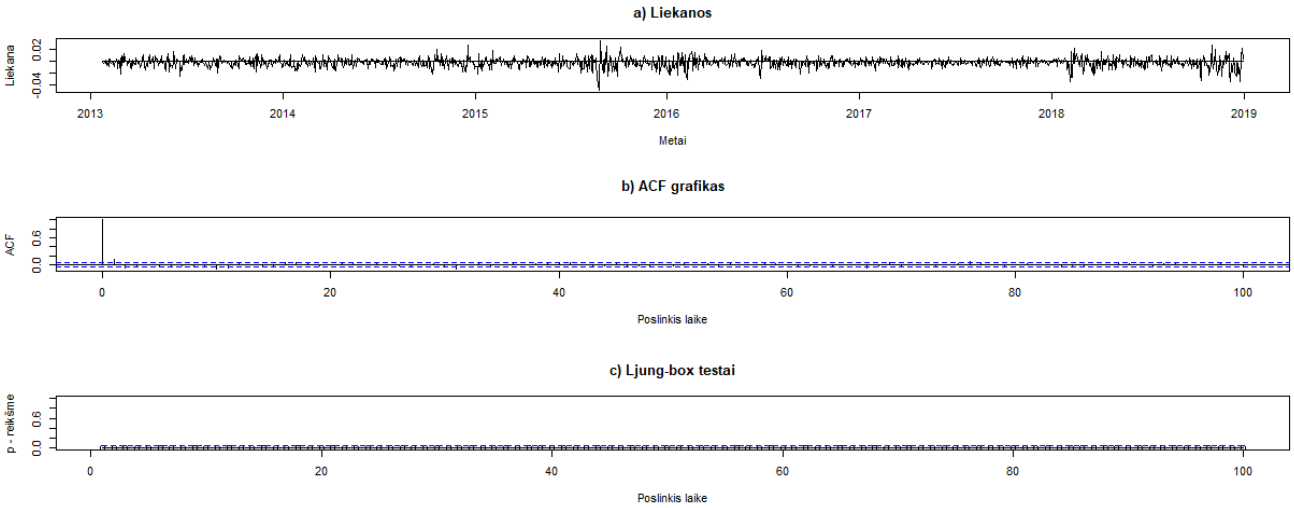
3.8 pav. Sukurto modelio VAR(10) su 6 endogeniniais kintamaisiais liekanų analizės rezultatai

3.1.5 LSTM

Šio tyrimo metu buvo sukurtas ir išanalizuotas LSTM modelio veikimas. Modelis sukurtas Microsoft palaikomos ir plėtojamos programinio paketu CNTK [38] aplinkoje. Kaip ir aukščiau aprašyti modeliai, taip ir šis modelis sukurtas panaudojant visą duomenų imtį ir modeliuojant turimus duomenis, norint nustatyti modelio gebėjimus charakterizuoti duomenis ir juos modeliuoti. Modelį sudaro Modelio įvesties parametrai:

- 8 praeities reikšmės:
    - integruotas ^GSPC indekso vidutinės dienos vertės pokyčio procesas;
    - integruotas ^GSPC indekso aukščiausios dienos vertės pokyčio procesas;
    - integruotas ^GSPC indekso žemiausios dienos vertės pokyčio procesas;
    - integruotas *LUV* akcijos vidutinės dienos kainos pokyčio procesas;
    - integruotas *DEL* akcijos vidutinės dienos kainos pokyčio procesas;
    - integruotas *AAL* akcijos vidutinės dienos kainos pokyčio procesas;
    - integruotas *UAL* akcijos vidutinės dienos kainos pokyčio procesas;
    - integruotas *DCOILWTICO* vidutinės dienos kainos pokyčio procesas;
    - sezoninė komponentė.

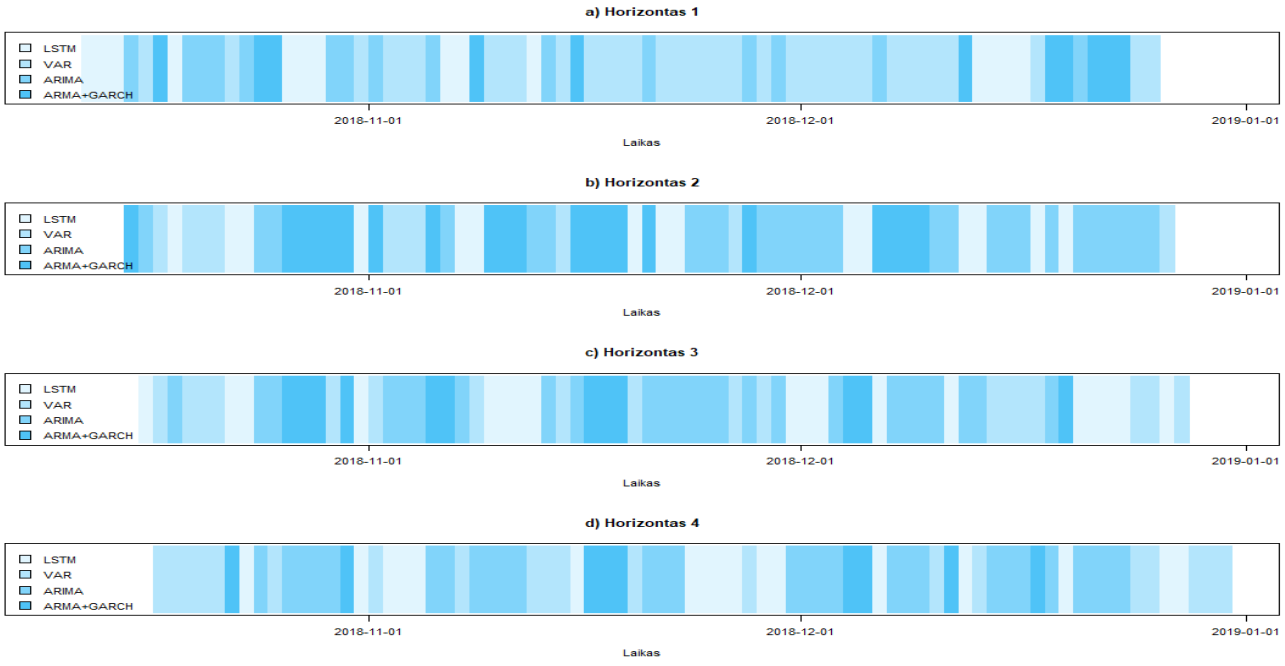
3.9 pav. matyti, jog modelio liekanos nėra balto triukšmo procesas – tai reiškia, jog liekanose vis dar yra struktūrų, kurias gali paaiškinti kiti statistiniai modeliai. Taip pat šio modelio tiek MAE, tiek MSE 0.0059 ir 6.27E-05 atitinkamai yra didžiausi, lyginant su kitais modeliais. Tai rodo, jog modelis prasčiausiai paaiškina proceso pokyčius laike. Norint pagerinti modelio kokybę galima šio modelio išvesties prognozes modeliuoti su statistiniais modeliais, pavyzdžiui, ARMA.



3.9 pav. Sukurto LSTM modelio liekanų analizės rezultatai

3.2 Modelių vienos dienos į priekį prognozės charakteristikos

Modelių išsami prognozių lyginimo analizė yra pateikta kitame skyriuje. Šiame skyriuje yra analizuojamos sukurtų modelių prognozavimo dinamikos. Dinamikoms analizuoti anksčiau aprašyti 4 modeliai (LSTM, ARIMA, ARMA+GARCH ir VAR) yra apmokomi naudojant 97% tiriamos duomenų imties ir atliekama 1, 2, 3 ir 4 dienos akcijos kainos ateities prognozė, tuomet apmokymo imtis padidinama 1 stebiniu, ir procesas kartojamas, kol gaunama 50 prognozių. 3.10 pav. pateiktas modelių prognozių palyginimas. Periodas yra nuspalvinamas tam tikra spalva, jeigu šiame intervale modelis padarė mažiausią spėjimo paklaidą. Pastebima, nors ir LSTM modelis padarė didžiausią paklaidą modeliuojant domenas, tačiau jo prognozės skirtingais periodais yra tiksliausios. Be to, matyti tendencija, jog tikslinga modelio prognozė dažnai kartojasi.



3.10 pav. Modelių prognozavimo charakteristikų palyginimas

3.3 Akcijų kainų prognozavimo tyrimas

Šiame poskyryje aprašyti modelių greitaveikos bei proceso ^*GSPC* prognozavimo tyrimų rezultatai. Pirmoje skyriaus dalyje pateikiami modeliams apmokyti reikalingi laiko kaštai. Tuomet aprašomos tiriamo proceso skirtingų modelių prognozavimo charakteristikos bei testų atlikimo aplinkybės. Paskutiniame poskyryje pateikiami rinkos simuliacijos rezultatai bei skirtingų strategijų investuojant į rinką palyginimai.

3.3.1 Greitaveikos tyrimas

Greitaveikos tyrimai yra atliekami kuriant modelius su įvairiomis (500 iki 1500) duomenų imties pjūviais ir apskaičiuojant vidutines laiko sąnaudas. Visi testai atlikti toje pačioje aplinkoje prie tų pačių sąlygų (naudojamų paketų versijos nebuvo keičiamos tyrimo metu). Aplinkos parametrai:

- Operacinė sistema: Windows 10 x64
- .NET versija: 4.7.2
- R versija: 3.6.0
- RAM: 16 GB RAM
- Procesorius: Intel i7-7700HQ

3.4 lent. pateikti tyrimų rezultatai. Kadangi modelių asamblėjoms sukurti reikia kitų modelių prognozių, todėl į duomenų paruošimo kaštus įeina ir prognozių generavimo kaštai. Pavyzdžiui, norint sukurti 1 dienos į priekį prognozę pasinaudojant asamblėją (AM), reikia pirmiausia perskelti apmokymui skirtų duomenų imtį į dvi dalis, sukurti prognozes, kurių kiekis lygus 50% apsimokymo imties su visais keturiais modeliais ($\frac{imties\ dydis \cdot 1026}{2}$), ir atitinkamai apmokyti atsitiktinio miško modelį su naujai sukurta apsimokymui skirtą duomenų imtį su naujomis prognozėmis, kas papildomai dar užtruks vidutiniškai 30 milisekundžių. Šie testai buvo atlikti, nes norėta įsitikinti, jog prognozėmis galima naudotis tyrime aprašytame šio skyriaus paskutiniame poskyryje, kuomet yra simuliuojama rinka ir prie kai kurių sąlygų prognozės kuriamos prekybos dienos pabaigoje. Kadangi duomenų paruošimas gali būti atliekamas iš anksto, reikia tik atitinkamai sugeneruoti vienos dienos į priekį prognozę, tai truks (žiūrėti į žemiau pateiktą lentelę) *Laiko sąnaudos prognozei generuoti + Laiko sąnaudos duomenims paruošti* šiuo atveju n yra lygu vienam, nes su n-1 imtimi galima atlikti reikalingus procesus prieš dienos pabaigą. Ilgiausiai užtruks modelio asamblėjos (LSTM) prognozės generavimas, kuris truks mažiau nei 2 sekundes. Šis rodiklis yra pakankamai mažas, jog tenkintų simuliacijos kriterijus.

3.4 lentelė. Greitaveikos tyrimo rezultatai

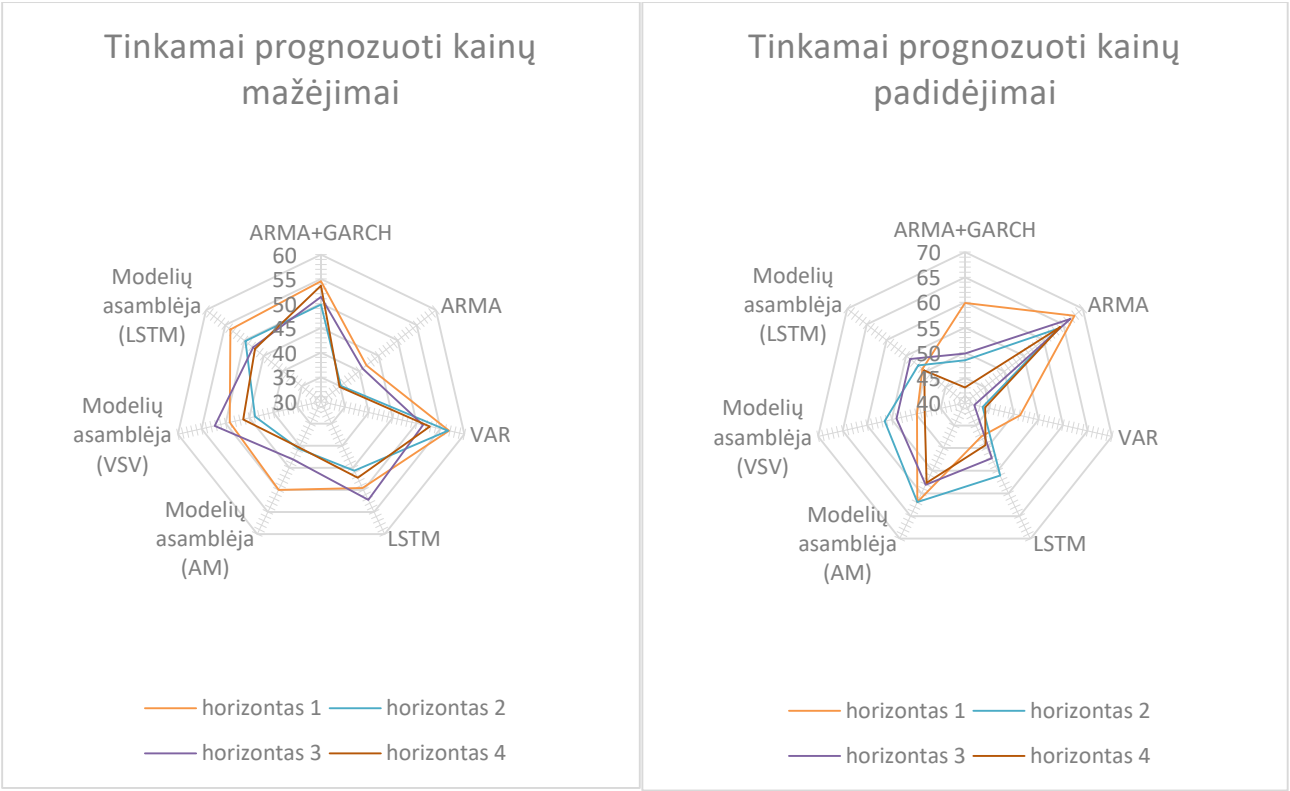
Modelio pavadinimas	**Laiko sąnaudos duomenims paruošti (n – 50 % apsimokymo imties)**	**Laiko sąnaudos prognozei generuoti (milisekundės)**
VAR	<10	80
ARMA	<10	118
ARMA+GARCH	<10	399
LSTM	<10	429
Modelių asamblėja (VSV)	1026 · n	<1
Modelių asamblėja (AM)	1026 · n	30
Modelių asamblėja (LSTM)	597 · n	1183

3.3.2 ^*GSPC* indekso prognozių analizė

Tiriamojo darbo metu buvo sukurti, ištirti ir palyginti 7 modeliai: LSTM, ARIMA, VAR, ARMA+GARCH, bei 3 modelių asamblėjos. Kadangi asamblėjų architektūros reikalauja, jog apsimokymo imtį papildomai būtų galima skaidyti į dvi lygias dalis, kur pirmoji dalis yra naudojama pirminiams modeliams apmokyti ir prognozėms generuoti, gautos prognozės pridedamos prie antrosios apsimokymui skirtų duomenų dalies ir jas panaudojant atliekamos atitinkamos procedūros, aprašytos 2.2 skyriuje. Todėl tiriama duomenų imtis, kurią sudaro 1509 stebiniai, dalijama į dvi dalis: pirmi 1000 įrašų (~66 %) panaudojami modelių sukūrimui, o likusi imtis panaudojama testavimui. Visi modeliai testuojami kryžminio patikrinimo metodu, kai apsimokymo imtis yra naudojama atliekant ateities prognozę, po to apmokymo imtis padidinama 1 stebiniu, kuris paimamas iš testavimui skirtų duomenų sąrašo galo (duomenys surikiuoti datos didėjimo tvarka), procesas kartojamas, kol testavimo imtyje nebėra duomenų. Remiantis 2.3 skyriaus įžvalgomis modelių sukurtos prognozės yra lyginamos panaudojant 4 kriterijus:

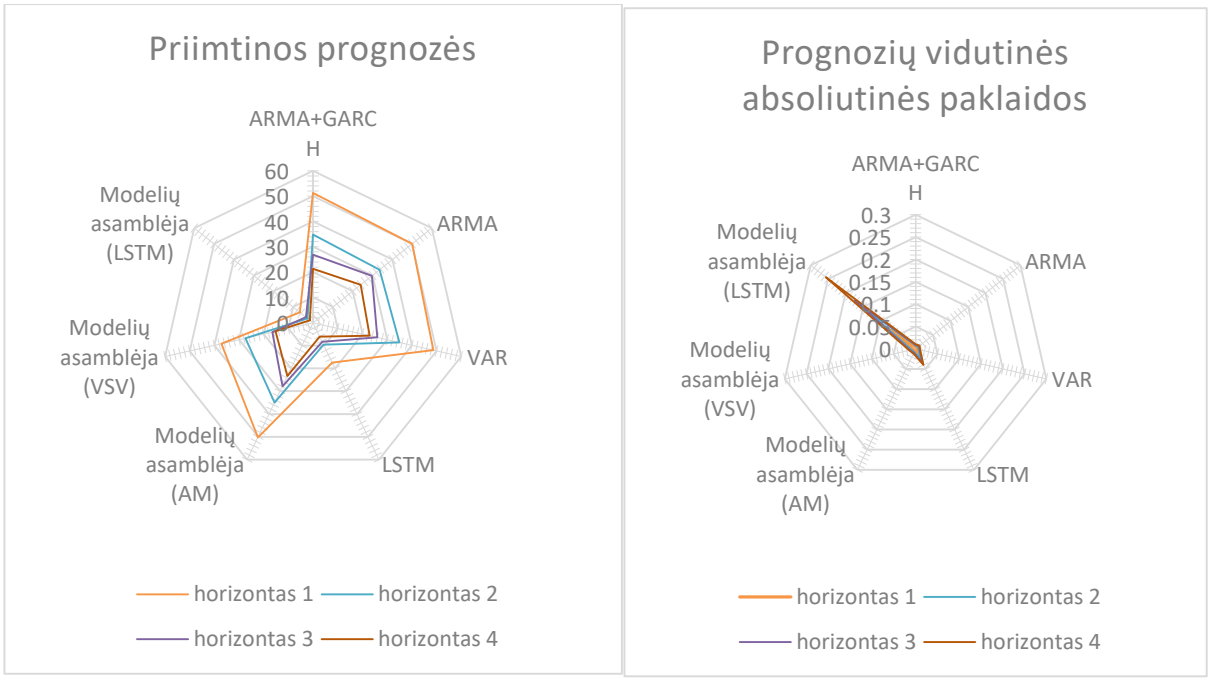
- teigiamo ženklo teisingas atpažinimas
- neigiamo ženklo teisingas atpažinimas
- absoliutinė vidutinė prognozių paklaida
- prognozių tinkamumas

Ženklas traktuojamas kaip indekso pokyčio kryptis, arba integruoto proceso reikšmės ženklas. Skaidomo pokyčio kryptis teigimas ir neigimas, nes skirtingoms investavimo strategijoms gali būti naudingos skirtingos metrikos. 3.11 grafike pateikti ženklų nustatymo rezultatai. Bandymų rezultatai išreikšti procentais. Ištirti 4 dienų į priekį prognozavimo horizontai. Pastebima, jog geriausiai kainų mažėjimus nuspėja su 57 % tikslumu VAR modelis, nuo jo tik 2 % atsilieka ARMA+GARCH modelis ir Modelių asamblėja (LSTM). Prognozuojant teigiamus pokyčius akivaizdžiai išsiskiria ARMA modelis, kurio teikiamų kainų pokyčių prognozavimo tikslumas beveik siekia 68 %, tačiau šio modelio mažas 42 % tikslumas prognozuojant akcijos kainos neigiamus ženklus. Antroje vietoje modelių asamblėja (AM), kurio tikslumas yra ~ 62% prognozuojant ateities kainų didėjimus.



3.11 pav. Tiriamų modelių ženklų prognozavimo palyginimas

Taip pat yra lyginamos modelių absoliutinės paklaidos prognozuojant nuo 1 iki 4 dienų indekso vidutines dienos reikšmes bei modelių prognozių tinkamumas. Tinkamumas yra vertinamas teigiamai, jei ateities kainos prognozė yra didesnė arba lygi prognozuojamos dienos rinkos indekso, įgaunama minimali reikšmė ir mažesnė arba lygi maksimaliai. Kadangi visų sukurtų modelių paklaidos yra labai panašios, tinkamumo metrika padeda geriau įvertinti absoliutinės vidutinės paklaidos svarbą. 3.12 pav. matyti šių metrikų palyginimo rezultatai. Akivaizdžiai išsiskiria du modeliai: LSTM ir asamblėja (LSTM). Jų prognozavimo paklaidos yra didžiausios ir atitinkamai jų tinkamumai mažiausi, tačiau šių modelių savybės yra tinkamai panaudojamos akcijų pirkimuose (toliau skaityti kitame skyriuje). Daugiausiai tinkamų prognozių sugeneravo ARMA+GARCH modelis, to buvo galima tikėtis, nes šis modelis parodė gerų rezultatų prognozuojant kainos pokyčio kryptį bei padarė mažiausią vidutinę absoliutinę paklaidą. Šis modelis padarė vidutinę 0.004529 *iGSPC* proceso reikšmės. Tai reiškia, jog vidutiniškai modelis blogai suprognozuoja 0.45 % rytojaus indekso reikšmės, arba 99.5 % tikslumu prognozuoja rytojaus indekso vertes. Tačiau net ir šio modelio tinkamų prognozių kiekis siekia tik 51.27 % testavimo imtyje, visi kiti modeliai sugebėjo prognozuoti ateities kainas tinkamai tik 50% ir mažiau atvejų. To priežastis prognozavimo laikotarpyje proceso sklaida yra 0.00453, todėl net naivus modelis (kur ateities prognozė yra dabarties kaina) gali pasiekti proceso prognozavimo tikslumą, kuris yra didelis (žiūrėti 1.8 skyrių), tačiau tai nėra pakankama metrika vertinant modelio kokybę. Taip pat pastebima, jog 2-jų ar daugiau dienų prognozės į priekį yra netinkamos, nes tinkamų prognozių kiekis mažėja sparčiai didinant prognozavimo horizontą. Padidinus prognozavimo horizontą nuo 1 iki 2 dienų pastebėtas 12 % vidutinis nuosmukis tinkamai prognozuoti indekso vertes ir atitinkamai 6 % be 3% nuostoliai, lyginant su ankstesniu horizontu, didinant horizontą iki 4 dienų. Šiame horizonte (4 dienų) geriausiai akcijų kainas prognozavo ARMA modelis, bet tik 24 % šio modelio prognozių buvo prilyginamos tinkamoms.



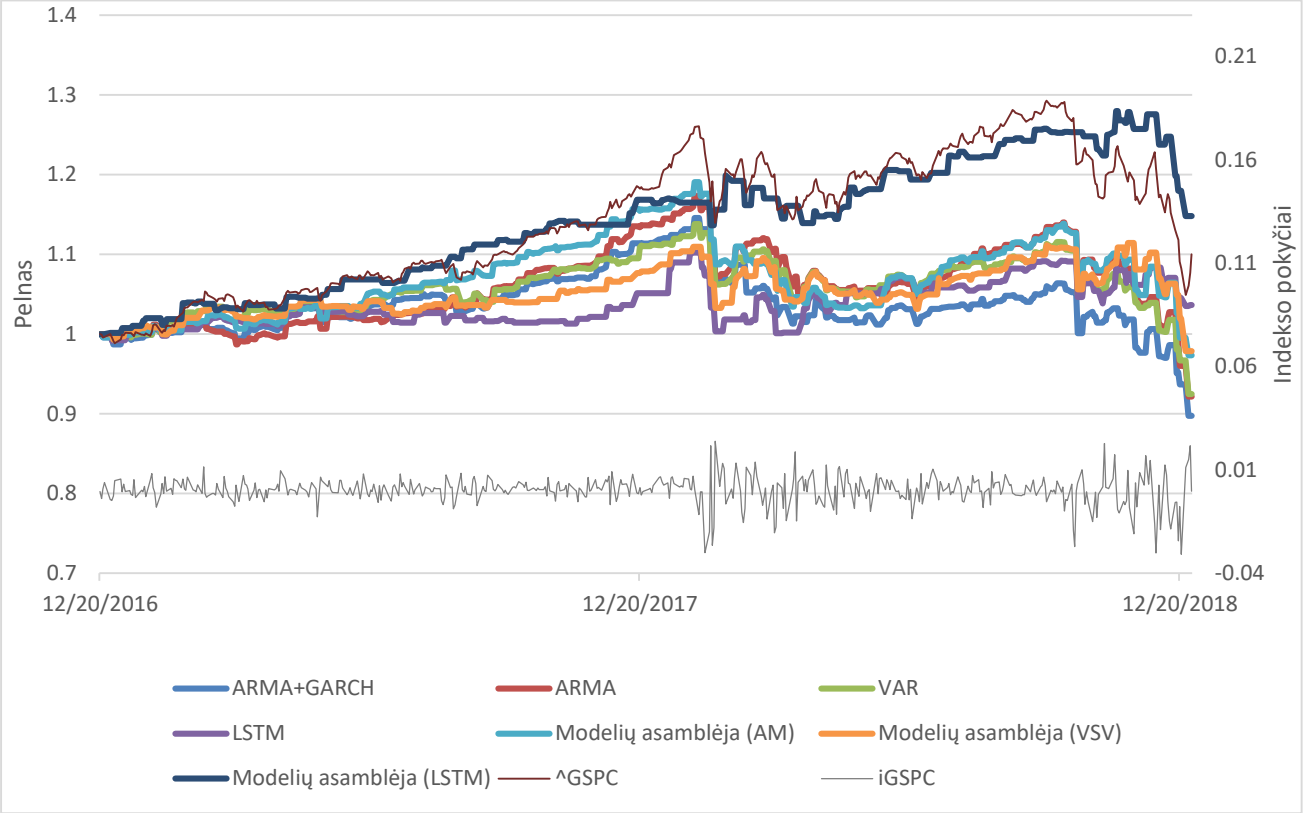
3.12 pav. Tiriamų modelių prognozavimo tinkamumo palyginimas

3.3.3 Rinkos simuliacijos tyrimo rezultatai

Norint dar objektyviau įvertinti kuriamų modelių prognozių atnešama naudą, šio darbo metu yra sukurta rinkos simuliacija. Atliekamas tyrimas, kai simuliuojama rinka ir atliekamos investicijos į ją panaudojanti strategiją *PirkIrParduok*, detaliai aprašytą 2.3.1 skyriuje. Kadangi prognozuojamas ^*GSPC* yra rinkos indeksas sudarytas iš pačių akcijų, jį patį galima traktuoti kaip akciją, o jo reikšmę *i*-tuoju laiko momentu kaip akcijos kainą. Tiesa, jei norima pasiekti pelną, kuris yra dokumentuotas tolimesniuose rezultatuose, reikia atlikti pirkimus ir pardavimus visų į indeksą įtrauktų įmonių akcijų, arba norint gauti panašių rezultatų reikia pirkti ETFs akciją, kuri siekia savo akcijų kainas „pririšti“ prie tiriamo indekso reikšmių (daugiau informacijos skaityti 1.1), tokiu atveju patartina ^*GSPC* tyrimo metu naudotą procesą pakeisti atitinkamu ETFs akcijų kainų procesu tam, kad būtų gaunamos kuo tikslesnės prognozės.

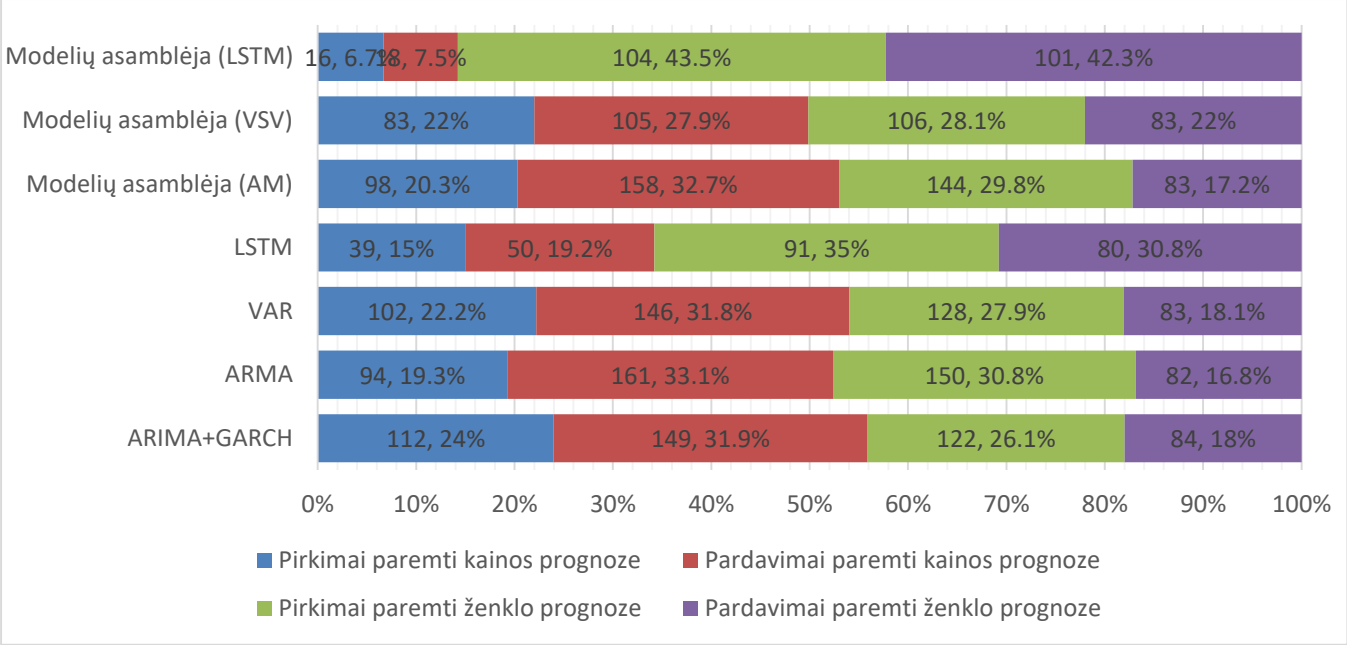
3.13 pav. pateiktame grafike matyti rinkos simuliacijos rezultatai. Grafike pavaizduoti rezultatai panaudojant strategijos *PirkIrParduok* su skirtingų modelių prognozių įverčius. Grafike iliustruota, kaip pelningumo rodmenys kinta laike simuliuojamame prekybos intervale. Pastebima, jog modelis asamblėja (LSTM), kurio prognozių paklaidų charakteristikos buvo prasčiausios, o ženklo nustatymo charakteristikos vidutinės, šiuo atveju parodė geriausius rezultatus. Per 2 metus šis modelis sugebėjo sugeneruoti 14.82 % prieaugį nuo investuotos sumos, kai tuo tarpu didžiausias šio modelio prieaugis tiriamame intervale yra lygus 27.97 % (2018 m. 7 mėn. 11 d.). Grafike pilka paploninta linija yra pavaizduotas *iGSPC* procesas (indekso pokytis išreikštas %). Matyti, jog dideli kainų sumažėjimai yra sekami didelių indekso verčių (atitinkamai akcijų kainų), padidėjimų. Šiuose rinkos nestabilumo perioduose tiriama *PirkIrParduok* strategija atitinkamai patiria didžiausią nuostolį ir / arba pelną priklausomai kaip gerai šiame periode atpažįsta pokyčio kryptis ir pokyčio dydį. Be to, šiame grafike taip pat pateikiami rezultatai strategijos *PirkIrLaikyk* (angl. *buy-and-hold*) paploninta tamsiai raudona linija. Šios strategijos esmė: akcijų įsigijimas ir jų laikymas ilgą laikotarpį (legendoje pavadinta ^GSPC), kur pelnas (grąža) yra skaičiuojamas pagal (6) formulę, jei akcijos būtų įsigytos simuliacijos pradžioje ir parduodamos atitinkamą dieną už vidutines tos dienos kainas. Pastebima, jog tik viena iš

visų pasiūlytų modelių prognozių buvo pakankamai naudinga *PirkIrParduok* strategijai, kuri pralenkė pelningumu *PirkIrLaikyk* strategiją.



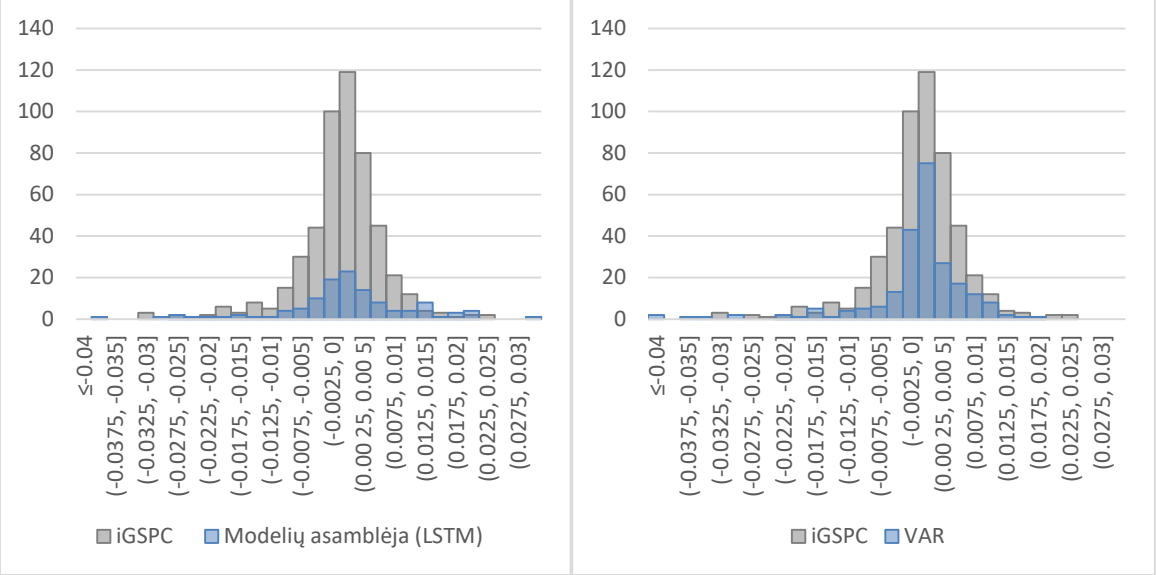
3.13 pav. Rinkos simuliacijos tyrimo rezultatai

Atlikus išsamesnę simuliacijos analizę pastebima to priežastis. 3.14 pav. pateiktuose rezultatuose matyti, prie kokių sąlygų buvo atlikti akcijų pirkimai ir pardavimai. Matyti, jog strategija *PirkIrParduok* panaudodama rinkos indekso prognozes, sugeneruotas modelių asamblėja (LSTM), atliko mažiausiai pirkimų ir pardavimų remiantis kainos (indekso) reikšmės prognozėmis, o daugiausiai rėmėsi modelio prognozuotomis ženklo prognozėmis. O tuo tarpu strategija, panaudojanti ARMA+GARCH prognozių rezultatus, beveik 56 % visų sprendimų atliko panaudodama indekso pokyčio prognozėmis ir tik likusius sprendimus atliko remdamasi ženklo prognoze. Atsižvelgiant į rezultatus, pateiktus 3.13 pav. ir 3.14 pav., galima daryti prielaidą, jog modeliai, kurie tinka duomenims (daugiau informacijos 3.1 skyriuje) ir minimizuoja šių modelių paklaidų vidutinę kvadratinę paklaidą, geba gerai ir tinkamai (metrika aprašyta 3.3.2 skyriuje), prognozuoti Indekso reikšmes, kurias galima panaudoti investavimui į rinką, kai rinka yra stabili. Tačiau laikotarpiuose, kai rinka yra nestabili (matyti padidėjusi sklaida), negeba kokybiškai įvertinti jos ateities reikšmių pokyčių bei pokyčių krypčių. Nestabilumo laikotarpių geriausius rezultatus parodė pasiūlytas modelių asamblėjos (LSTM) modelis, kuris nesugeba tinkamai prognozuoti pokyčio dydį, lyginant su statistiniais modeliais, bet geba nustatyti pokyčio tendencijas.



3.14 pav. Tiriamų modelių charakteristikos

Žemiau esančiuose grafikuose pateiktos histogramos, kur grupavimo intervalas 0.0025 %, detalizuojantis simuliacijos rezultatus. Pilka spalva pavaizduoti S&P 500 indekso pokyčiai, išreikšti procentais (*iGSPC*) simuliacijos intervale, ir atitinkamai mėlyna spalva nuspalvinti pelno pokyčiai, apskaičiuoti rinkos simuliacijos metu panaudojant *PirkIrParduok* strategiją su atitinkamo modelio prognozėmis. Šioje diagramoje akivaizdžiai matyti, kaip skirtingos modelių prognozių metrikos atspindi simuliacijos rezultatuose. Modelio ARMA+GARCH, kuris padarė mažiausią MAE iš tiriamų modelių prognozuojant vienos dienos į priekį indekso reikšmes ir pasižymi geromis ženklų atspėjimo charakteristikų pelningumo pykčių histograma, yra panaši į *iGSPC* proceso, matyti, jog yra atliekama dvigubai (51%) daugiau sprendimų, panaudojant šio modelio prognozes, lyginant su Modelių asamblėja (LSTM). To priežastis: modelis geba tiksliau prognozuoti indekso pokyčio dydį, todėl net maži pokyčiai prognozuojamame indekse gali nulemti akcijų pirkimą ar pardavimą, pardavimai atliekami beveik kas antrą dieną. Modelių asamblėjos (LSTM) geriau atpažįsta rinkos tendenciją, t.y. mažiau reaguoja į mažus indekso pokyčius, bet teisingai identifikuoja didesnius, grafike matyti, jog VAR modelis kelis kartus iš eilės netinkamai identifikavo pokyčio kryptį tiriamajame intervale, kai rinka buvo nestabili, ir 2 kartus patyrė 4% nuostolį. Iš viso strategija *PirkIrParduok* panaudodama atitinkamų Modelių asamblėjos (LSTM) ir ARMA+GARCH prognozes atliko atitinkamai 11 ir 19 investicijų, kurios padarė didesnį arba lygu 1% nuostoliui, ir atitinkamai 21 ir 11, kurie padarė didesnį nei 1 % pelną. Šios atitinkamų skirstinių galuose esančios reikšmės yra svarbios, nes didžioji abiejų modelių sprendimų dalis sugeneruoja labai mažą prieaugį: nuo 0 iki 0,25 % prieaugio (kitų sukurtų modelių histogramos pateiktos 1 priede).



3.15 pav. iGSPC bei pelno pokyčių histogramos

Išvados

1. Darbo metu identifikuotos tiriamo proceso S&P 500 (*GSPC*) savybės leido parinkti DNT bei statistinius modelius, gebančius modeliuoti bei prognozuoti šį procesą. Tiriant integruotą *GSPC,* panaudojant ACF bei PAC identifikuoti praeities reikšmių poveikiai dabarties reikšmėms bei periodiškumo požymiai, šią proceso savybę nuspręsta modeliuoti SARMA modeliu. Analizuojant integruoto tiriamo proceso reikšmių grafiką pastebėta sąlyginė sklaida, kurią pasirinkta modeliuoti ARMA + GARCH modeliu. Kadangi akcijų procesai reaguoja į impulsus (naujos informacijos atsiradimą) skirtingai, pasirinktas VAR modelis ir procesai, kurie galį turėti poveikį tiriamajam procesui. Taip pat pasirinktas LSTM modelis procesui modeliuoti, todėl šis modelis geba prisiminti įvykius bei nustatyti jų priklausomybes.
2. Tyrimo metu nustatyta, jog rinkos modeliai skirtingais laiko intervalais yra vienas už kitą pranašesni (padaro mažiausią absoliutinę paklaidą), kitaip pasakius, jų tikslumas priklauso nuo rinkos savybių, kurios nėra pastovios tiriamame laiko intervale. Remiantis šiuo pastebėjimu pasiūlytos 3 modelių asamblėjų architektūros, siekiant panaudoti šią savybę. Visos sukurtos architektūros naudoja 1 punkte nurodytų sukurtų modelių tiriamo proceso prognozes, kuriant naujas prognozes. Pirmoji pasiūlyta modelių prognozių gerinimo strategija – prognozių vidurkiniams panaudojant vienodų svorių vidurkinimą (sukurtas modelis pasinaudojant šia strategija pavadintas modelių asamblėja (VSV)), antroji pasiūlyta strategija panaudojant klasifikatorių prognozuoti prieš atliekant kiekvieną tiriamo proceso ateities prognozę, kuris iš tyrimų modelių padarys mažiausią paklaidą ir naudoti šio modelio prognozes. Kadangi modelių kiekis baigtinis, o procesą aprašančių charakteristikų kiekis didelis, parinktas klasifikatorius atsitiktinis miškas (modelis pavadintas modelių asamblėja (AM)). Trečioji pasiūlyta strategija prie įvesties domenų LSTM modeliui pridėti statistinių modelių tiriamo proceso prognozes, bei atlikti šio modelio parametrų optimizaciją gerinant ne MSE, o proceso pokyčio krypties nustatymo metriką (modelis pavadintas modelių asamblėja (LSTM)).
3. Atlikus sukurtų modelių greitaveikos testus nustatyta, jog ilgiausiai 4 dienų į priekį prognozėms sukurti užtrunka modelių ansamblis (LSTM) – beveik 2 sekundes. Šis rezultatas yra pakankamai mažas tam, kad būtų galima pasinaudojant šią dieną surinktais duomenimis sukurti rytojaus prognozes ir atlikti finansinius sprendimus, likus 1 minutei iki rinkos darbo dienos pabaigos.
4. Analizuojat ištirtų modelių prognozių rezultatus nustatyta, jog geriausiai S&P 500 indekso mažėjamus (lyginant su vakarykšte diena) nuspėja su 57 % tikslumu VAR modelis nuo jo tik 2 % atsilieka ARMA+GARCH modelis ir Modelių asamblėja (LSTM). Prognozuojant teigiamus pokyčius akivaizdžiai išsiskiria ARMA modelis, kurio teigiamų kainų pokyčių prognozavimo tikslumas beveik siekia 68 %, tačiau šio modelio mažas 42 % tikslumas prognozuojant indekso pokyčio proceso neigiamus pokyčius. Antroje vietoje Modelių asamblėjos (AM) modelis, kurio tikslumas yra ~ 62% prognozuojant indekso didėjimą. Čia teigiamai vertinama pokyčio prognozė, jei pokyčio kryptis sutampa su prognozės įverčio ženklu.
5. Nustatyta, jog tiriamam procesui prognozuoti tinkamiausias modelis ARMA+GARCH. Tinkamumas yra vertinamas teigiamai, jei ateities kainos prognozė yra didesnė arba lygi prognozuojamos dienos rinkos indekso įgaunamai minimaliai reikšmei ir mažesnė arba lygi maksimaliai, o neigiamai priešingu atveju. Remiantis šia metrika galima išskirti du modelius: LSTM ir modelių asamblėją (LSTM). Jų prognozavimo paklaidos yra didžiausios ir atitinkamai

jų tinkamumai mažiausi. Daugiausiai tinkamų prognozių sugeneravo ARMA+GARCH modelis, to buvo galima tikėtis, nes šis modelis parodė gerus rezultatus prognozuojant kainos pokyčio kryptį bei padarė mažiausią vidutinę absoliutinę paklaidą. Šis modelis padarė vidutinę absoliutinę 0.004529 paklaidą prognozuojant *iGSPC* proceso rekšmes. Tačiau net ir šio modelio tinkamų prognozių kiekis siekia tik 52.3 % testavimo imtyje, visi kiti modeliai sugebėjo prognozuoti ateities kainas tinkamai tik 50 % ir mažiau atvejų. Taip pat pastebima, jog 2-jų ar daugiau dienų prognozės į priekį yra netinkamos, nes tinkamų prognozių kiekis mažėja sparčiai didinant prognozavimo horizontą. Padidinus prognozavimo horizontą nuo 1 iki 2 dienų pastebėtas 12 % vidutinis nuosmukis tinkamai prognozuoti indekso vertes ir atitinkamai 6 % be 3% nuostoliai, lyginant su ankstesniu horizontu, didinant horizontą iki 4 dienų. Šiame horizonte (4 dienų) geriausiai akcijų kainas prognozavo ARMA modelis, bet tik 24 % šio modelio prognozių buvo prilyginamos tinkamoms.

6. Atliekant rinkos simuliacijos bandymus strategija „Pirk ir parduok“ buvo pelninga tik panaudojant modelių asamblėjos (LSTM) ir LSTM ateities prognozes, visais kitais atvejais buvo nuostolinga. Didžiausias pelnas pasiektas panaudojant modelių asamblėją (LSTM), per 2 metus sugeneruoti 14.82 % prieaugis nuo investuotos sumos, kai tuo tarpu didžiausias šio modelio prieaugis tiriamame intervale yra lygus 27.97 % (2018 m. 7 mėn. 11 d.). Analizuojant simuliacijos rezultatus pasitelkiamos histogramos, kur grupavimo intervalas 0.0025 %, pastebėta, jog šis modelis geriau atpažįsta rinkos tendenciją, t.y. mažiau reaguoja į mažus indekso pokyčius, bet teisingai identifikuoja didesnius lyginant su kitais ištirtais modeliais.
7. Tyrimo metu nustatyta, jog metrikos, tokios kaip MSE, MAE, bei teisingo pokyčio ženklo nustatymas nėra pakankamas norint teisingai įvertinti modelio sukurtų prognozių tinkamumui atliekant investavimą į akcijų rinką. Atlikus rinkos simuliacijos tyrimą pastebima, jog „Pirk ir parduok“ strategija panaudodama modelių prognozes, kurios yra gana tikslios (ARMA+GARCH vidutinė absoliutinė paklaida 0.45 % rytojaus indekso reikšmės), nesugeba sugeneruoti pelno tiriamame laikotarpyje. Nors ir modelis geba „įsisavinti“ rinkos prieaugį, kai rinka reliatyviai stabili, nestabilumo periode visas gautas pelnas yra prarandamas.

Literatūra

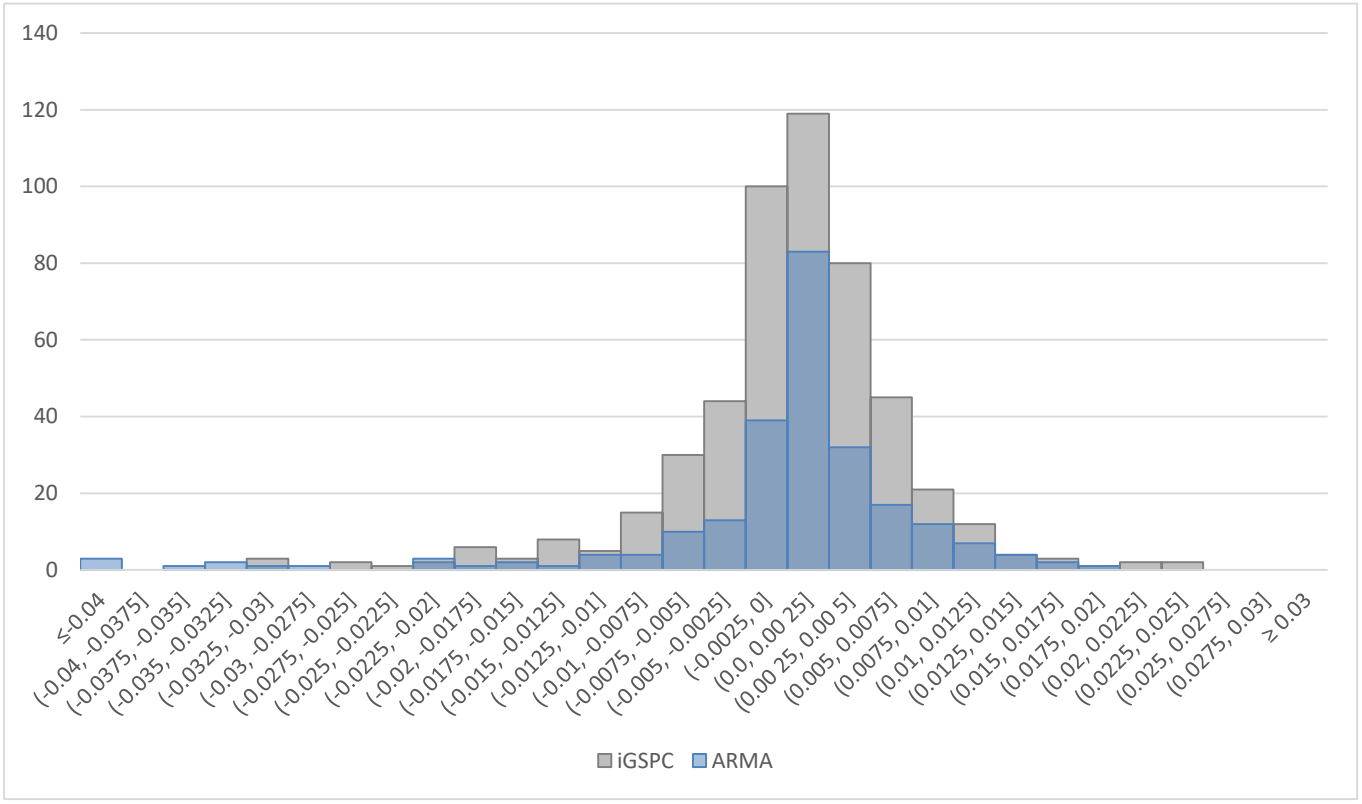
1. THE WORLD BANK. *Market capitalization of listed domestic companies* [interaktyvus]. [žiūrėta 2019-04-27]. Prieiga per internetą: https://data.worldbank.org/indicator/cm.mkt.lcap.cd
2. YIN-WONG CHEUNG, MENZIE D. CHINN, ANTONIO GARCIA PASCUAL, YI ZHANG. *Working Paper Series*. ISBN 978-92-899-2740-6 [interaktyvus]. [žiūrėta 2018-02-12]. Prieiga per internetą: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp2018.en.pdf?4a7f5e47892951a024a3cedd13f9a459
3. YAHOO FINANCE. *Business Finance, Stock Market, Quotes, News*. [interaktyvus]. [žiūrėta 2018-03-13] https://finance.yahoo.com/
4. FRED. *Federal Reserve Economic Data*. [interaktyvus]. [žiūrėta 2018-03-13]. https://fred.stlouisfed.org/
5. D. VENUGOPAL SETTY, T. M. RANGASWAMY, K. N. SUBRAMANYA. *A Review on Data Mining Applications to the Performance of Stock Marketing, International Journal of Computer Applications*. 2010 , vol. 1, no. 3, pp. 24-34. DOI: 10.5120/88-187 [interaktyvus]. [žiūrėta 2018-02-12]. Prieiga per internetą: https://www.researchgate.net/publication/43655911_A_Review_on_Data_Mining_Applications_to_the_Performance_of_Stock_Marketing
6. WORLD-EXCHANGES. [interaktyvus] [žiūrėta 2018-01-06] Prieiga per internetą https://www.world-exchanges.org/home/index.php/statistics/monthly-reports
7. SRICHANDER RAMASWAMY. *Market structures and systemic risks of exchange-traded funds*. 2011, *BIS Working Papers*, no 343. ISBN 1682-7678 [interaktyvus]. [žiūrėta 2019-02-21]. Prieiga per internetą: https://www.bis.org/publ/work343.pdf
8. LISA KEALY, KIERAN DALY, ANDREW MELVILLE, PIERRE KEMPENEER, MATT FORSTENHAUSLER, MARK D. MICHEL, JULIE KERR: *Global ETF Research*. EYG, no. 06335-174GBL 1709-2413498 [interaktyvus]. [žiūrėta 2019-03-13]. Prieiga per internetą: https://www.ey.com/Publication/vwLUAssets/ey-global-etf-survey-2017/$FILE/ey-global-etf-survey-2017.pdf
9. SNEHA SONI. *Applications of ANNs in Stock Market Prediction: A Survey. International Journal of Computer Science & Engineering Technology*. Vol. 2, no. 3. ISSN: 2229-3345 [interaktyvus]. [žiūrėta 2017-9-20]. Prieiga per internetą: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.478.4538&rep=rep1&type=pdf
10. NASDAQ. *NASDAQ Composite*. [interaktyvus]. [žiūrėta 2017-9-20]. Prieiga per internetą:https://indexes.nasdaqomx.com/docs/FS_COMP.pdf
11. ANNE SRADERS. *What Is the Dow Jones Industrial Average*. [interaktyvus]. [žiūrėta 2017-9-20] Prieiga per internetą: https://www.thestreet.com/investing/what-is-dow-jones-industrial-average-14769048
12. WILL KENTON, CHRIS B MURPHY. *S&P 500 Index – Standard & Poor's 500 Index Definition.* [interaktyvus]. [žiūrėta 2017-9-20]. Prieiga per internetą: https://www.investopedia.com/terms/s/sp500.asp
13. JONATHAN CLARKE, TOMAS JANDIK, GERSHON MANDELKER. *The Efficient Markets Hypothesis*. 2000 [interaktyvus]. [žiūrėta 2018-01-06]. Prieiga per internetą: http://www.e-m-h.org/ClJM.pdf
14. FAMA, F. EUGENE. *Random Walks In Stock Market Prices*. *FINANCIAL ANALYSTS JOURNAL*. 1965, vol. 21, no. 5. Doi:10.2469/faj.v21.n5.55 [interaktyvus]. [žiūrėta 2018-01-06]. Prieiga per internetą: https://www.chicagobooth.edu/~/media/34F68FFD9CC04EF1A76901F6C61C0A76.PDF

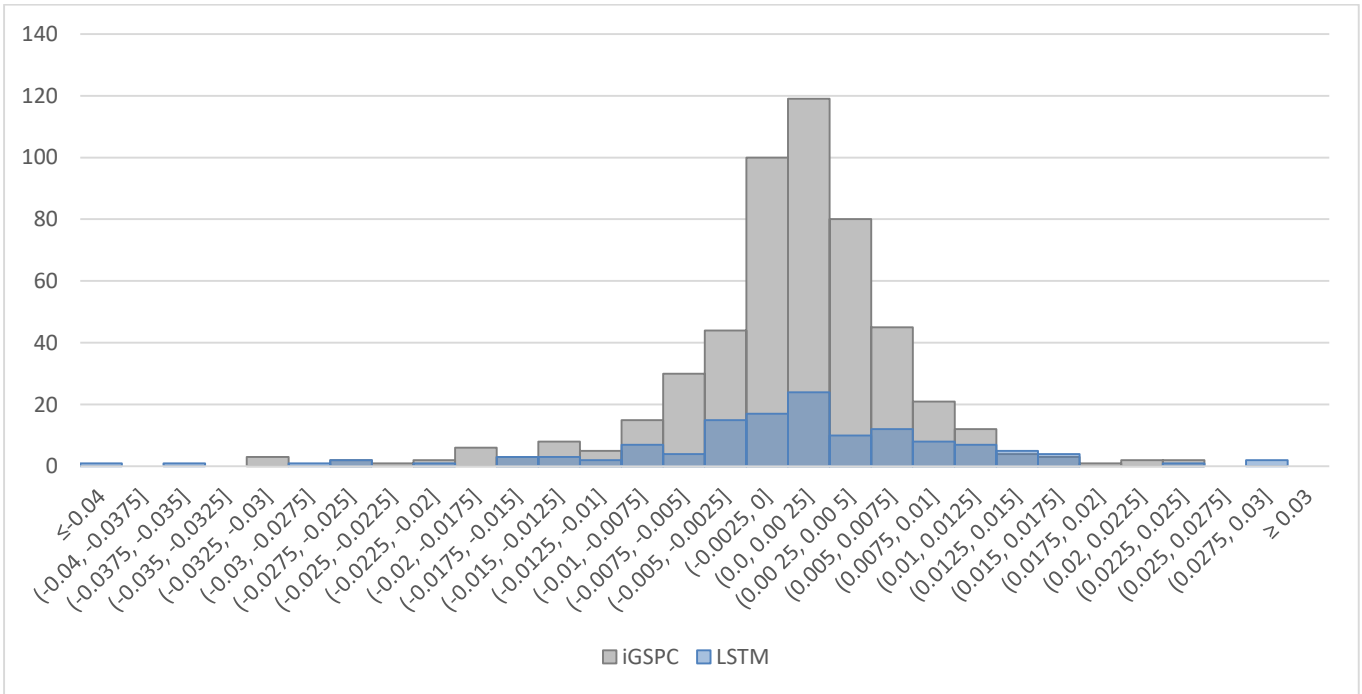
15. F. PEGAH: *Stock trend prediction using news articles a text mining approach*. 2007. ISSN: 1653-0187 [žiūrėta 2017-10-12]. Prieiga per internetą: http://www.diva-portal.org/smash/get/diva2:1019373/FULLTEXT01.pdf
16. STEPHEN W. BIGALOW: *THE MAJOR CANDLESTICKS SIGNALS*. The Candlestick Forum LLC. [interaktyvus]. [žiūrėta 2017-10-12]. Prieiga per internetą: https://stephenbigalow.com/pdfs/MajorSignals.pdf
17. SCOTT M. CARNEY. *Harmonic Trading: Volume One*. 2010. ISBN-10: 0-13-705150-6 [interaktyvus]. [žiūrėta 2017-10-12]. Prieiga per internetą: https://library.cryptotradercentral.com/Harmonic%20Trading%20Volume%201.pdf
18. THOMSETT , MICHAEL C: *Mastering Fundamental Analysis*. 1998. ISBN 0-7931-2873-0. [interaktyvus]. [žiūrėta 2017-10-12]. Prieiga per internetą: https://kupdf.com/download/thomsett-michael-c-mastering-fundamental-analysis-1998-working-pdf_58e75b74dc0d60d011da9812_pdf
19. FRANÇOIS DELOCHE: *Unfolded basic recurrent neural network*. [interaktyvus].[žiūrėta 2017-12-05]. Prieiga per: https://en.wikipedia.org/wiki/Recurrent_neural_network#/media/File:Recurrent_neural_network_unfold.svg
20. A.A.PETROSIANA, D.V.PROKHOROV, W.LAJARA-NANSONA, R.B.SCHIFFERA. *Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG*. *Clinical Neurophysiology*. 2001, vol 112, no. 8, 1378-1387 [interaktyvus] [žiūrėta 2017-12-03]. Prieiga per internetą: http://www.sciencedirect.com/science/article/pii/S138824570100579X
21. SEPP HOCHREITER, JÜRGEN SCHMIDHUBER. *Long Short-Term Memory. Neural Computation*. 1997, vol. 9, no.8, 1723-1780 [interaktyvus] [žiūrėta 2017-12-06]. Prieiga per internetą: http://www.bioinf.jku.at/publications/older/2604.pdf
22. CHRISTOPHER OLAH: *Understanding LSTM Networks.* [interaktyvus]. [žiūrėta 2018-01-20]. Prieiga per internetą: http://colah.github.io/posts/2015-08-Understanding-LSTMs/
23. WEI-YIN LOH. *Classification and regression trees*. *Data Mining and Knowledge Discovery*. Vol 1, no. 1. [interaktyvus]. [žiūrėta 2018-01-20]. Prieiga per internetą: https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.8
24. YOGESH KAKDE, SHEFALI AGRAWAL. *Predicting Survival on Titanic by Applying Exploratory Data Analytics and Machine Learning Techniques*. 2018, *International Journal of Computer Applications*. DOI: 10.5120/ijca2018917094 [interaktyvus]. [žiūrėta 2018-01-21]. Prieiga per internetą: https://www.researchgate.net/publication/325228831_Predicting_Survival_on_Titanic_by_Applying_Exploratory_Data_Analytics_and_Machine_Learning_Techniques
25. J. R. QUINALO. *Induction of Decision Trees*. 1986, *Machine Learning 1*, 81-106 [interaktyvus]. [žiūrėta 2018-01-20]. Prieiga per internetą: http://hunch.net/~coms-4771/quinlan.pdf
26. TREVOR HASTIE, ROBERT TIBSHIRANI, JEROME FRIEDMAN. *The Elements of Statistical Learning*. 2017, antras leidimas. ISBN-13: 978-0387848570 [interaktyvus]. [žiūrėta 2018-01-21]. Prieiga per internetą: https://web.stanford.edu/~hastie/ElemStatLearn//printings/ESLII_print12.pdf
27. JAMIE SHOTTON, ANDREW FITZGIBBON, MAT COOK ,TOBY SHARP, MARK FINOCCHIO, RICHARD MOORE, ALEX KIPMAN, ANDREW BLAKE. *Real-Time Human Pose Recognition in Parts from Single Depth Images*. 2013. ISBN: 978-3-642-28660-5 [interaktyvus]. [žiūrėta 2017-12-06]. Prieiga per internetą: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/BodyPartRecognition.pdf
28. RUEY S. TSAY. *Analysis of Financial Time Series*. ISBN-13: 978-0470414354.
29. DOUGLAS C. MONTGOMERY, CHERYL L. JENNINGS, MURAT KULAHCI. *Introduction to Time Series Analysis and Forecasting*. ISBN-10: 1118745116.

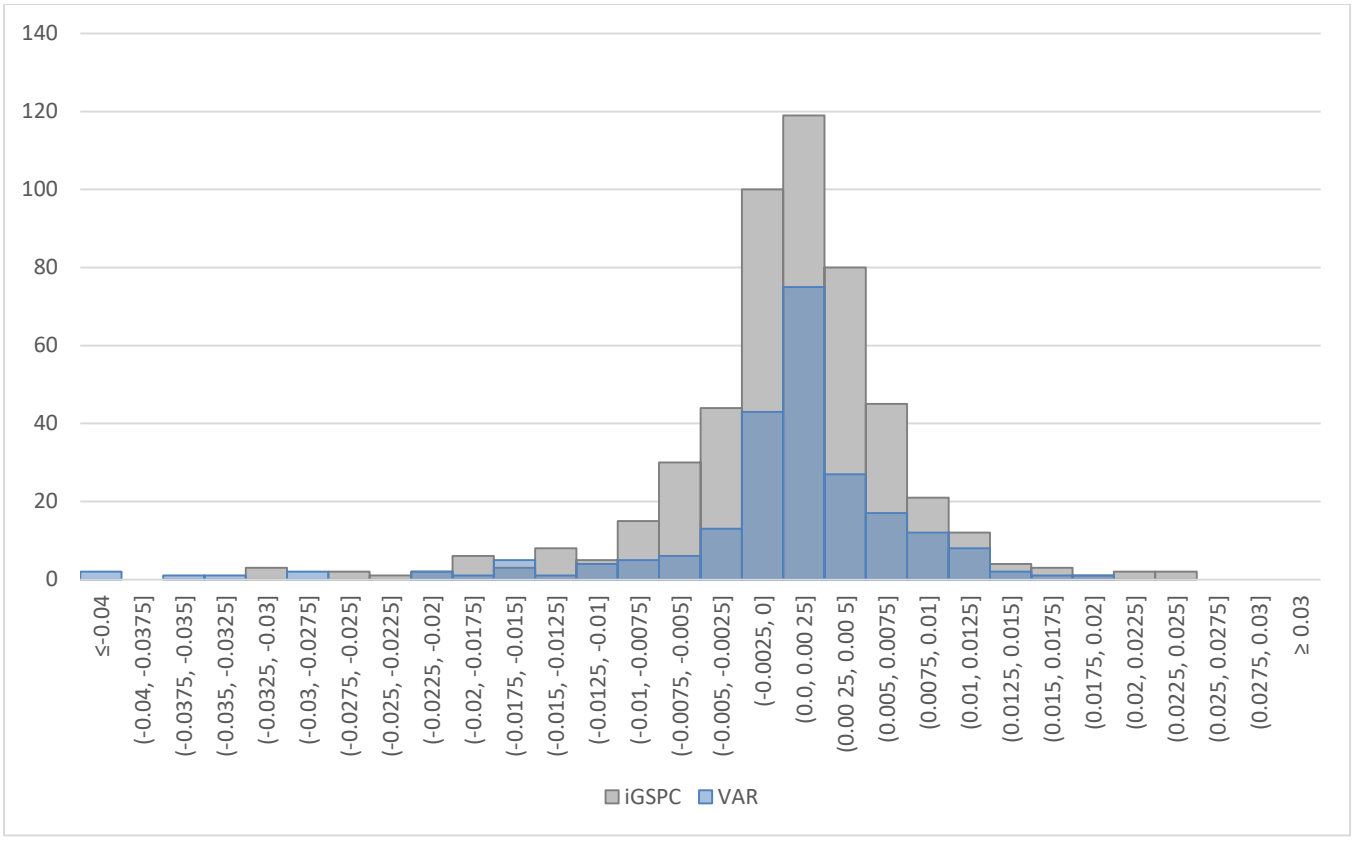
30. AGHABABAEYAN R., N. TAMANNASIDDIQUI, NAJEEBAHMADKHAN. *Forecasting the Tehran Stock Market by Artificial Neural Network*. 2011, *International Journal of Advanced Computer Science and Applications.* DOI: 10.14569 [interaktyvus]. [žiūrėta 2017-12-06]. Prieiga per internetą: https://thesai.org/Downloads/SpecialIssueNo3/Paper%203-Forecasting%20the%20Tehran%20Stock%20Market%20by%20Artificial%20Neural%20Network.pdf
31. Khan, Z. H., T. S. Alin, A. Hussain. *Price Prediction of Share Market using Artificial Neural Network (ANN)*. 2011, *International Journal of Computer Applications*, vol. 22, no. 2. DOI: 10.5120/2552-3497 [interaktyvus]. [žiūrėta 2017-12-16]. Prieiga per internetą: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.4394&rep=rep1&type=pdf
32. GENE SHER. *Evolving Chart Pattern Sensitive Neural Network Based Forex Trading Agents.* 2011. [interaktyvus]. [žiūrėta 2017-12-16]. Prieiga per internetą: https://www.researchgate.net/publication/51959188_Evolving_Chart_Pattern_Sensitive_Neural_Network_Based_Forex_TradingAgents
33. LOANN DESBOULETS. *A Review on Variable Selection in Regression Analysis*. 2018, *Econometrics*. DOI: 10.3390/econometrics6040045 [interaktyvus]. [žiūrėta 2017-12-12]. Prieiga per internetą: https://www.mdpi.com/2225-1146/6/4/45
34. G. WILLIAM SCHWERT, PAUL J. SEGUIN. *Heteroskedasticity in Stock Returns. Econometrics*. 1990, *The Journal of Finance*, vol. XLV, no. 4, 1129-1155. DOI: 10.3386/w2956 [interaktyvus]. [žiūrėta 2017-12-12]. Prieiga per internetą: https://www.nber.org/papers/w2956
35. ROB J HYNDMAN, GEORGE ATHANASOPOULOS. *Forecasting: principles and practice.* 2018. ISBN: 0987507117.
36. BOX, G.E.P., DRAPER, NORMAN R. *Empirical Model-building and Response Surfaces*. 1987, Wiley, New York, ISBN-13: 978-0471810339.
37. CEES G. H. DIKS, JASPER VRUGT: *Comparison of point forecast accuracy of model averaging methods in hydrologic applications*. 2010, *Stochastic Environmental Research and Risk Assessment*, no. 24, 809-820. DOI: 10.1007/s00477-010-0378-z [interaktyvus]. [žiūrėta 2017-12-06]. Prieiga per internetą: https://www.researchgate.net/publication/226072791_Comparison_of_point_forecast_accuracy_of_model_averaging_methods_in_hydrologic_applications
38. MICROSOFT COGNITIVE TOOLKIT (CNTK). *An open source deep-learning toolkit.* [interaktyvus]. [žiūrėta 2018-02-12]. Prieiga per internetą: https://github.com/Microsoft/CNTK

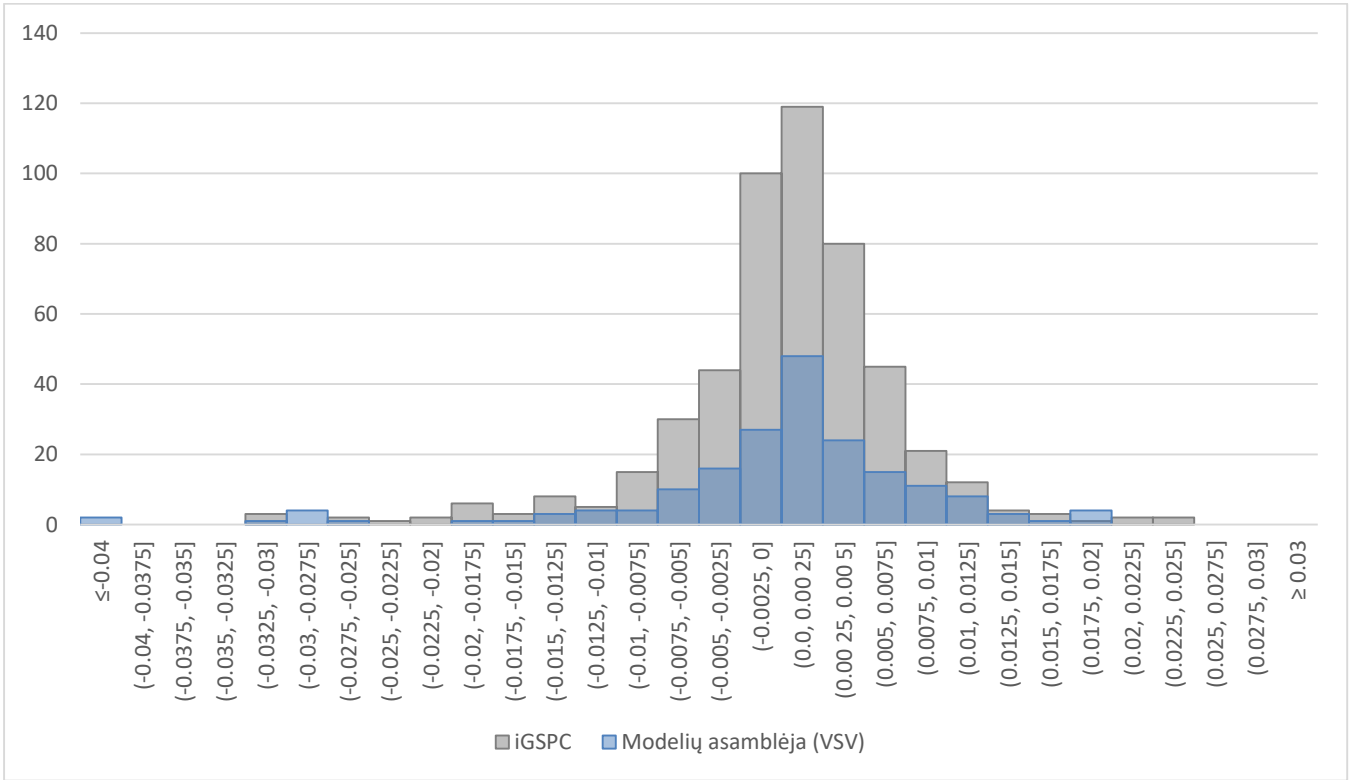
Priedai

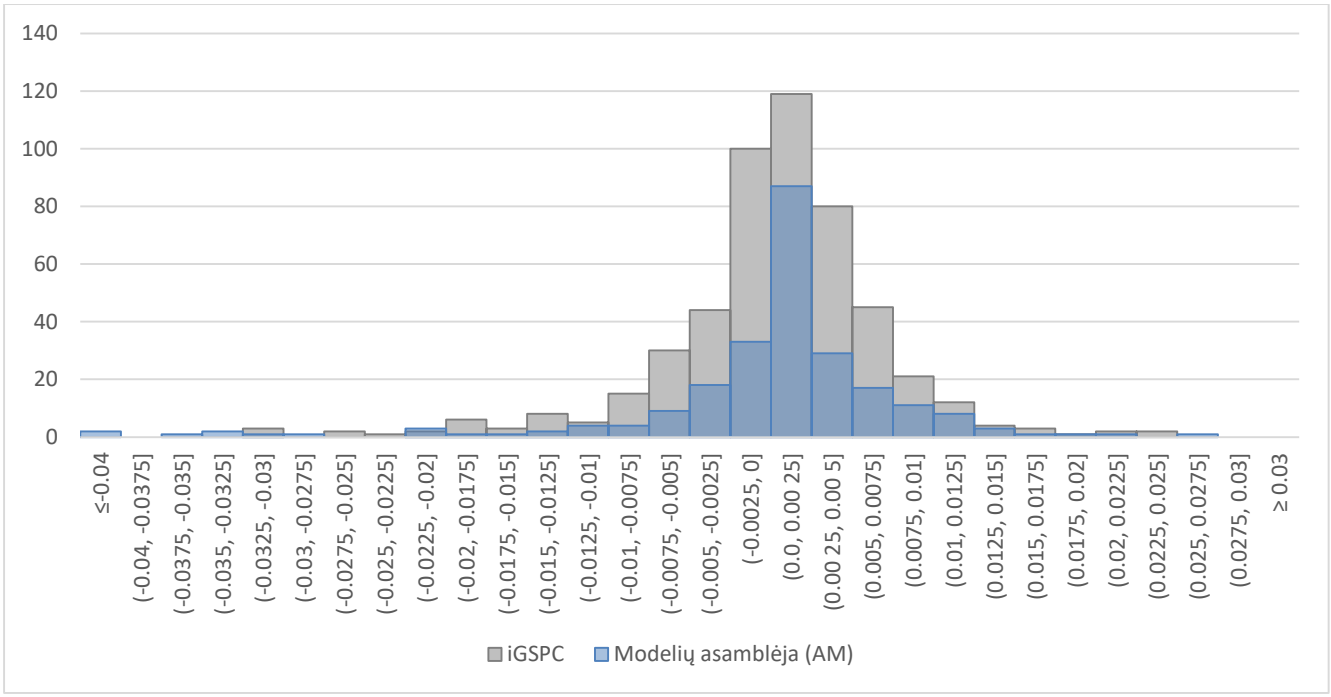
1 Priedas. iGSPC bei pelno pokyčių histogramos






140
120
100
80
60
40
20
0
≤-0.04
(-0.04, -0.0375]
(-0.0375, -0.035]
(-0.035, -0.0325]
(-0.0325, -0.03]
(-0.03, -0.0275]
(-0.0275, -0.025]
(-0.025, -0.0225]
(-0.0225, -0.02]
(-0.02, -0.0175]
(-0.0175, -0.015]
(-0.015, -0.0125]
(-0.0125, -0.01]
(-0.01, -0.0075]
(-0.0075, -0.005]
(-0.005, -0.0025]
(-0.0025, 0]
(0.0, 0.00 25]
(0.00 25, 0.00 5]
(0.005, 0.0075]
(0.0075, 0.01]
(0.01, 0.0125]
(0.0125, 0.015]
(0.015, 0.0175]
(0.0175, 0.02]
(0.02, 0.0225]
(0.0225, 0.025]
(0.025, 0.0275]
(0.0275, 0.03]
≥ 0.03
iGSPC
VAR


140
120
100
80
60
40
20
0
≤-0.04
(-0.04, -0.0375]
(-0.0375, -0.035]
(-0.035, -0.0325]
(-0.0325, -0.03]
(-0.03, -0.0275]
(-0.0275, -0.025]
(-0.025, -0.0225]
(-0.0225, -0.02]
(-0.02, -0.0175]
(-0.0175, -0.015]
(-0.015, -0.0125]
(-0.0125, -0.01]
(-0.01, -0.0075]
(-0.0075, -0.005]
(-0.005, -0.0025]
(-0.0025, 0]
(0.0, 0.00 25]
(0.00 25, 0.00 5]
(0.005, 0.0075]
(0.0075, 0.01]
(0.01, 0.0125]
(0.0125, 0.015]
(0.015, 0.0175]
(0.0175, 0.02]
(0.02, 0.0225]
(0.0225, 0.025]
(0.025, 0.0275]
(0.0275, 0.03]
≥ 0.03
iGSPC
Modelių asamblėja (VSV)


140
120
100
80
60
40
20
0
≤-0.04
(-0.04, -0.0375]
(-0.0375, -0.035]
(-0.035, -0.0325]
(-0.0325, -0.03]
(-0.03, -0.0275]
(-0.0275, -0.025]
(-0.025, -0.0225]
(-0.0225, -0.02]
(-0.02, -0.0175]
(-0.0175, -0.015]
(-0.015, -0.0125]
(-0.0125, -0.01]
(-0.01, -0.0075]
(-0.0075, -0.005]
(-0.005, -0.0025]
(-0.0025, 0]
(0.0, 0.00 25]
(0.00 25, 0.00 5]
(0.005, 0.0075]
(0.0075, 0.01]
(0.01, 0.0125]
(0.0125, 0.015]
(0.015, 0.0175]
(0.0175, 0.02]
(0.02, 0.0225]
(0.0225, 0.025]
(0.025, 0.0275]
(0.0275, 0.03]
≥ 0.03
iGSPC
Modelių asamblėja (AM)