



Kauno technologijos universitetas

Ekonomikos ir verslo fakultetas

**Mašininio mokymosi metodais grindžiamas „OMX Vilnius“
gražos indekso prognozavimas**

Baigiamasis magistro projektas

Tomas Kvietkus

Projekto autorius

Doc. dr. Andrius Grybauskas

Vadovas

Kaunas, 2024



Kauno technologijos universitetas

Ekonomikos ir verslo fakultetas

Mašininio mokymosi metodais grindžiamas „OMX Vilnius“ gražos indekso prognozavimas

Baigiamasis magistro projektas

Ekonomika (6211JX040)

Tomas Kvietkus

Projekto autorius

Doc. dr.

Andrius Grybauskas

Vadovas

Doc. dr.

Vilda Gižienė

Recenzentė

Kaunas, 2024



Kauno technologijos universitetas

Ekonomikos ir verslo fakultetas

Tomas Kvietkus

Mašininio mokymosi metodais grindžiamas „OMX Vilnius“ gražos indekso prognozavimas

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdamas kitų asmenų autoriaus ar kitų teisių, laikydamasis Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs;
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalintas iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Tomas Kvietkus

Patvirtinta elektroniniu būdu

Kvietkus, Tomas. Mašininio mokymosi metodais grindžiamas „OMX Vilnius“ gražos indekso prognozavimas. Magistro baigiamasis projektas / vadovas Doc. dr. Andrius Grybauskas; Kauno technologijos universitetas, Ekonomikos ir verslo fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Ekonomika, Socialiniai mokslai.

Reikšminiai žodžiai: mašininis mokymasis, „OMX Vilnius“ gražos indeksas, ARIMA, ARDL, LSTM, GRU, giluminis mokymasis.

Kaunas, 2024. 64 p.

Santrauka

Šiame projekte yra lyginamas mašininio mokymosi modelių, tokių kaip LSTM ir GRU, efektyvumas su klasikiniiais prognozavimo metodais, tokiais kaip ARIMA ir ARDL, prognozuojant „OMX Vilnius“ gražos indeksą. Projekto metu yra atliekama išsami literatūros apžvalga, nagrinėjanti „OMX Vilnius“ gražos indekso struktūrą, skaičiavimą ir šio indekso panaudojimą įvairiuose moksliniuose šaltiniuose. Taip pat yra atliekama plati gilaus mašininio mokymosi ir klasikinių prognozavimo metodų apžvalga. Įvertinami labiausiai naudojami modeliai šių laikų literatūroje, aprašomi šių modelių veikimo principai, parametrų konfigūravimas. Tyrimui atlikti yra naudojama 2004 -2024m. „OMX Vilnius“ gražos indekso ir 41 makroekonominio kintamojo istoriniai dieniniai duomenys. Reikšmingų kintamųjų atrinkimui yra naudojami „Granger“ priežastingumo ir „Pearson“ koreliacijos analizės tyrimai. Pasitelkus tinklo paiešką yra įvertinami įvairūs mašininio mokymosi modelių parametrai, iš kurių atrenkami geriausi. Naivusis, LSTM, GRU, ARIMA ir ARDL modeliai yra pritaikomi dieniniams ir mėnesiniams, vienmačiams ir daugiamačiams duomenims, tam, kad įvertinti visapusišką modelių efektyvumą. Iš viso yra pritaikyta 12 skirtingų modelių, kurių tikslumas yra vertinamas su naiviuoju metodu apskaičiuotais rezultatais. Galiausiai, radus tiksliausią modelį, juo remiantis yra atliekama tikra ateinančio pusmečio (30/04/2024 - 30/09/2024) prognozė. Apibendrinus, šis projektas išsamiai analizuoja „OMX Vilnius“ gražos indeksą, taip pat įvairius mašininio mokymosi ir klasikinius modelius, kurių pranašumas yra įvertinamas prognozuojant „OMX Vilnius“ gražos indeksą.

Kvietkus, Tomas. Machine Learning-Based Forecasting of the OMX Vilnius Gross Index. Master's Final Degree Project / supervisor Assoc. prof. dr. Andrius Grybauskas; School of Economics and Business, Kaunas University of Technology.

Study field and area (study field group): Economics, Social Science.

Keywords: machine learning, OMX Vilnius gross index, ARIMA, ARDL, LSTM, GRU, deep learning.

Kaunas, 2024. 64 pages.

Summary

This project compares the performance of machine learning models such as LSTM and GRU with classical forecasting methods such as ARIMA and ARDL in predicting the OMX Vilnius return index. The project includes a comprehensive literature review on the structure, calculation and use of the OMX Vilnius return index in various academic sources. An extensive review of deep machine learning and classical forecasting methods. The most common models in recent studies are reviewed, their basic principles explained, and the parameters used are described. The study uses historical daily data for the OMX Vilnius return index 2004-2024 and 41 other macroeconomic variables. Granger causality and Pearson correlation analysis are used to select the most significant variables. Various parameters of the machine learning models are evaluated by means of grid search, from which the best ones are selected. The Naive, LSTM, GRU, ARIMA and ARDL models are applied to daily and monthly, univariate and multivariate data in order to assess the full performance of these models. A total of 12 different models are fitted, after which the accuracy of each one is evaluated and compared to the results estimated by the naive method. Finally, once the most accurate model has been found, it is used as the basis for a actual forecast for the next six months (30/04/2024 - 30/09/2024). In summary, this project provides a comprehensive analysis of the OMX Vilnius gross index, as well as of the various machine learning and classical models whose comparative advantage is evaluated in forecasting the OMX Vilnius gross index.

Turinys

Lentelių sąrašas	6
Paveikslų sąrašas	7
Santrumpų ir terminų sąrašas	8
Įvadas.....	9
1. „OMX Vilnius“ gražos indekso ir gilaus mašininio mokymosi aktualumas ir problematika šių dienų kontekste.	10
1.1. Ekonominių rodiklių prognozavimo svarba Lietuvos atžvilgiu	10
1.2. Gilaus mašininio mokymosi tyrimų svarba.....	12
2. Gilaus mašininio mokymosi ir klasikinių modelių, prognozuojančių „OMX Vilnius“ gražos indekso ir kitų ekonominių rodiklių vertes, teoriniai aspektai	15
2.1. „OMX Vilnius“ gražos indeksas ir susiję moksliniai tyrimai	15
2.2. Mašininio mokymosi konceptas, jo tyrimai ir nauda, prognozuojant įvairius biržos rodiklius	17
2.3. Tradiciniai prognozavimo metodai.....	22
2.4. Dažniausiai naudojami mašininio mokymosi modeliai.....	25
2.5. Bendriniai mašininio mokymosi parametrai	29
3. Gilaus mašininio mokymosi ir klasikinių modelių, prognozuojančių „OMX Vilnius“ gražos indeksą, metodika	33
4. Gilaus mašininio mokymosi ir klasikinių modelių lyginamoji analizė, prognozuojant „OMX Vilnius“ gražos indeksą.....	40
4.1. „Granger“ priežastingumo, „Pearson“ koreliacijos ir geriausių mašininio mokymosi modelių parametrų analizė.....	40
4.2. Modelių vertinimas remiantis vienmačiais duomenimis	44
4.3. Modelių vertinimas remiantis daugiamačiais duomenimis	50
4.4. Modelių palyginimas ir tiksliausio modelio pritaikymas	54
5. Išvados	59
Literatūros sąrašas	62
Priedai.....	65
1 priedas. Dieninių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu	65
2 priedas. Mėnesinių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu	66
3 priedas. Dieninių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu.....	67
4 priedas. Mėnesinių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu	68

Lentelių sąrašas

1 lentelė. „OMX Vilnius“ gražos indekso dedamosios 2024 metų balandžio 1 dienai (NasdaqOMXNordic, 2024)	11
2 lentelė. Įvairių akcijų prognozavimo modelių RMSE tikslumas (Xiao, Feng, Yan ir Ma, 2022) .	21
3 lentelė. Tinklo paieškos parametrų derinimo imtis (Ngoc, Dai ir Phuc, 2021)	32
4 lentelė. Tyrimo metu naudojamų kintamųjų sąrašas pagal kategoriją.	34
5 lentelė. „Pearson“ koreliacijos reikšmių interpretacijos (Balabonienė, Bliekienė ir Stundžienė, 2013).....	35
6 lentelė. Tyrimo metu naudojami prognozavimo modeliai	36
7 lentelė. Tyrimo metu naudojama tinklo paieškos parametrų derinimo imtis.....	38
8 lentelė. Reikšmingų dieninių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu.....	40
9 lentelė. Reikšmingų mėnesinių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu.....	41
10 lentelė. Reikšmingų dieninių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu	42
11 lentelė. Reikšmingų mėnesinių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu.....	42
12 lentelė. Geriausi rasti tinklo paieškos parametrai mašininio mokymosi modeliams	43
13 lentelė. Naiviuoju metodu apskaičiuotų prognozių tikslumas	44

Paveikslų sąrašas

1 pav. Istorinė „OMX Vilnius“ gražos indekso vertė (NasdaqBaltic, 2024).....	12
2 pav. Mašininio mokymosi modelio konceptuali analizė (sudaryta autoriaus, remiantis Kumbure, Lohrman, Luukka ir Porras, 2022; Cerqueira, Torgo ir Mozetic 2020; Nevasalmi 2020).....	18
3 pav. Grafinė 1 paslėpto sluoksnio neuronų tinklų reprezentacija (Chan ir Mátyás, 2022).....	25
4 pav. Grafinė RNN modelio architektūra (Dautel, Härdle, Lessmann ir Seow, 2020)	26
5 pav. RNN, LSTM ir GRU modelių vidutinės absoliutinės procentinės paklaidos (Saud ir Shakya, 2020).....	26
6 pav. Grafinė RNN modelio ląstelės struktūra (Olah, 2015).....	27
7 pav. Grafinė LSTM modelio ląstelės struktūra (Olah, 2015)	28
8 pav. Grafinė GRU modelio ląstelės struktūra (Shen ir kt., 2018).....	29
9 pav. Tiksliausias dieninių duomenų pagal AIC kriterijų išgautas modelis, naudojant „auto.arima“	45
10 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį SARIMAX(3,1,1)(2,0,0)[7] modelį.....	45
11 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį LSTM modelį	46
12 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį GRU modelį	47
13 pav. Tiksliausias dieninių duomenų pagal AIC kriterijų išgautas modelis, naudojant „auto.arima“	48
14 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį SARIMAX(1,1,1)(0,1,1)[7] modelį.....	48
15 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį LSTM modelį	49
16 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį GRU modelį	49
17 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį ARDL(12,13,8,4) modelį	51
18 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį LSTM modelį	51
19 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį GRU modelį	52
20 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį ARDL(12,1,6,2) modelį	53
21 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį LSTM modelį.....	53
22 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį LSTM modelį.....	54
23 pav. Prognozavimo modelių MAPE rodiklio palyginimas prognozuojant dienes „OMX Vilnius“ gražos indekso vertes.....	55
24 pav. Prognozavimo modelių MAPE rodiklio palyginimas prognozuojant mėnesines „OMX Vilnius“ gražos indekso vertes.	56
25 pav. „OMX Vilnius“ gražos indekso mėnesinių duomenų prognozė 31/03/2024 - 30/09/2024 periodui naudojant LSTM modelį	57

Santrumpų ir terminų sąrašas

Santrumpos:

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

ARDL – autoregresinis paskirstyto vėlavimo modelis (angl. *autoregressive distributed lag*);

ARIMA - autoregresinis integruotas slenkamasis vidurkis (angl. *autoregressive integrated moving average*);

GRU – Sklendžių rekurentinių vienetų modelis (angl. *gated recurrent units model*);

LSTM – ilgalaikės-trumpalaikės atminties tinklas (angl. *long short-term memory networks*);

MAPE – vidutinė absoliuti procentinė paklaida (angl. *mean absolute percentage error*);

Min-Max metodas – minimalaus-maksimalaus mastelio keitimo metodas (angl. *Min-Max scaler*).

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

RMSE – vidutinė kvadratinės šaknies paklaida (angl. *root mean squared error*);

RNN – rekurentiniai neuroniniai tinklai (angl. *recurrent neural network*);

SVM - atraminių vektorių klasifikatorius (angl. *support vector machine*);

Terminai:

Epochos - tai yra skaičius, kuris nustato, kiek kartų modelis turi praeiti visą duomenų imtį (angl. *epochs*);

Lookback – istorinių duomenų imtis, naudojama kiekvienai prognozuojamai vertei;

Naivusis prognozavimo metodas – tai vertinimo metodas, kai praėjusio laikotarpio faktiniai duomenys naudojami kaip šio laikotarpio prognozė, jų nekoreguojant ir nebandant nustatyti priežastinių veiksnių (angl. *naive*)

Optimizacijos algoritmai - tai algoritmai arba metodai, naudojami neuroninio tinklo atributams, pavyzdžiui, svoriams ir mokymosi greičiui, keisti, siekiant sumažinti modelio prognozavimo paklaidą (angl. *optimizer algorithms*)

Partijos dydis - tai duomenų imties, apdorotų prieš atnaujinant modelio vidinius parametrus, skaičius (angl. *batch size*);

Paslėptas dydis – neuronų kiekis kiekviename paslėptame sluoksnyje (angl. *hidden size*);

Paslėpti sluoksniai - tarpiniai sluoksniai tarp įvesties ir išvesties, kurie apdoroja duomenis iš ankstesnio sluoksnio ir perduoda juos kitam sluoksniui (angl. *hidden layers*);

Perteklinis pritaikymas – tai kai modelis vietoje to, kad išmoktų bendrus kintamojo judėjimus, juos tiesiog įsidėmi, tai lemia, kad mokymosi procese modelis būna labai tikslus, o testavimo procese priešingai (angl. *overfit*).

Įvadas

Aktualumas. Ekonominių rodiklių prognozavimas yra svarbus tiek privačiu, tiek visuomeniniu lygiu, ar tai būtų įvairios įmonės, sekančios savo apyvartą, ar tai būtų politikos formuotojai, vertinantys valstybės ekonominę būklę, sudaryti tikslias prognozes yra būtina, norint tinkamai pasiruošti būsimiems įvykiams. Pastaruoju metu mašininis mokymasis vis dažniau naudojamas kaip vertingas įrankis įvairiose srityse dėl savo puikių analizės galimybių, ne išimtis ir ekonomikos mokslo disciplina (Jaya, 2023; Nevasalmi, 2020; Omar, Huang, Salameh, Khurram ir Fareed, 2022). Ši sparčiai tobulėjanti technologija siūlo alternatyvą klasikiniams prognozavimo modeliams, tokiems kaip slenkamieji vidurkiai, tiesinė regresija ar autoregresinis integruotas slenkamasis vidurkis. Atsižvelgiant į tai, jog mašininis mokymasis šiais laikais vis labiau populiarėja (Kumbure, Lohrman, Luukka ir Porras, 2022), yra svarbu įvertinti ar mašininio mokymosi paremti prognozavimo modeliai iš tiesų gali pateikti tikslesnius rezultatus nei klasikiniai metodai. Todėl norint sužinoti ar mašininio mokymosi modeliai gali prognozuoti tiksliau nei tradiciniai modeliai, pirmiausia reikia atrasti tinkamiausius mašininio mokymosi modelius ekonominių rodiklių prognozavimui. Tam, kad šis tyrimas būtų aktualus ir Lietuvos rinkai, tiek mašininio mokymosi, tiek klasikinio prognozavimo modeliai bus vertinami prognozuojant „OMX Vilnius“ gražos indeksą.

Mokslinė problema – Ar gilusis mašininis mokymasis gali pasiekti tikslesnes „OMX Vilnius“ gražos indekso prognozes lyginant su klasikiniiais prognozavimo metodais?

Projekto tikslas – atlikti skirtingų gilaus mašininio mokymosi ir klasikinių prognozavimo modelių palyginamąją analizę, prognozuojant „OMX Vilnius“ gražos indeksą.

Tyrimo uždaviniai:

1. Įvertinti „OMX Vilnius“ gražos indekso ir gilaus mašininio mokymosi aktualumą šių dienų kontekste;
2. Atlikti mokslinės literatūros analizę, nagrinėjant teorinius ir empirinius mašininio mokymosi ir klasikinių prognozavimo modelių panaudojimo tyrimus, „OMX Vilnius“ gražos indekso ir kitų ekonominių rodiklių atžvilgiu;
3. Remiantis literatūros analize pasirinkti tyrimo metodus „OMX Vilnius“ gražos indekso prognozavimui ir modelių vertinimui;
4. Atlikti gilaus mašininio mokymosi ir klasikinių prognozavimo modelių palyginamąją analizę „OMX Vilnius“ gražos indekso atžvilgiu;
5. Remiantis tiksliausiu rastu modeliu atlikti „OMX Vilnius“ gražos indekso realią prognozę ateinančiam laikotarpiui.

Tyrimo metodai. Projekto metu buvo atlikta mokslinės literatūros analizė ir sisteminimas, statistinių duomenų rinkimas, jų analizė ir perdirdimas, įvairių prognozavimo modelių vertinimas, palyginimas ir testavimas, įvairių duomenų imčių vertinimas prognozių tikslumui, tiksliausio modelio atrinkimas ir jo pritaikymas prognozuojant „OMX Vilnius“ gražos indeksą. Tyrimui atlikti buvo naudojamos tokios programos kaip „Excel“, „Python“ ir „Eviews“.

1. „OMX Vilnius“ gražos indekso ir gilaus mašininio mokymosi aktualumas ir problematika šių dienų kontekste.

Ekonominių rodiklių prognozavimas net ir šiais laikais išlieka labai sudėtinga užduotis. Taip yra dėl to, kad akcijų, obligacijų, užsienio valiutų, nekilnojamo turto bei kitos rinkos yra paremtos daugybės makro ir mikro ekonomikos veiksnių. Kuo rinkos kapitalizacija didesnė, tuo daugiau kintamųjų daro įtaką jos judėjimui. Nors tiksliai prognozuoti šiuos rodiklius atrodo neįmanoma, prancūzų matematikas, astronomas ir fizikas Pierre-Simon Laplace (1902, p. 4) yra pasakęs „Duodant akimirką intelektui, kuris suvokia visas gamtos jėgas, ir būtybių, kurios jas sudaro, situacijas – intelektas, pakankamai platus, galintis šiuos duomenis pateikti analizei – jis apimtų toje pačioje formulėje didžiausių visatos kūnų ir lengviausių atomų judėjimus; jam niekas nebūtų abejotina, ir ateitis, ir praeitis būtų jam prieš jo akis“. Taigi, ateities nuspėjimas susideda iš dviejų dedamųjų - pakankamo kiekio duomenų ir pakankamai protingo intelekto. Kadangi visų duomenų, nuo atomų judėjimo, šiais laikais dar nėra įmanoma surinkti, o intelekto, žinančio visas visatos paslaptis, dar nėra, tenka apsiriboti tuo, ką turime. O arčiausiai to, šiais laikais, yra mašininio mokymosi paremti modeliai.

1.1. Ekonominių rodiklių prognozavimo svarba Lietuvos atžvilgiu

Ekonominės analitikos srityje tikslus ekonominių duomenų prognozavimas yra kritiškai svarbus ekonomikos planavimo ir strateginių investicinių sprendimų priėmimo atžvilgiu. Sparčiai besivystančiai Lietuvos ekonomikai gebėjimas numatyti ekonomines tendencijas yra ne tik naudingas, bet ir būtinas. Toks prognozavimas suteikia vertingų įžvalgų apie šalies ekonomikos būklę, padeda formuoti investicinius sprendimus ir atlieka lemiamą vaidmenį formuojant monetarinę ir fiskalinę politiką. Lietuvos kontekste itin svarbu sutelkti dėmesį į vietinių rodiklių, tokių kaip „OMX Vilnius“ gražos indekso, prognozavimą. Būdamas vienas pagrindinių Lietuvos vertybinių popierių rinkos indeksų, „OMX Vilnius“ gražos indeksas atspindi bendrą Lietuvos vertybinių popierių rinkos dinamiką. Tikslios „OMX Vilnius“ gražos indekso prognozės gali suteikti investuotojams, tiek Lietuvoje, tiek užsienyje, kritiškai svarbios informacijos, būtinos priimant pagrįstus investavimo sprendimus. Tokios įžvalgos taip pat yra naudingos ir Lietuvos ekonomikai, kadangi jos skatina rinkos tinkamą pasiruošimą ateities įvykiams ir ekonominį stabilumą.

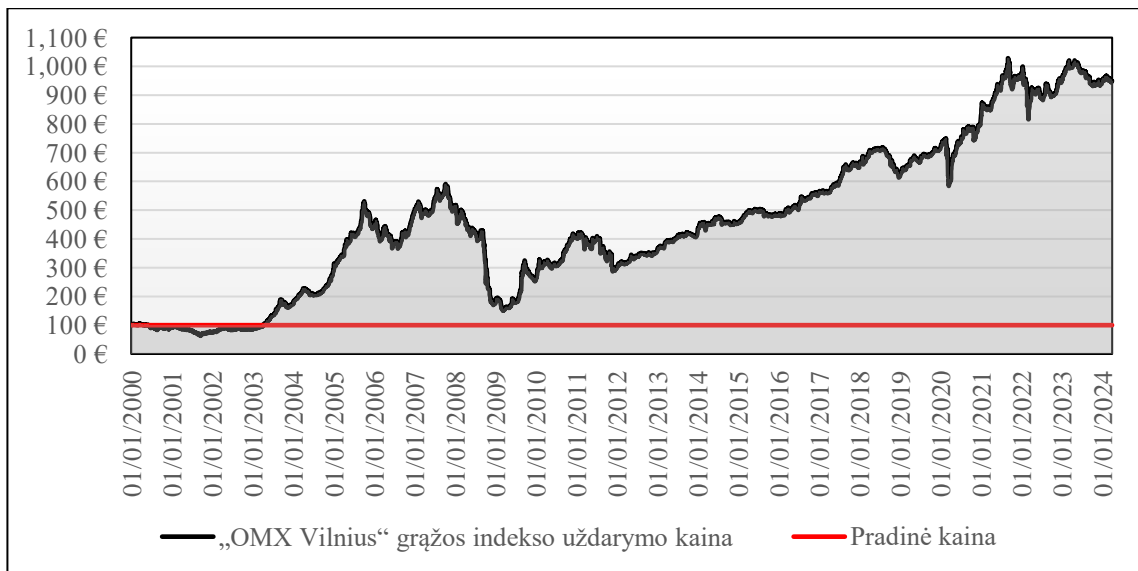
Atsižvelgiant į Lietuvos ekonominius rodiklius, o ypač į akcijų biržą, bene geriausiai Lietuvos įmonių veiklą atspindintis rodiklis, kaip ir minėta, būtų „OMX Vilnius“ gražos indeksas. Šis indeksas įtraukia visas listinguojamas akcijas Vilniaus akcijų biržoje. Šio indekso tikslas yra parodyti bendrus Lietuvos ekonomikos pokyčius. „OMX Vilnius“ gražos indeksas mėgdžioja visų Vilniaus akcijų biržoje esančių akcijų judėjimą, neįtraukiant likvidumo ar stabilumo reikalavimų. Kadangi nėra likvidumo filtravimo, patį indeksą gali būti sudėtinga lyginti su jo sudedamosiomis dalimis, kadangi gali pasitaikyti nedažnai prekiaujančių akcijų tarp jo dedamųjų (Nasdaq, *n. d. a*). Remiantis NasdaqBaltic (*n. d.*) gražos indeksas parodo bendrąją į jį įtrauktų akcijų gražą. Jis atspindi tiek kainų pokyčius, tiek mokamus dividendus, remiantis prielaida, kad į akcijas bus pakartotinai investuojama. Todėl šis indeksas yra laikomas išsamesne rinkos rezultatų priemone nei kainos indeksas. Taip pat yra pabrėžiama, kad yra įtraukiamos tik tos akcijos, kurių individualūs akcininkai turi mažiau nei 90% išleistų akcijų. (NasdaqBaltic, *n. d.*). Remiantis 2024 metų balandžio 28d. duomenimis, į šio indekso apskaičiavimą buvo įtraukta 21 įmonė (žr. 1 lentelę).

1 lentelė. „OMX Vilnius“ gražos indekso dedamosios 2024 metų balandžio 1 dienai (NasdaqOMXNordic, 2024)

Nr.	Įmonės pavadinimas	Akcijos trumpinys
1.	Akola group	AKO1L
2.	Apranga	APG1L
3.	AUGA group	AUG1L
4.	AB Grigeo	GRG1L
5.	Ignitis grupė AB	IGN1L
6.	INVL TECHNOLOGY	INC1L
7.	INVL BALTIC FARMLAND	INL1L
8.	INVL Baltic Real Estate	INR1L
9.	Invalda INVL AB	IVL1L
10.	KN Energies	KNE1L
11.	Novaturas AB	NTU1L
12.	Panevėžio Statybos Trestas	PTR1L
13.	Pieno Žvaigždės	PZV1L
14.	Rokiškio Sūris	RSU1L
15.	Šiaulių Bankas	SAB1L
16.	Telia Lietuva	TEL1L
17.	Utenos Trikotažas	UTR1L
18.	Vilniaus Baldai	VBL1L
19.	Vilkyškių Pieninė	VLP1L
20.	Žemaitijos Pienas	ZMP1L

Kaip matyti 1 lentelėje, šis indeksas įvertina įvairių sektorių veiklą, nuo agrokultūros iki telekomunikacijų technologijų. Todėl „OMX Vilnius“ gražos indeksas, apimantis įvairias Lietuvos įmonių akcijas, parodo visapusišką rinkos rezultatų vaizdą ir veiksmingai atspindi sudėtingą šalies ekonomikos dinamiką, naudodamas bendrą gražos rodiklį.

Žvelgiant iš rinkos vertės pusės, kaip matoma istoriniuose duomenyse (žr. 1 pav.), kadangi yra apskaičiuojama bendra akcijų graža nuo laikotarpio pradžios, indekso vertė metai iš metų kilo. Kol kas vienintelis reikšmingas vertės kritimas buvo matomas tik 2008-2009 metų ekonominės krizės metu, po kurio „OMX Vilnius“ gražos indeksas užtruko beveik 9 metus atsistatyti iki buvusios gražos. Nors artimesniu laiku įvykęs nuosmukis dėl COVID - 19 neatrodo toks reikšmingas kaip buvusios krizės, matoma, jog dėka prasidėjusio karo Ukrainoje, „OMX Vilnius“ gražos indeksas dar neatsistatė iki savo aukštumų. Bendrai žiūrint, globali ekonominė situacija dėl įvairių geopolitinių veiksmų šiuo metu vis dar nėra labai stabili. Dėl šios priežasties yra naudinga tinkamai vertinti ir nuspėti ateities įvykius, liečiančius Lietuvos ekonomiką ir jos įmones.



1 pav. Istorinė „OMX Vilnius“ gražos indekso vertė (NasdaqBaltic, 2024)

Nors daugelis šaltinių atliekant prognozes mėgsta vertinti didelius JAV ekonominius vienetus, tokius kaip „S&P 500“, „Dow Jones“, „Nasdaq“ (Kumbure, Lohrman, Luukka ir Porrás, 2022), kuo rinkos rodiklis apima didesnę vertę, tuo jis daugiau yra paremtas jį veikiančių kintamųjų. Todėl tai buvo viena iš priežasčių orientotis į šiek tiek siauresnę, mažesnę rinką, kuri yra pakankamai didelė, kad būtų daroma įtaka tiek makro, tiek mikro faktoriams, tačiau nebūtų tokia didelė, jog kintamųjų skaičius būtų begalinis. Apibendrinus, „OMX Vilnius“ gražos indeksas apima ir parodo bendrą Lietuvos ekonomiką įmonių vertės atžvilgiu. Atsižvelgiant į dar ne visiškai stabilią ekonominę situaciją dėl praėjusios pandemijos ir pasaulyje esančių nestabilumų, yra naudinga ne tik žvelgti ir vertinti istorinius įvykius, tačiau ir sudaryti galimas ateities prognozes. Dėl šių priežasčių darbo metu buvo pasirinkta analizuoti ir prognozuoti būtent šį indeksą.

1.2. Gilaus mašininio mokymosi tyrimų svarba

Bėgant metams ir mašininio mokymosi metodams tampant vis labiau prieinamiems, prognozavimas, pasitelkiant šiuos įrankius, vis populiarėja. Viena iš plačiau nagrinėjamų sričių yra rinkos rodikliai. Nepaisant to, jog tikslios rinkos rodiklių prognozės gali suteikti milžinišką vertę tiek investuotojams, tiek įmonėms, ši sfera išlieka populiari ir dėl savo kompleksiško. Rinkos rodikliai tipiškai priklauso nuo daugybės mikro ir makro kintamųjų, iš kurių daugelio šiais laikais net neįmanoma įvertinti. Atsižvelgiant būtent į akcijų rinką, ji susideda iš neapibrėžtumo ir kintamumo, kas apsunkina tikslios prognozės gavimą (Zou ir kt., 2022). Apibendrinus, šios srities tyrimai yra naudingi dėl dviejų priežasčių: pirma - tikslios prognozės gali suteikti milžinišką finansinę vertę tiek individualiems asmenims, tiek įmonėms, antra - ši sritis skatina inovacijas ir naujus sprendimo būdus, kadangi gauti tikslias prognozes, neturint visų įmanomų kintamųjų, yra labai sudėtinga užduotis.

Analizuojant šią temą, buvo įvertinti įvairūs šaltiniai, kurių autoriai pabrėžė tam tikras nagrinėtinas sritis ateities tyrimams. Jaya (2023) atliko išsamų tyrimą nagrinėjant mašininio mokymosi naudą finansinių rinkų prognozavime. Šio šaltinio autorius nagrinėjo logistinės regresijos, SVM, LSTM ir dvikrypčio transformacinio kodavimo būdo (angl. *bidirectional encoder representations from transformers*, BERT) bei finBERT (iš anksto apmokytas modelis skirtas finansinio teksto sentimentams analizuoti) apjungtą modelių tikslumą, prognozuojant metų ir pusmečio „Microsoft“, „Amazon“, „Apple“ ir „S&P“ rinkos akcijų vertės pokyčio kryptį. Nuspėjant krypties trajektoriją

autorius pasitelkė ne tik tradicinę logistinės regresijos ar gilaus mašininio mokymosi modelius kaip LSTM ir SVM, tačiau pasitelkė ir išmaniuosius kalbos modelius, kurie galėtų sekti ir analizuoti rinkos sentimentus, nagrinėjant „Reuters“ straipsnius. Tarp gautų rezultatų, pasitelkiant SVM ir BERT modelius, autorius sugebėjo pasiekti net 99,5% prognozės tikslumą nuspėjant vertės kitimo trajektoriją. Tačiau net ir su tokio tikslumo rezultatais yra tobulintinų vietų. Šaltinyje yra siūloma taip pat panagrinėti tikslumus nesujungiant SVM/LSTM modelių su BERT kalbos modeliais. Yra rekomenduojama pasitelkti ir paros duomenis, kas galimai galėtų suteikti dar tikslesnius rezultatus. Galiausiai yra siūloma eksperimentuoti ir su modelio struktūra bei jo parametrais, tikintis dar labiau optimizuoti modelį ir išgauti dar tikslesnius rezultatus.

Omar'as, Huang'as, Salameh'as, Khurram'as ir Fareed'as (2022) atliko tyrimą, kurio metu pasitelkiant mašininio mokymosi modelius buvo atliekama Karachi Stock Exchange (KSE-100) indekso dieninių duomenų prognozė prieš ir per Covid -19 pandemiją. Atlikus tyrimą autoriai padarė kelias išvadas. Viena iš jų - giliojo neuronų tinklo (angl. *deep neural network*, DNN) modelių pranašumas prognozėse, ypač lyginant su ARIMA. Visais laiko tarpais atliekant tyrimą ARIMA rezultatai buvo mažiausiai tikslūs. Taip pat tyrimo metu buvo išaiškinta, kad kol DNN modeliai veikia gerai su dideliu kiekiu duomenų, jų tikslumas yra mažesnis nei atsitiktinio miško (angl. *random forest*, RF) modelių, jeigu duomenų kiekis yra pakankamai mažas. Todėl buvo padaryta išvada, kad RF modeliai labiau tinkami mažam kiekiui duomenų, kol DNN modeliai tinkami didesniai kiekiui duomenų. Atsižvelgiant į tolimesnių tyrimų rezultatus, autoriai rekomendavo toliau atlikti panašaus pobūdžio tyrimus, ypač nagrinėjant kitus mašininio mokymosi modelius, tokius kaip SVM, LSTM ar GRU modelius. Taip pat prognozė buvo atlikta tik Azijos rinkoje, todėl autoriai rekomendavo atlikti tolimesnius tyrimus ir kitose akcijų rinkose (Omar, Huang, Salameh, Khurram ir Fareed, 2022).

Atsižvelgiant į akcijų gražos prognozes, Nevasalmi'io (2020) atliktas tyrimas nagrinėjo naują daugianarį (angl. *multinomial*) gražos apskaičiavimo būdą. Įprastai finansinėje literatūroje akcijų graža yra lyginama tarp aukštos (pvz. S&P 500) ir žemos (3 mėn. vyriausybės obligacijos) rizikos kainų skirtumo. O sprendimai yra priimami vertinant tik akcijos pirkimą arba pardavimą. Tačiau šio tyrimo metu į lygtį buvo įtraukta daugiau kintamųjų, buvo vertinamas ne tik pirkimas ar pardavimas, tačiau ir akcijos laikymas, vietoje to, kad orientuotis į nedidelius dieninius kainų kitimus, buvo labiau kreipiamas dėmesys į didelius kainų pokyčius. Tyrimo metu analizuojant duomenis paaiškėjo, kad didžiausią gražą nešantis modelis buvo gradiento didinimo modelis (angl. *gradient boosting model*, GBM), tačiau, net ir turint veikiantį modelį, autoriai pabrėžia tolimesnes tyrimo sritis, kurios galėtų pagerinti būsimų tyrimų rezultatus. Autoriai rekomendavo tarp kintamųjų įtraukti „bull“ ir „bear“ sertifikatus, kurie padėtų patikslinti prognozių modelius. Taip pat buvo rekomenduojama nenaikinti išskirčių, kadangi jose būna didžiausi akcijų judėjimai. Galiausiai autoriai pabrėžia, kad prognozavimo, paremto statistinių duomenų vertinimu, sąsaja su prekiavimo strategijų ekonominiu pelningumu turėtų būti giliau nagrinėjama (Nevasalmi, 2020).

Atsižvelgiant į lietuvių atliktus tyrimus, Džikevičius ir Stabužytė (2012) atliko net labai panašaus pobūdžio tyrimą į šį. Autoriai taip pat prognozavo „OMX Vilnius“ gražos indekso vertes, pasitelkiant neuronų tinklus. Skirtumas tik tas, kad tai buvo 2012 metai, technologijos per 12 metų pasikeitė, atsirado naujų modelių, galingesnių kompiuterių ir patogesnių įrankių šiai užduočiai atlikti. Tyrimo metu buvo analizuojami dieniniai ir mėnesiniai 01.01.2000 – 30.04.2012 metų duomenys, kurie buvo normalizuoti ir paruošti daugiasluoksnio perceptrono (angl. *multi-layer perceptron*, MLP) modelio apmokymui. Duomenys buvo padalinti į mokymo (90%) ir testavimo (10%) grupes, o tikslumo vertinimui buvo pasirinktas vidutinis absoliutus nuokrypis (angl. *mean absolute deviation*, MAD),

kvadratinės paklaidos suma (angl. *sum of squared errors*, SSE), MSE ir MAPE. Nors tyrimo rezultatai ir nebuvo įspūdingi prognozuojant dienuosius duomenis (geriausias modelis sugebėjo pasiekti tik 53,06% tikslumą prognozuojant judėjimo trajektoriją, 0,74 MAPE prognozuojant kainą), MLP modelis vis tiek sugebėjo būti tikslesnis nei daugialypis regresijos modelis arba paprastasis slenkamasis vidurkis. Tačiau ne visi rezultatai buvo teigiami, tyrimo metu paaiškėjo, kad mėnesinės prognozės buvo labai prastos. Šioje vietoje daugialypė regresija sugebėjo pranokti visus MLP prognozuojant kainą. Mėnesinėse prognozėse MLP modelis pranoko tik nuspėjant judėjimo trajektoriją. Atlikus tyrimą autoriai taip pat rekomendavo tiriamąsias sritis, buvo siūloma nagrinėti daugiau kintamųjų, ne tik vėlintas „OMX Vilnius“ gražos indekso reikšmes. Taip pat šis šaltinis nagrinėja tik MLP modelį, dėl šios priežasties yra rekomenduojama įvertinti ir kitus modelius (Dzikevičius ir Stabužytė, 2012). Žinoma, kadangi šis šaltinis yra 2012 metų, mašininio mokymosi prognozavimo sfera smarkiai pakito per pastaruosius 12 metų. Ir iš nagrinėjamų kitų šaltinių galima pastebėti, kad pasitelkus sudėtingesnius ir platesnius modelius tikrai galima išgauti geresnius rezultatus. Todėl būtų naudinga, atnaujinti šį tyrimą ir papildyti jį naujais giliuoju mašininio mokymosi paremtais prognozavimo būdais.

Apibendrinus literatūrą, mašininio mokymosi prognozavimo tema yra plati, tačiau nepakankamai gausi, kad visi klausimai jau būtų išspręsti. Nors yra atlikta daug tyrimų, dėl šios srities dinamikos, naujos technologijos, nauji sprendimo būdai ir vis didėjanti duomenų gausa skatina ne tik atnaujinti senus tyrimus, tačiau ir kurti naujus, modifikuojant įvairius parametrus ir pritaikant naujus sprendimo būdus. Dėl šių priežasčių ir buvo nuspręsta toliau gilinti šią temą.

Taigi, apibendrinus nagrinėtus šaltinius, mašininis mokymasis yra nuolat tobulėjanti technologija, kuri, ypač per pastaruosius metus, dėl technologinio prieinamumo, naujų inovacijų ir vis gerėjančių rezultatų, smarkiai išpopuliarėjo (Kumbure, Lohrman, Luukka ir Porras, 2022). Dzikevičius ir Stabužytė (2012) savo darbe pabrėžia, jog tolimesni tyrimai gali būti pagerinti naudojant daugiau kintamųjų arba koreguojant vėlavimo metodus. Taip pat yra siūloma panagrinėti naujesnius neuronų tinklų modelius prognozuojant „OMX Vilnius“ gražos indeksą. Jaya (2023) pateikia siūlymą prognozuoti taip pat naudojant LSTM modelį, tačiau šįkart naudojant tik vieno požymio rinkinį arba padidinus duomenų dažnumą į valandinius/minutinius intervalus. Omar‘as, Huang‘as, Salameh‘as, Khurram‘as ir Fareed‘as (2022) rekomenduoja atlikti tyrimus pasitelkiant SVM, LSTM ir GRU modelius, taip pat atsižvelgti ne tik į Azijos, tačiau ir kitas rinkas atliekant mašininio mokymosi grįstas prognozes. Nevassami‘is (2020) rekomenduoja ateities tyrimuose į daugianarius duomenis įtraukti „Bull“ ir „Bear“ sertifikatus, kurie galėtų padidinti prognozių tikslumą. Taip pat atsižvelgus į Lietuvos rinką, „OMX Vilnius“ gražos indeksas yra svarbus tuo, kad jis atspindi šalies akcijų rinkos būklę ir judėjimo tendencijas. Dėl šios priežasties tikslios šio indekso prognozės suteikia naudą ne tik priimant pagrįstus investicinius sprendimus, bet ir rengiant ekonomines strategijas, kurios gali atlikti svarbų vaidmenį formuojant monetarinę ir fiskalinę Lietuvos politiką. Todėl apibendrinus šių šaltinių poreikius tirti minėtomis temomis ir atsižvelgiant į aktualumą Lietuvos rinkai, buvo pasirinkta nagrinėti mašininio mokymosi metodais grįsto „OMX Vilnius“ gražos indekso prognozavimo temą.

2. Gilaus mašininio mokymosi ir klasikinių modelių, prognozuojančių „OMX Vilnius“ gražos indekso ir kitų ekonominių rodiklių vertes, teoriniai aspektai

Pasitelkiant įvairią literatūrą šiame darbe bus analizuojamas „OMX Vilnius“ gražos indeksas ir jo prognozavimas, taip pat dėl nedidelės tyrimų gausos, susijusios būtent su šiuo indeksu, bus vertinami ir kiti rinkos rodikliai prognozių atžvilgiu. Didelis dėmesys taip pat skiriamas analizuoti mašininio mokymosi įvairovę, bus nagrinėjamos jo metodikos, taikymo sritys ir pažanga. Galiausiai bus nagrinėjami pagrindiniai šios srities tyrimai ir indėlis, pateikiama įžvalgų apie mašininio mokymosi metodų panaudojimą ir jų vertę prognozuojant ekonominius rodiklius.

2.1. „OMX Vilnius“ gražos indeksas ir susiję moksliniai tyrimai

Pradedant nuo „OMX Vilnius“ gražos indekso, svarbu panagrinėti kaip jis yra skaičiuojamas, tam, kad geriau suprasti, kas galėtų turėti įtakos jo judėjimui. Taigi, atsižvelgiant į indekso apskaičiavimą, remiantis NasdaqBaltic (2016a) ir NasdaqBaltic (2016b) jis yra apskaičiuojamas pasitelkiant „Nasdaq“ bendrosios gražos (angl. *gross total return*, GTR) formulę. Kaip ir minėta, šis indeksas apskaičiuoja bendrą akcijų judėjimo pokytį per prekiaujamą dieną. Pirmiausia, kad galima būtų apskaičiuoti GTR, reikia pasiskaičiuoti kainos grąžą (angl. *price return*, PR). Remiantis NasdaqBaltic (2016b) akcijos kainos grąža yra apskaičiuojama pagal (2.1.1) ir (2.1.2) formules:

$$PRI_t = \frac{PR\ Index\ Market\ Value_t}{PR\ Index\ Divisor_t} \quad (2.1.1)$$

čia PRI_t – kainos gražos indeksas laikotarpiu t ;

$$PR\ Index\ Market\ Value_t = \sum_{i=1}^n q_{i,t} * p_{i,t} * r_{i,t} \quad (2.1.2)$$

čia $q_{i,t}$ – i akcijų kiekis, įtrauktas į indeksą laiko momentu t ;

$p_{i,t}$ – kaina kotiruotės valiuta akcijos i laiko momentu t ;

$r_{i,t}$ – užsienio valiutos keitimo kursas indekso akcijos i kotiruotės valiutai konvertuoti į indekso valiutą laiko momentu t ;

Apskaičiavus kainos grąžą yra apskaičiuojama bendra rinkos grąža, kuri ir yra naudojama „OMX Vilnius“ gražos indekso apskaičiavimui. GTR apskaičiuojama pagal (2.1.3), (2.1.4) ir (2.1.5) formules:

$$GTRI_t = GTRI_{t-1} \frac{PRI_t + IDP_t}{PRI_{t-1}} \quad (2.1.3)$$

čia $GTRI_t$ – bendros gražos indekso vertė laikotarpiu t ;

$GTRI_{t-1}$ – bendros gražos indekso vertė laikotarpiu $t-1$;

PRI_t – kainos gražos indekso vertė laikotarpiu t ;

PRI_{t-1} – kainos gražos indekso vertė laikotarpiu $t-1$;

$$IDP_t (\text{Index Dividend Points}) = \frac{\text{Index Dividend Market Value}_t}{\text{Index Divisor}_t} \quad (2.1.4)$$

$$\text{Index Dividend Market Value}_t = \sum_{i=1}^n q_{i,t} * d_{i,t} * r_{i,t-1} \quad (2.1.5)$$

čia $q_{i,t}$ – i akcijų kiekis, įtrauktas į indeksą laiko momentu t ;

$d_{i,t}$ – paprastieji indekso dividendai laiko momentu t ;

$r_{i,t-1}$ – užsienio valiutos keitimo kursas indekso akcijos i kotiruotės valiutai konvertuoti į indekso valiutą laiko momentu $t-1$;

Atsižvelgiant į indekso daliklį (angl. *Index Divisor*), jis kiekvienam indeksui nėra viešai skelbiamas. Kadangi „Nasdaq“ yra privatūs indeksų tiekėjai, dažnai atvejais ne viskas būna atskleista. Taip ir šiuo atveju, tikslus indekso daliklis nėra skelbiamas. Tačiau Nasdaq (2016a) yra aprašoma, kam yra skirtas šis daliklis ir pagal kokius principus jis yra skaičiuojamas. Pagrindinė jo esmė yra tinkamai reprezentuoti akcijų indeksus. Kadangi šie susidaro iš kelių įmonių akcijų, kurios nuo savo rinkos kapitalizacijos, nuosavybės teisės ar veiklos sričių gali smarkiai kisti, buvo sukurtas šis daliklis, kuris tinkamai įvertintų, jeigu vienos įmonės akcijos indekso skaičiavime būtų pakeistos kitomis. Pavyzdžiui, „S&P 500“ susideda iš 500 didžiausių pagal rinkos kapitalizaciją prekiaujamų akcijų Amerikos biržose (S&P Dow Jones Indices, 2024). Laikui bėgant akcijos, esančios indekso viduje, kis, kadangi atsirastų naujos įmonės, kurios beaugant pateks tarp minėtų 500 ir išstums kitas. Pasikeitus šioms įmonėms, indekse turės būti apskaičiuojamas kitoks įmonės akcijų kiekis ir kitokia akcijų vertė. Būtent dėl šios priežasties akcijų indekso vidiniai kintamieji gali nuolat keistis, todėl yra apskaičiuojamas ir pritaikomas tam tikras, viešai neskelbiamas, indekso daliklis.

Įvertinus, kas yra „OMX Vilnius“ gražos indeksas ir kaip jis yra apskaičiuojamas, kitas svarbus žingsnis yra įvertinti su šiuo rodikliu susijusius tyrimus. Kitų mokslininkų atlikti tyrimai gali ne tik suteikti vertingų įžvalgų apie šio indekso subtilybes, tačiau ir išryškinti ateityje tiriamas vietas.

Teresienė (2009) atliko tyrimą vertinant įvairių apibendrintų autoregresinių sąlyginio heteroskedastiškumo (angl. *generalized autoregressive conditional heteroskedasticity*, GARCH) modelių efektyvumą, prognozuojant Lietuvos akcijų rinką. Tam, kad įvertinti Lietuvos rinką, pagrindinis šio tyrimo objektas buvo parinktas „OMX Vilnius“ gražos indeksas. Autorė įvardija investuotojų psichologiją kaip vieną pagrindinių veiksnių, darančių įtaką kainų svyravimams. Tyrime taikomi keli GARCH modeliai, tokie kaip GARCH (1,1), EGARCH (1,1), TARARCH (1,1), IGARCH (1,1) ir PARARCH (1,1), siekiant įvertinti jų efektyvumą modeliuojant „OMX Vilnius“ gražos indekso vertę. Rezultatai rodo, kad skirtingų GARCH modelių veiksmingumas nevienodas. Atlikus tyrimą nustatyta, kad EGARCH (1,1) yra ypač veiksmingas „OMX Vilnius“ gražos indeksui. Tyrimo metu taip pat buvo identifikuotas svėro efektas (angl. *leverage effect*). Tai, anot autorės, reiškia, kad investuotojai labiau reaguoja į blogas nei į geras naujienas. Apibendrinus, Teresienės (2009) tyrimas ne tik atrado tinkamiausią GARCH modelį „OMX Vilnius“ gražos indekso prognozavimui, tačiau ir įvertino svėro efektą ir išryškino investuotojų psichologijos svarbą akcijų kainų svyravimuose.

Plinkaus (2009) tyrime nagrinėjamas „OMX Vilnius“ indekso ir įvairių makroekonominių kintamųjų ryšys, siekiant nustatyti jų tarpusavio įtaką. Straipsnyje, taikant „Granger“ priežastingumo testus, nagrinėjama 40 makroekonominių rodiklių ir „OMX Vilnius“ gražos indekso sąveika nuo 1999 m. gruodžio mėn. iki 2008 m. kovo mėn. Tyrimo rezultatai rodo, kad tam tikri makroekonominiai kintamieji gali nuspėti šio indekso judėjimus, kitais atvejais, atvirkščiai, „OMX Vilnius“ gražos indeksas gali būti naudojamas kaip kintamasis nuspėti makroekonominių kintamųjų judėjimus. Pavyzdžiui, bendroji užsienio skola, BVP defliatorius, pramonės produkcijos gamintojų kainų indeksas, remiantis „Granger“ priežastingumu, lemia „OMX Vilnius“ gražos indekso judėjimus. Tyrime taip pat yra matomas ir abipusis dviejų kintamųjų ryšis tarp tokių rodiklių kaip M1 ir M2

pinigų pasiūla. Autorius taip pat pabrėžia, kad mokslininkai ne visiškai sutaria ar makroekonominiai kintamieji daro įtaką akcijų rinkos gražai. Apibendrinus, tyrimo metu buvo atlikti „Granger“ priežastingumo testai įvertinti, ar egzistuoja ryšys arba statistinis priežastingumas tarp „OMX Vilnius“ gražos indekso ir makroekonominių kintamųjų.

Valakevičius ir Vaznelytė (2012) tyrime „Daugiapakopių investavimo strategijų kūrimas, įgyvendinimas ir vertinimas“ nagrinėja daugiapakopių investavimo strategijų taikymą Lietuvos finansų rinkoje. Tam, kad tinkamai įvertinti investavimo strategijas, kaip bazinis standartas buvo pasirinktas „OMX Vilnius“ gražos indeksas, dėl savo reprezentuojamos bendros akcijų gražos Lietuvoje. Tyrime buvo taikomos modifikuotos klasikinių investavimo modelių, tokių kaip Markowitz modelis ir kapitalo turto įkainojimo modelis (ang. capital asset pricing model, CAPM), versijos. Autoriai šiuos modelius įgyvendina daugiapakopėje investavimo strategijoje, kuri apima turto atranką, portfelio sudarymą ir grįžtamąjį patikrinimą pagal istorinius duomenis. Gauta portfelio graža palyginus su „OMX Vilnius“ indekso graža buvo net 1,625 karto aukštesnė. Panašų tyrimą atliko ir Aleksaitė (2021), kuri savo tyrimu ne tik tyrė CAPM, tačiau ir naudojo „OMX Vilnius“ gražos indeksą kaip bazinį standartą, apibūdinantį bendrą Lietuvos akcijų rinką. Remiantis šiais šaltiniais galima daryti išvadą, kad „OMX Vilnius“ gražos indeksas yra naudingas įrankis norint įvertinti Lietuvos ekonominę būklę įvairių akcijų atžvilgiu.

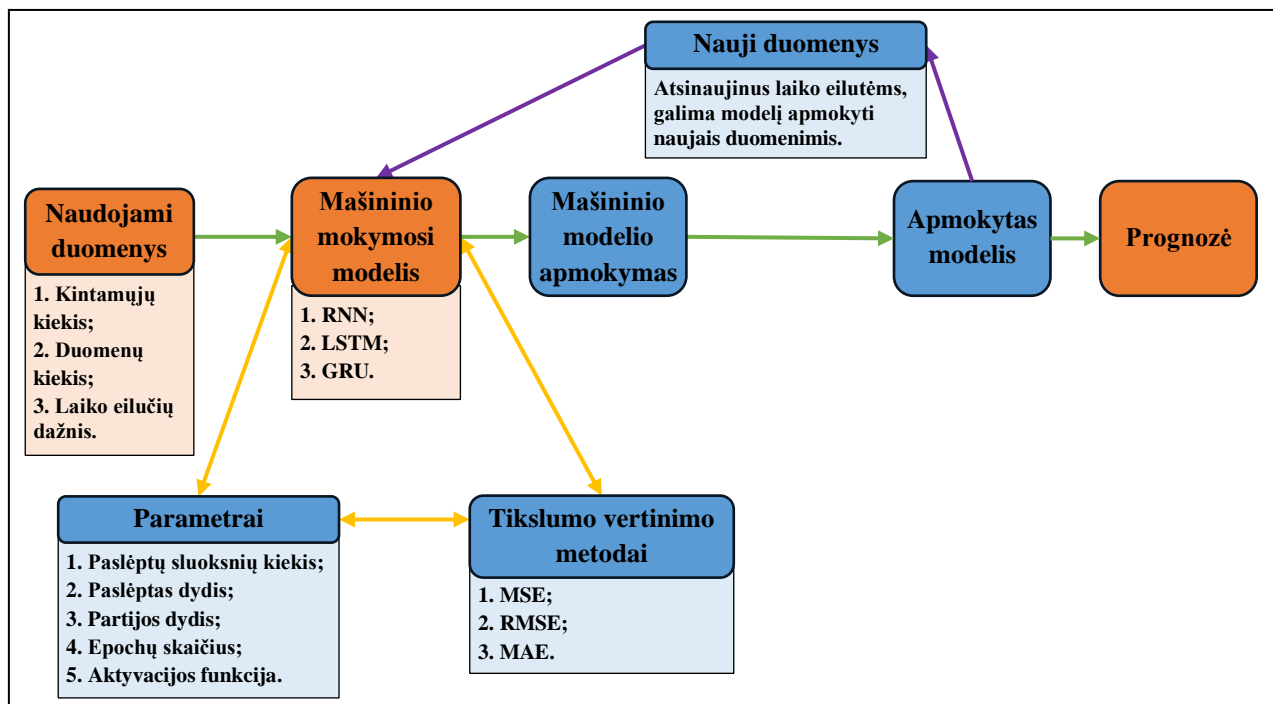
Tvaronavičienės ir Michailovos (2006) tyrime yra nagrinėjamas įvairių makroekonominių veiksnių poveikis vertybinių popierių vertėms. Tam, kad šias vertybinių popierių vertes tinkamai įvertinti, buvo naudojamas „OMX Vilnius“ gražos indeksas, kaip jas apibendrinantis vienetas. Tyrime buvo taikoma koreliacinė-regresinė analizė, siekiant kiekybiškai įvertinti „OMX Vilnius“ gražos indekso ir kelių makroekonominių rodiklių, tokių kaip tiesioginės užsienio investicijos, valstybės biudžeto rodikliai, BVP, vartotojų kainų indeksas, M2 pinigų pasiūla, vidutinis vyriausybės obligacijų pelningumas ir infliacija, sąsajas. Išvados rodo, kad makroekonominiai kintamieji daro reikšmingą įtaką „OMX Vilnius“ gražos indeksui. Visų pirma, tyrimas atskleidžia, kad tokie veiksniai kaip augantis BVP, M2 ir vyriausybės obligacijų pelningumas daro teigiamą poveikį „OMX Vilnius“ gražos indekso vertės didėjimui, kol tiesioginės užsienio investicijos, valstybės biudžeto rodikliai ir infliacija turi neigiamą poveikį. Apibendrinus, tyrimas patvirtina, kad „OMX Vilnius“ gražos indeksas yra jautrus ekonominiams pokyčiams.

Baranauskas (2010) tyrimo metu vertino Jungtinių Amerikos Valstijų ir Jungtinės Karalystės makroekonominių rodiklių įtaką „OMX Vilnius“ gražos indeksui. Tam, kad būtų įvertintas šis ryšio stiprumas, autorius pasitelkė koreliacinės analizės metodą. Tyrimo metu nustatyta, kad „OMX Vilnius“ gražos indeksas labiausiai koreliuoja su JK nedarbo lygiu, JAV parduotų prekių indeksu ir JAV naujų namų pardavimo indeksu. Atlikus tyrimą buvo padaryta išvada, kad įvairūs makroekonominiai rodikliai daro tiesioginę įtaką akcijų rinkai, įskaitant ir „OMX Vilnius“ Gražos indeksą.

2.2. Mašininio mokymosi konceptas, jo tyrimai ir nauda, prognozuojant įvairius biržos rodiklius

Mokslinėje literatūroje skiriama vis daugiau dėmesio mašininio mokymosi naudojimo paskirčių plėtimui ir efektyvumo vertinimui, įskaitant ir naudojimą ekonominiams rodikliams prognozuoti (Kumbure, Lohrman, Luukka ir Porrás, 2022). Nors tradiciniai prognozavimo metodai yra vis dar taikomi (Hyndman ir Athanasopoulos, 2018; Cerqueira, Torgo ir Mozetic 2020; Shmueli ir

Lichtendahl 2018), atsiradus naujoms technologijoms yra svarbu analizuoti naujas galimybes. Tačiau tarp Lietuvos mokslininkų mašininio mokymosi taikymas ekonominėse prognozėse yra vis dar retas, ypač vertinant Lietuvos ekonominius rodiklius. Atsižvelgiant į OMX Vilniaus gražos indekso prognozavimą pasitelkiant mašininį mokymąsi, buvo rastas tik vienas minėtas Džikevičiaus ir Stabužytės 2012 metų mokslinis straipsnis. Todėl atliekant tyrimą buvo analizuojama tarptautinė literatūra.



2 pav. Mašininio mokymosi modelio konceptuali analizė (sudaryta autoriaus, remiantis Kumbure, Lohrman, Luukka ir Porras, 2022; Cerqueira, Torgo ir Mozetic 2020; Nevasalmi 2020)

Remiantis išanalizuota literatūra buvo parengtas konceptualus mašininio mokymosi modelis (žr. 2 pav.). Bene visuose šaltiniuose yra išskirtos 2 pagrindinės mašininio mokymosi sritys, norint gauti tinkamas ir tikslias prognozes. Pirmoji sritis yra duomenys, naudojami modeliui išmokyti. Nuo pasirinktų duomenų priklauso tiek mašininio mokymosi modelis, tiek prognozės tikslumas. Pasirinkti analizuojami duomenys lemia, kiek susijusių kintamųjų galima bus išgauti, kokia bus duomenų imtis, koks laiko eilučių dažnis. Visi šie elementai yra svarbūs norint gauti patikimą prognozę. Jaya (2023) rekomenduoja surinkti kaip įmanoma didesnę duomenų kiekį, kad ši imtis būtų dar didesnė yra rekomenduojama pereiti iš dieninių verčių į valandines ar net minutes. Antroji pagrindinė sritis yra tinkamo mašininio mokymosi modelio pasirinkimas. Įprastai jis priklauso nuo duomenų, kurie yra naudojami. Kadangi bus naudojami laiko eilučių duomenys, tokio tipo mašininio mokymosi modeliai ir bus nagrinėjami. Remiantis Kumbure'e, Lohrman'u, Luukka'a ir Porras'u (2022) paaiškėjo, kad patikimiausi ir dažniausiai naudojami laiko eilučių prognozavimo modeliai yra LSTM, GRU ir SVM. Apibendrinus kitą literatūrą, tarp papildomų sričių, kurias reikia įvertinti, buvo pabrėžti modelio parametrai, tikslumo vertinimo metodai ir pakartotinis modelio apmokymas, kiekvieną kartą pasitelkiant naujus duomenis.

Prieš nagrinėjant mašininio mokymosi veikimo principus, pirmiausia yra verta pasigilinti į jau atliktus tyrimus bei jų gautus rezultatus. Kadangi planuojamas nagrinėti „OMX Vilnius“ gražos indeksas yra Lietuvos akcijų indeksas, nagrinėjami šaltiniai buvo parinkti būtent orientuojantis į akcijų biržos rodiklių prognozes. Atlikus šaltinių analizę buvo išskirti šie penki moksliniai straipsniai:

1. Li, Kamaruddin'as, Yuhaniz'as ir Al-Jaifi'as (2024) tyrimo metu, pasitelkiant simbolinį genetinį programavimą (angl. *symbolic genetic programming*, SGP) ir integruojant jį į giliuosius neuronų tinklus, bandė pagerinti mašininio mokymosi grindžiamus modelius prognozuojant Kinijoje prekiaujamų akcijų grąžą. Tyrimas parodė, kad prognozavimas gerokai pagerėjo, o rezultatai pralenkė populiariausių Kinijos indeksų grąžą. Hibridinis modelis tarp SGP ir LSTM taip pat pademonstravo puikius rezultatus, nuolat pranokdamas įprastą CSI 300 ir 500 indeksų grąžą. Šaltinis taip pat pabrėžia, kad šio tyrimo rezultatai taip pat išryškina galimybę tokio tipo analizes panaudoti ir valdant investicinius portfelius. Galiausiai yra aptariama, jog modelis naudojo tik savaitinius ir didesnio intervalo laiko eilučių duomenis, tačiau šiomis dienomis dėl technologijų pažangos galima naudoti ir kur kas smulkesnį duomenų intervalą, pavyzdžiui dieninius ar net valandinius. Tyrime yra rekomenduojama analizuoti ne tik istorinius rodiklių duomenis, tačiau ir sujungti šiuos modelius su medija, naujienomis ar net naujuoju generatyviu priešpriešiniu tinklu (angl. *generative adversarial network* – GAN), pagal kurį buvo sukurtas ChatGPT.
2. Atsižvelgiant į giliojo mašininio mokymosi tobulėjimą pastaraisiais metais, Dautel'is, Härdle'as, Lessmann'as ir Seow'as 2020 metų moksliniame straipsnyje nagrinėja šios technologijos galimybes valiutų kursų judėjime. Tyrimo metu buvo lyginami LSTM ir GRU modelių prognozavimo tikslumai su tradiciniais RNN ir grįžtamojo neuroninio tinklo (angl. *feedforward networks*, FNN) modeliais. Tyrimas prognozavo keturias valiutų poras EUR/USD, GBP/USD, USD/JPY, USD/CHF, kurių tikslumas (tai yra vienas iš klasifikavimo modelių vertinimo rodiklių, kuris rodo dalį teisingai atliktų prognozių, angl. *Accuracy*, ACC) svyravo nuo 0,5062 (FNN) iki 0,5107 (GRU), o AUC kriterijus - nuo 0,5061 (FNN) iki 0,5088 (LSTM). Tyrimo rezultatai įrodo, jog GRU ir LSTM modeliai yra pranašesni, tačiau rezultatai nėra ypač įspūdingi. Dėl šių giliojo mašininio mokymosi modelių kompleksškumo ir sudėtingumo koreguoti architektūrą, autoriai teigia, kad kai kuriais atvejais naudojant paprastesnius modelius, ypač prognozuojant pelnus, galima būtų gauti netgi tikslesnius rezultatus. Autoriai tai pat pabrėžia, kad šie testuojami modeliai dar daug kur gali būti tobulinami. Kitokia modelio struktūra, parametų koregavimas, pelno transformavimas į binarinius duomenis, naujesnių ir galingesnių mašininio mokymosi modelių panaudojimas taip pat gali duoti geresnius rezultatus. Apibendrinus, nors gauti rezultatai nėra ypač įspūdingi, autoriai pabrėžia, jog yra dar daug tobulintinų vietų, o atsižvelgiant į modelių tikslumus galima teigti, kad LSTM ir GRU modeliai pasiekia geresnių rezultatų nei įprasti RNN ir FNN modeliai.
3. M. Hansson'as 2017 metais parašė disertaciją apie akcijų grąžos prognozavimą pasitelkiant LSTM modelius. Savo darbe autorius nagrinėjo „S&P 500“, „Bovespa 50“ ir „OMX 30“ indeksus, pasitelkiant LSTM „Softmax“ modelius, kurių rezultatus vėliau lygino su ARMA-GJRGARCH prognozavimo rezultatais. Atsižvelgiant į gautus tyrimo rezultatus autorius teigia, kad vis dėlto nebuvo rastas vienas geras būdas akcijų biržos rezultatams prognozuoti, pasitelkiant LSTM modelį. Atlikto tyrimo metu nepavyko gauti patikimų „S&P 500“ ir „Bovespa 50“ prognozių, tačiau įdomu tai, jog 3 iš 4 nagrinėjamų modelių vis dėlto pasiekė ganėtinai neblogus rezultatus prognozuojant Švedijos OMX 30 indeksą. Nors dėl didelio parametų kiekio tikslios priežasties šiems rezultatams pasakyti negalima, manoma, jog modelio mokymosi ir testavimo duomenys buvo pranašesni, todėl modeliui pavyko tinkamai įsisavinti tam tikras tendencijas. Tyrimo metu taip pat buvo padaryta išvada, kad mašininio

mokymosi modeliai vis dėlto gali konkuruoti ar net pasiekti geresnius rezultatus nei tradiciniai prognozavimo modeliai, tokie kaip ARMA-GJRARCH. Rezultatai taip pat buvo teigiami prognozuojant ne tik grąžą, tačiau ir rinkos judėjimo kryptį. M. Hansson'o (2017) disertacija įrodo, kad LSTM modeliai gali būti naudingi prognozuojant akcijų biržos duomenis, tačiau rezultatai priklauso ne tik nuo naudojamų modelių, bet ir nuo prognozuojamų duomenų.

4. Atsižvelgiant į galimus uždirbti pelnus iš akcijų biržos pasitelkiant mašininį mokymąsi, Nevasalmi'is (2020) išsamiai išnagrinėjo šią sritį. Akcijų pelningumas iš ekonominės pusės yra vertinamas per visą laikotarpį, ši pozicija angliškai dar vadinama *buy-and-hold* (pirkti ir laikyti). Šiuo būdu yra lengva apskaičiuoti bendrą akcijos pelningumą per ilgalaikį laikotarpį, tačiau šis pelno apskaičiavimas yra naudingas tik bendram vaizdui arba investuotojams, kurie sutinka visą laikotarpį nekeisti savo pozicijų. Tačiau, kaip žinoma, akcijos kartais juda neracionaliai, o šiuose judėjimuose yra nemažai vietos uždirbti papildomo pelno. Būtent tai Nevasalmi'is (2020) tyrimo metu ir nagrinėjo pasitelkdamas daugianarį prognozavimo modelį. Užuoat modelį apmokus nuspėti akcijos būsimą vertę, modelis buvo apmokytas kiekvienu laiko momentu priimti vieną iš trijų pozicijų – pirkti, parduoti ar laikyti turimą poziciją. Šiuo būdu modelis sugebėjo uždirbti pinigų ne tik iš bendro akcijos vertės didėjimo, tačiau ir tarp įvairių vertės svyravimų. Atsižvelgiant į rezultatus, tarp 2007 ir 2019 metų „S&P 500“ pirkimo ir laikymo pozicija iš 100 USD išaugo į 195,86 USD poziciją, taigi galima teigti, kad per šį periodą bendra „S&P 500“ vertė išaugo beveik dvigubai, todėl šios investicijos pelnas buvo 95,86\$. Tačiau pasitelkus mašininio mokymosi daugianarį prognozavimo būdą pelnas buvo dar didesnis, k artimiausių kaimynų algoritmo (angl. *k nearest neighbors algorithm*, k-NN) modelis tyrimo metu pasirodė geriau, 100 USD vertės investiciją pavertęs į 202,95 USD, kol GBM modelis pozicijos vertę sugebėjo pakelti net iki 351,54 \$. Apibendrinus šio tyrimo gautus rezultatus galima teigti, jog mašininio mokymosi grindžiami prognozavimo būdai sugeba aplenksti ilgalaikę akcijų biržos grąžą. Taip pat atsižvelgiant į pačius mašininio mokymosi modelius, GBM modelis šiame tyrime dominavo, kadangi tarp antrą (dirbtinių neuronų tinklų (angl. *artificial neural network*, ANN) ir pirmą (GBM) vietą užėmusių prognozavimo modelio yra net 107 USD skirtumas.
5. Ferreira ir Medeiros'as (2021) taip pat atliko rinkos prognozavimą, pasitelkiant mašininį mokymąsi. Tik šiuo atveju tyrimas išsiskyrė tuo, kad autoriai naudojo minutinius „S&P 500“ ir svyravimo indekso (angl. *volatility index*, VIX) duomenis nuo 2005 m. sausio mėn. iki 2016 m. gruodžio mėn.. Tyrime buvo nagrinėjamas LSTM ir RF modelių tikslumas prognozuojant rinkos grąžą. Tam, kad būtų įvertintas modelių patikimumas, rezultatai taip pat buvo lyginami su mažiausiųjų kvadratų regresija. Prognozės buvo atliekamos taikant VIX, vėlintą 5 minučių „S&P 500“ grąžos indeksą ir agreguotą svyravimo indekso, svyravimo pokyčio ir dispersijos rizikos premijos (angl. *variance risk premium*, VRP) duomenis. Sulyginus gautus rezultatus buvo pastebėta, jog LSTM modelis su visais duomenimis sugebėjo išgauti tiksliausias prognozes, aplenkiant ir tradicinius regresijos modelius, o atsitiktinio miško modelio rezultatai buvo prasčiausi. Apibendrinus, Ferreira'os ir Mederos'o (2021) tyrimas parodo, kad LSTM modeliai yra pranašesni prognozavime, ypač prognozuojant VIX rodiklį.
6. Xiao, Feng'as, Yan'as ir Ma (2022) atliko tyrimą, kurio metu palygino ARIMA ir LSTM modelio rezultatus prognozuojant dienes „Google“, „Apple“, „Amazon“ ir „Netflix“ vertes. Kaip autoriai rašo, jų darbo tikslas yra išanalizuoti skirtumus ir panašumus tarp klasikinių,

kaip ARIMA, ir gilaus mašininio mokymosi, kaip LSTM, modelių. Autoriai tyrimo metu naudojo aukščiausias dienos kainas (angl. *high*), nuo 2010 iki 2018 metų, kurias gavo iš „Yahoo finance“ duomenų bazės. Duomenys buvo padalinti į mokymosi, validacijos ir testavimo grupes. Treniravimui buvo paskirti 2010/01/01 – 2015/12/31 duomenys, 2016/01/01 – 2017/12/31 validavimui, o 2018/01/01 – 2018/12/31 testavimui. Tam, kad rezultatai tarp ARIMA ir LSTM modelių būtų tinkamai įvertinti, autoriai tokiu pat būdu sugrupavo duomenis ir ARIMA modelių. Taip pat verta paminėti, kad geriausia ARIMA modeliui rasti autoriai pasitelkė AIC ir BIC kriterijus. Abejems modeliams buvo naudojami 60 buvusių istorinių duomenų (angl. *lookback*), tam, kad prognozuoti 1 periodą į priekį. Ruošiant LSTM modelį autoriai taip pat normalizavo duomenis pasitelkiant minimalaus-maksimalaus mastelio keitimo metodą (angl. *min-max scaler*), kuris suvienodindamas kintamųjų mastelį palengvina mašininio mokymosi procesą. LSTM modelio treniravimosi metu Xiao, Feng‘as, Yan‘as ir Ma (2022) taip pat eksperimentavo su įvairiais parametrais, buvo koreguojami atsitiktinio praretinimo transformacijos (angl. *dropout*), paslėptų sluoksnių kiekio (tai tarpiniai sluoksniai tarp įvesties ir išvesties, kurie apdoroja duomenis iš ankstesnio sluoksnio ir perduoda juos kitam sluoksniui, angl. *hidden layers*) ir paslėpto dydžio (tai neuronų kiekis kiekviename paslėptame sluoksnyje, angl. *hidden size*) parametrai, o rezultatai buvo vertinami pagal netekties funkciją (angl. *loss function*), kuri gali būti apskaičiuota pagal (2.2.1) formulę:

$$E = \sum_k t_k \log y_k \quad (2.2.1)$$

čia y – prognozuojami testuojamų duomenų rezultatai;

t – teisingi testuojamų duomenų rezultatai;

k – duomenų dimensija;

Galiausiai, atlikus visus testus ir atradus tinkamiausius LSTM modelio parametrus, autoriai apsiskaičiavo gautų rezultatų RMSE rodiklį, kurį palygino tarp įvairių akcijų ARIMA ir LSTM modelių (žr. 2 lentelę).

2 lentelė. Įvairių akcijų prognozavimo modelių RMSE tikslumas (Xiao, Feng, Yan ir Ma, 2022)

RMSE pagal akcijas ir prognozavimo modelį	Google	Apple	Amazon	Netflix
ARIMA	3.732	4.284	4.178	3.368
LSTM	2.766	3.124	3.124	2.135

Kaip matyti iš gautų rezultatų (žr. 2 lentelę), Xiao, Feng‘o, Yan‘o ir Ma (2022) tyrimas rodo, kad prognozuojant šias akcijas, visais atvejais LSTM modelis savo gautais rezultatais aplenkia klasikinį ARIMA prognozavimo modelį. Todėl remiantis šiuo šaltiniu galima daryti prielaidą, kad LSTM modelis ganėtinai neblogai veikia prognozuojant akcijas. Dėl šios priežasties tyrimo metu būtų verta įvertinti šio modelio rezultatus ir prognozuojant „OMX Vilnius“ gražos indeksą.

- Madziwa, Pillalamarry‘ė ir Chatterjee (2022) savo tyrime prognozavimui pasitelkė šiais laikais labiau klasikinį - ARDL modelį. Tyrimo metu buvo siekiama tinkamai suprognozuoti aukso kainą, remiantis aukso paklausa, izdo vekselių palūkanų norma ir buvusių laikotarpių

aukso kainomis. Tam buvo įvertinti stochastinis, ARDL ir autoregresinis slenkamasis vidurkis (angl. *autoregressive moving average*, ARMA) modeliai. Iš tyrimo gautų rezultatų, remiantis MAD, MAPE ir RMSE, paaiškėjo, kad ARDL modelis savo rezultatais aplenkė tiek ARMA, tiek stochastinio modelio rezultatus. Modeliai taip pat tenkino visas teisingo ARDL modelio hipotezes, tai įrodo, jog modelis buvo parinktas tinkamas ir yra patikimas. Autorių teigimu, būtent šio tyrimo atveju daugiamatis modelis pasirodė geriau nei vienmatis, o gauti rezultatai yra svarbūs aukso kasimo ir išgavimo rinkai. Apibendrinus, remiantis šiuo atliktu tyrimu galima daryti išvadas, kad rinkose prekiaujamus įrankius yra įmanoma prognozuoti reliatyviai tiksliai, o daugiamatiai modeliai, bent jau prognozuojant aukso vertę, yra tikslesni nei vienmatiai.

2.3. Tradiciniai prognozavimo metodai

Tam, kad tinkamai įvertinti mašininų modelių prognozavimo tikslumą, naudinga šias prognozes palyginti su labiau klasikiniais metodais. Todėl remiantis prielaida, jog bus nagrinėjamos vienmatės ir daugiamatės laiko eilutės, remiantis nagrinėtais literatūros šaltiniais buvo pasirinkta įvertinti ARIMA ir ARDL modelius.

Autoregresinio integruoto slenkamo vidurkio modelis (ARIMA) ir eksponentinio išlyginimo modeliai yra vieni daugiausiai naudojamų modelių laiko eilučių prognozavimui. ARIMA modelio pavadinimas yra autoregresinio integruoto slenkamo vidurkio akronimas. Kaip galima suprasti iš pavadinimo, modelis susideda iš 3 dalių (Hyndman ir Athanasopoulos, 2018):

1. **Autoregresinio modelio (AR(p))** – tai statistinės analizės modelis, pagal kurį prognozuojamos būsimos reikšmės, remiantis to paties kintamojo ankstesnėmis reikšmėmis. Šis autoregresijos modelis svarbus statistikoje, nes padeda prognozuoti ateities tendencijas pagal istorinius duomenis, tai yra ypač aktualu atliekant laiko eilučių analizę ekonomikos ar finansų srityse. Narių skaičius p modelio AR dalyje nurodo, kiek kintamojo vėlinimų reikia įtraukti į lygtį. Remiantis Hyndman'u ir Athanasopoulos'u (2018), šis modelis gali būti apskaičiuojamas pagal (2.3.1) formulę:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.3.1)$$

čia c yra konstanta;

ϕ – besikeičiantys parametrai;

y_t – vėlintos priklausomo kintamojo reikšmės;

ε_t – paklaida.

2. **Integruotumo (I(d)).** Tam, kad ARIMA modelis veiktų tinkamai, duomenys turi būti stacionarūs. Stacionari laiko eilutė yra tokia, kurios savybės nepriklauso nuo laiko. Taigi, laiko eilutės su tendencijomis arba sezoniškumu nėra stacionarios, todėl tendencija ir sezoniškumas turės įtakos laiko eilutės vertei skirtingu laiku. Jeigu duomenys nėra stacionarūs, juos reikėtų pasidaryti stacionarius, vienas iš šių būdų yra apskaičiuoti skirtumus tarp vienas po kito einančių verčių. Šis procesas dar yra vadinamas diferencijavimu ir yra apskaičiuojamas pagal (2.3.2) formulę (Hyndman ir Athanasopoulos, 2018):

$$y'_t = y_t - y_{t-1} \quad (2.3.2)$$

čia y_t yra kintamasis atitinkamu laiko momentu.

ARIMA modelyje integravimo eilė d (I) nurodo, kiek kartų duomenys buvo diferencijuoti. Jei $d=0$, duomenys nebuvo diferencijuojami. Jei $d=1$, duomenys diferencijuoti vieną kartą ir t. t. Tai labai svarbu, nes nestacionarūs duomenys gali lemti nepatikimus ir netikslius modelius.

3. **Slenkančio vidurkio (MA(q)).** Vietoje to, kad regresijoje būtų naudojamos ankstesnės prognozuojamo kintamojo reikšmės, slenkančio vidurkio modelyje naudojamos ankstesnės prognozės paklaidos. Kiekviena praeities prognozės paklaida dauginama iš koeficiento ir pridama prie dabartinės vertės prognozės. Narių skaičius q modelio MA dalyje nurodo, kiek praeities prognozės klaidų reikia įtraukti. Šis metodas padeda išlyginti atsitiktinius laiko eilutės duomenų svyravimus, dar kitaip vadinamus „triukšmu“ (angl. *white noise*). Slankusis vidurkis yra apskaičiuojamas pagal (2.3.3) formulę (Hyndman ir Athanasopoulos, 2018):

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.3.3)$$

čia c yra konstanta;

ε_t – paklaida;

θ – besikeičiantys parametrai.

Sudėjus šiuos procesus yra gaunamas bendras ARIMA modelis, kuris yra apskaičiuojamas pagal (2.3.4) formulę:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.3.4)$$

čia c yra konstanta;

ϕ – besikeičiantys parametrai;

y_t – vėlintos priklausomo kintamojo reikšmės;

θ – besikeičiantys parametrai;

ε_t – paklaida.

Sezoninis ARIMA modelis, dažnai žymimas ARIMA(p,d,q)(P,D,Q)_m, išplečia ARIMA modelį, kad būtų atsižvelgta ir į sezoniskumą. Tai yra ypač aktualu, kai yra nagrinėjamas ilgas periodas duomenų su pasikartojančiais sezoniniais judėjimais. Sezoninė modelio dalis žymima didžiosiomis raidėmis P, D ir Q (Hyndman ir Athanasopoulos, 2018):

- P - sezoninės autoregresijos (SAR) narių, kurie atitinka eilutės sezoninės dalies autoregresijos komponentę, skaičius.
- D - sezoninio diferencijavimo eiliškumas, kuriuo pašalinamos sezoninės tendencijos, siekiant stabilizuoti sezoninį slankųjį vidurkį.
- P - sezoninio slankiojo vidurkio (SMA) narių skaičius, analogiškas MA komponentei, tačiau taikomas eilučių sezoniniams skirtumams.

Sezoninis ARIMA modelis gali būti apskaičiuotas pagal (2.3.5) ir (2.3.6) formules (Hyndman ir Athanasopoulos, 2018):

$$(1 - \phi_1 B)(1 - \Phi_1 B^m)(1 - B)(1 - B^m)y_t = (1 - \theta_1 B)(1 - \Theta_1 B^m)\varepsilon_t \quad (2.3.5)$$

čia ϕ , Φ , θ ir Θ yra konstanta;

$$B - B^m y_t = y_{t-m} \quad (2.3.6)$$

m – pasikartojančio sezoniškumo trukmė;

ε_t – paklaida.

Kaip galima matyti iš pateiktų formuliu, tinkamo ARIMA modelio radimo procesas gali būti ganėtinai sudėtingas. Tačiau kaip Hyndman'as ir Athanasopoulos'as (2018) savo knygoje aprašo, proceso palengvinimui ir automatizavimui galima pasitelkti „auto.arima“ įrankį. Šis įrankis, tinkamai nustačius, išbando visas įmanomas ARIMA modelio variacijas ir remiantis mažiausiu AIC parenka tinkamiausius modelio parametrus. Šis procesas sumažina rankinį darbą ir pagreitina geriausio modelio radimo procesą. Pasitelkus šį įrankį ir gavus tinkamiausią modelį belieka tik patikrinti, ar gautos modelio paklaidos yra tik nereikšmingas „triukšmas“. Tai galima padaryti patikrinant 4 hipotezes:

1. **Paklaidos yra nekoreliuotos.** Jei tarp paklaidų yra koreliacijos, tuomet jose lieka informacijos, kuri turėtų būti naudojama prognozėms apskaičiuoti;
2. **Paklaidų** vidurkis lygus nuliui. Jei liekanų vidurkis yra kitoks nei nulis, prognozės yra neobjektyvios;
3. **Paklaidos** turi pastovią dispersiją;
4. **Paklaidos** pasiskirstę pagal normalųjį skirstinį.

Jeigu ARIMA modelis yra atliktas pagal aprašytas taisykles ir tenkina visas aprašytas hipotezes, modelis yra tinkamas naudoti. Apibendrinant, ARIMA modeliai yra dažnai naudojami dėl savo galimybės tinkamai įvertinti istorinius kintamojo laiko eilučių duomenis, tai leidžia ganėtinai tiksliai prognozuoti būsimas vertes. O remiantis anksčiau aprašyta literatūra, ARIMA modelių rezultatai statistinėse laiko eilučių prognozėse taip pat pateikia ganėtinai tikslius rezultatus.

Autoregresinis paskirstyto vėlavimo modelis (ARDL), remiantis EViews (2017), tai ekonometrikoje taikomas statistinis metodas, skirtas analizuoti ryšiams tarp laiko eilučių duomenų. Šis modelis apima tiek tiesioginį, tiek vėluojantį vieno ar daugiau kintamųjų poveikį priklausomam kintamajam. Autoregresinis (AR) modelis jau buvo paaiškintas aptariant ARIMA, o atsižvelgiant į vėlavimo (DL) modelį, tai yra funkcija, kuri į skaičiavimą įtraukia pasirinktus kintamųjų vėlinimus. Remiantis Karpuškiene ir Lastausku (2012), ARDL modelis yra apskaičiuojamas pagal (2.3.7) formulę.

$$Y_t = \beta_0 + \sum_i \beta_i X_{t-1} + \sum_j \theta_j Y_{t-j} + u_t \quad (2.3.7)$$

čia β_0 yra konstanta;

$\beta_i X_{t-1}$ – yra paskirstytas vėlavimas egzogeninio kintamojo X_t laiko eilutėje.

$\theta_i Y_{t-1}$ – yra autoregresinis Y_t endogeninis kintamasis su vėlinimu.

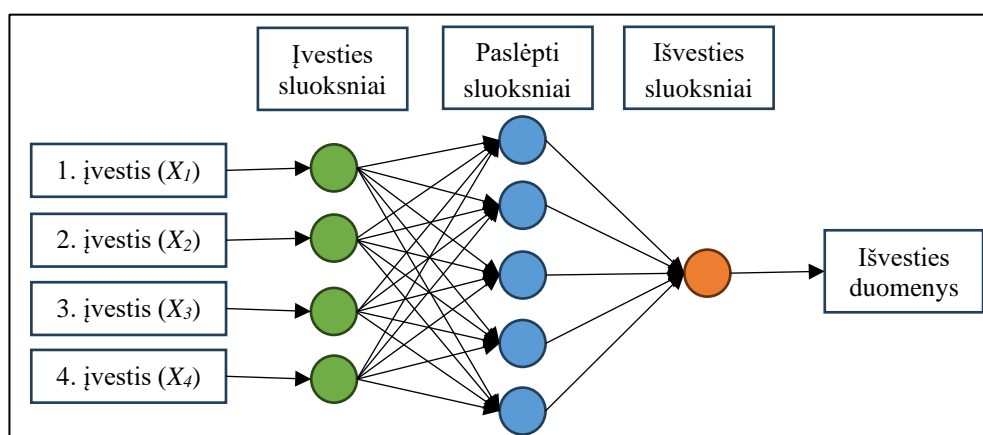
u_t – paklaida.

Nors šis modelis šiais laikais savo funkcionalumu yra ganėtinai paprastas, jo didžiausia vertė yra ta, kad jis įvertina daugiau nei vieną kintamąjį. Dėl šios priežasties jis yra vienas iš modelių, kurių pasitelkus galima prognozuoti remiantis ne tik priklausomų, tačiau ir nepriklausomų kintamųjų istorinėmis vertėmis.

2.4. Dažniausiai naudojami mašininio mokymosi modeliai

Išanalizavus klasikinius prognozavimo metodus, kitas žingsnis yra gilinti žinias į pačius mašininio mokymosi gristų prognozavimo modelių veikimo principus. Kadangi šių modelių šiais laikais yra daug, remiantis Kumbure'e, Lohrman'u, Luukka'a ir Porras'u (2022) buvo išskirti dažniausiai naudojami mokslinėje literatūroje mašininio mokymosi modeliai.

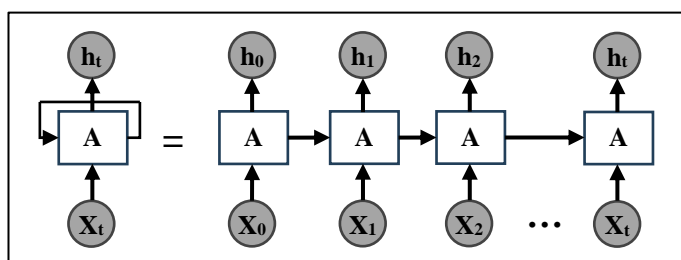
Rekurentinio neuroninio tinklo modelis (RNN). Taigi, prieš apmokant mašininio mokymosi modelį prognozuoti „OMX Vilnius“ gražos indeksą, reikėtų aptarti kas yra neuronų tinklai ir kaip jie veikia. Neuronų tinklai (angl. *neural networks*) susideda iš kelių susijusių duomenų sluoksnių, vadinamų neuronais. Neuronų tinklas gauna pradinis (įvesties) duomenis ir pasitelkiant matricos sujungimus ir nelineines transformacijas pateikia atsakymą. Visi neuronų tinklai susideda iš įvesties duomenų, išvesties duomenų ir vieno ar daugiau paslėptų sluoksnių, kurie susideda iš kelių neuronų, kurių kiekis priklauso nuo pasirinkto paslėpto dydžio. Kiekvienas sąryšis tarp neuronų ir šių paslėptų sluoksnių turi skirtingą svorį, kuris mašininio mokymosi metu yra nuolat keičiamas. Mašininio mokymosi metu iš esmės yra koreguojamas šių sąryšių svoris, atsižvelgiant į gaunamus rezultatus. Tinklo tikslas yra iš įvesties duomenų nuspėti priklausomąjį kintamąjį. Jei rezultatai yra artimi realiai reikšmei, šiam sąryšiui būna skiriamas didesnis svoris tolimesniam mokymuisi. Taip keičiantis iteracijoms, modelis išmoksta vis daugiau sąsajų tarp kintamųjų (Dautel, Härdle, Lessmann ir Seow, 2020). Remiantis Chan'u ir Mátyás'u (2022) 3 paveiksle yra pateikta vieno paslėpto sluoksnio tradicinio neuronų tinklo vizualizacija.



3 pav. Grafinė 1 paslėpto sluoksnio neuronų tinklo reprezentacija (Chan ir Mátyás, 2022)

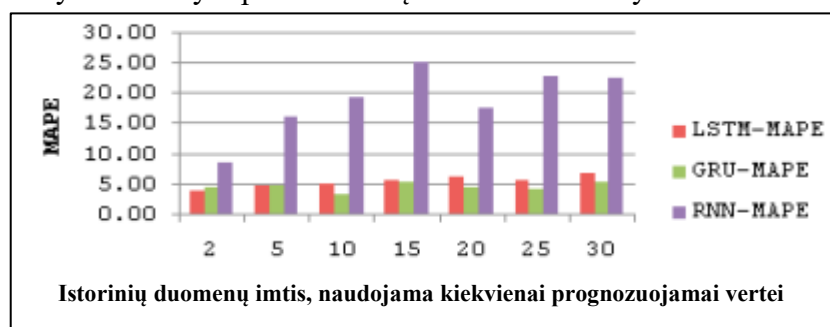
Atsižvelgiant į RNN, šie modeliai yra išskirtiniai tuo, kad jie pritaiko grįžtamąjį ryšį. RNN sugeba išlaikyti ir prisiminti tinkamiausių sąryšių svorius ir diskontuoti juos tinklo prognozėse. Žiūrint laiko atžvilgiu, RNN tinklai primena grandinę (žr. 4 pav.), kuri yra tarsi viena kitos kopija, perduodanti

informaciją savo palikuonims (angl. *decendant*). Kitais žodžiais tariant, RNN modeliai į savo skaičiavimą įtraukia buvusio laiko žingsnio informaciją (Dautel, Härdle, Lessmann ir Seow, 2020).



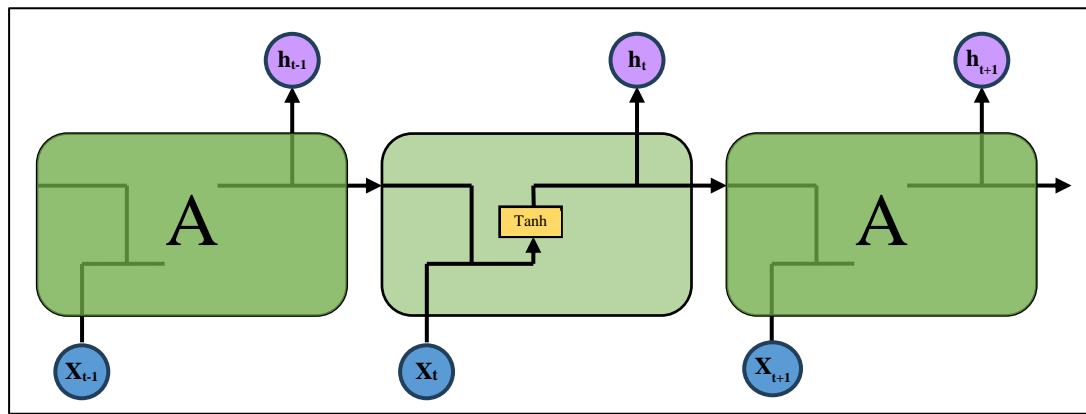
4 pav. Grafinė RNN modelio architektūra (Dautel, Härdle, Lessmann ir Seow, 2020)

Neuronų tinklas yra naudojamas optimizavimo funkcijos paklaidai mažinti. Todėl mokymosi procese, kuomet yra koreguojami paslėptų sluoksnių ir neuronų sąryšių svoriai, pokyčio nauda yra įvertinama pagal tai, ar prognozuojami rezultatai yra arčiau, ar toliau realių duomenų. Jei skirtumas tarp gautų ir realių duomenų mažėja, ši informacija yra perduodama į tolimesnį mokymosi procesą. Jei skirtumas didėja, tinklo sąryšių svoriai yra atitinkamai koreguojami. Nesigilinant į matematinius procesus, RNN modelius galima apibendrinti kaip įrankį, kuris paima įvesties duomenis ir, taikydamas atsitiktinius skirtingus svorius ir reikšmes šių duomenų sąsajoms, gauna išvesties duomenis, kuriuos vėliau lygina su realiais duomenimis. Jeigu skirtumas mokant modelį tarp realių ir prognozuojamų duomenų mažėja, tam tikri taikyti svoriai yra perduodami į tolimesnius mokymus.



5 pav. RNN, LSTM ir GRU modelių vidutinės absoliutinės procentinės paklaidos (Saud ir Shakya, 2020)

Atsižvelgiant į praktinę pusę, RNN modeliai jau yra ganėtinai nusistovėję akcijų prognozavime. Daugelyje atliekamų tyrimų šis modelis yra naudojamas ne tam, kad būtų gauti patys tiksliausi rezultatai, tačiau kaip pradžios taškas, su kuriuo kitų modelių rezultatai galėtų būti lyginami. Tai yra ypač naudinga todėl, kad vėliau aptariamieji modeliai bus šio RNN modelio patobulinta atšaka. Nors šio tipo modelis lyginant su kitais yra mažiau tikslus, turint tinkamus duomenis jis gali vis tiek sugeneruoti ganėtinai tikslias prognozes. Remiantis Saud'e ir Shakya (2020), tyrimo metu buvo palygintas RNN, MAPE tikslumas su LSTM ir GRU prognozavimo metodais (žr. 5 pav.) ir, kaip galima matyti remiantis šiuo šaltiniu, RNN modelis turėjo prasčiausius rezultatus. Todėl tyrimo metu taip pat galima tikėtis, jog RNN modelis turėtų pateikti palyginti prasčiausius prognozės rezultatus.



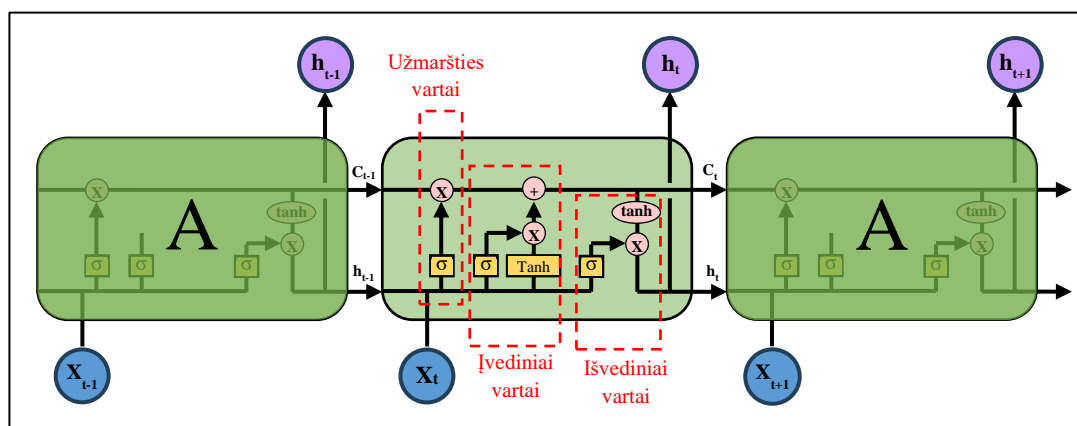
6 pav. Grafinė RNN modelio ląstelės struktūra (Olah, 2015)

Nors šis mašininio mokymosi modelis gali atrodyti naudingas, vis dėlto jis turi tam tikrų trūkumų. Kaip matyti iš 6 paveikslo, RNN modelio ląstelės struktūra yra ganėtinai paprasta. Šio paprastumo pagrindinė problema yra ta, kad keičiantis iteracijoms reikšmės praranda savo svorį. Taip atsitinka todėl, kad RNN naudoja tokias aktyvacijos funkcijas kaip hiperbolinį tangentą (angl. *hyperbolic tangent*) ar logistinę sigmoidę (angl. *logistic sigmoid*), kurie rezultatus talpina į 0 ir 1 intervalą. Tai reiškia, jog modeliui treniruojantis, tam tikrų reikšmių svoris palaipsniui mažėja. Kuo daugiau iteracijų, tuo mažesni svorį kiekviena iteracija gali užimti apskaičiuotose reikšmėse. Apibendrinus, RNN modeliams sunkiai sekasi prognozuoti ilgalaikes priklausomybes, nes svoriai pirmuose etapuose laiku bėgant tampa per maži, kad juos galima būtų tinkamai įvertinti, todėl šios vertės nuolat juda link nulio arba begalybės, šis įvykis dar yra vadinamas kaip nykstantis gradientas (angl. *vanishing gradient*). Dėl šios priežasties modeliui sunkiai sekasi išmokti priklausomybes tarp verčių, ypač kai yra naudojami dideli neuronų tinklai arba yra naudojimas didesnis kiekis mokymosi žingsnių (Zou ir kt., 2022).

Ilgalaikės-trumpalaikės atminties modeliai (LSTM) yra vieni iš populiariesnių mašininio mokymosi prognozavimo modelių, skirtų laiko eilučių prognozavimui (Jaya, 2023). Remiantis Ferreira‘a ir Medeiros‘u (2021) šis modelis yra RNN modelio atšaka, su galimybe tinkamai įvertinti buvusias laiko būsenas. LSTM modelis, kaip ir apsakoma jo pavadinime, turi galimybę atpažinti tam tikras prognozavimo sekas tiek trumpalaikėje, tiek ilgalaikėje perspektyvoje. Paprasčiau tariant, jis sugeba išvelgti ne tik tam tikras sekas ir atitikimus per visą nagrinėjamą laikotarpį, tačiau ir įvertinti pačius naujausius duomenis ir tai, kas vyksta šiuo metu. Dėl šių galimybių šis prognozavimo modelis ypač išpopuliarėjo laiko eilučių prognozavime.

Aptarus RNN pagrindus ir šio modelio esamas limitacijas, verta pereiti prie šiek tiek labiau pažengusių šio modelio variacijų – LSTM. Šis modelis išsiskiria tuo, kad, kaip ir minėta, sugeba išsaugoti ir išvelgti ilgalaikes kintamųjų priklausomybes. Jį sukūrė Hochreiter‘is ir Schmidhuber‘is 1997 metais, o vėliau jis buvo tobulinimas ir daugelio kitų mokslininkų. LSTM modeliai veikia labai gerai sprendžiant įvairias problemas, dėl to šiais laikais jie yra ypač plačiai naudojami. Priešingai nei įprastas RNN modelis, šio modelio struktūra buvo sukurta su tikslu išvengti ilgalaikės priklausomybės. Kaip ir minėta anksčiau, visi RNN modeliai susideda iš grandininės struktūros, ne išimtis ir LSTM modeliai (žr. 7 pav.). Pagrindinė LSTM funkcija yra horizontalioji linija, einant per visą diagramos viršų. Ji leidžia buvusių laikotarpių informacijai lengvai persikelti į būsimas iteracijas ir dar yra vadinama ląstelės būsena (angl. *cell state*). Kadangi LSTM modeliai neturi galimybės pridėti ar išimti informacijos, ji reguliuoja naują informaciją, pasitelkiant vartus (angl. *gates*).

Pasitelkiant sigmoidę (σ) modelis reguluoja, kiek informacijos turi būti praleista, 0 – reiškia nieko nepraleisti, 1 – praleisti viską (Olah, 2015).

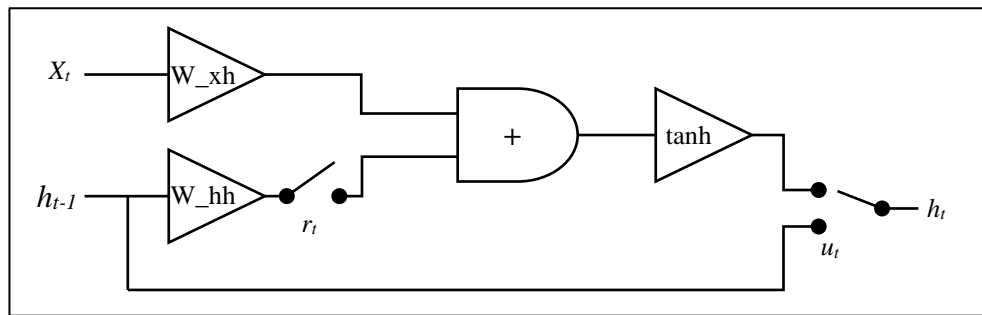


7 pav. Grafinė LSTM modelio ląstelės struktūra (sudaryta autoriaus, remiantis Olah, 2015)

LSTM modelis susideda iš trijų pagrindinių vartų: užmaršties, įvesties ir išvesties vartų. Pirmieji, užmaršties vartai, nusprendžia, kokia informacija bus nenaudojama koreguoti ląstelės būseną. Tai yra atliekama pasitelkiant jau minėtą sigmoidinę funkciją, kuri įvertina h_{t-1} ir x_t išvesdama 0 arba 1 kiekvienam ląstelės būsenos skaičiui C_{t-1} ir nuspręsdama, kurią informaciją palikti, o kurią išimti. Antrieji, įvesties vartai, nusprendžia, kokią naują informaciją įrašysime į ląstelės būseną, tai susideda iš dviejų dalių. Pirma, sigmoidinė funkcija nusprendžia, kurias reikšmes reikėtų atnaujinti. Antra, tanh sluoksnis sukuria naujų kandidatų vektorių \tilde{c}_t , kuriuos būtų galima įtraukti į ląstelės būseną. Tuomet atnaujintos reikšmės ir nauji kandidatai yra sujungiami ir pridedami prie ląstelės būsenos ją atnaujinant. Galiausiai turime nustatyti savo išvestį, kuri bus modifikuota ląstelės būsenos versija. Iš pradžių sigmoidinis sluoksnis pasirenka, kurios ląstelės būsenos dalys bus įtrauktos į išvestį. Po to ląstelės būseną apdorojama naudojant tanh funkciją (kuri pakoreguoja reikšmes taip, kad jos būtų tarp -1 ir 1) ir gauti rezultatai padauginami iš sigmoidinių išvesties vartų. Taip gaunant ir prognozuojamą vertę, ir ląstelės pradinius duomenis kitos iteracijos skaičiavimams atlikti (Olah, 2015).

Apibendrinant, ilgosios trumpalaikės atminties (LSTM) modelis yra ypač aktualus ekonomikos prognozavimo srityje dėl savo prognozavimo galimybių. Šio modelio architektūra išsprendžia gradientinio nykimo problemą, tai leidžia veiksmingai fiksuoti ilgalaikes priklausomybes laiko eilučių duomenyse. Dėl LSTM modelio gebėjimo mokytis ir įsiminti dideles istorinių duomenų sekas, jis itin gerai tinka ekonominiams rodikliams, vartotojų tendencijoms ir finansų rinkų svyravimams prognozuoti. Be to, šis modelis yra plačiai naudojamas ekonominėms prognozėms rengti dėl savo universalumo apdoroti tiek vienmačius, tiek daugiamačius duomenis su mažomis paklaidomis. Šios savybės padeda kurti sudėtingus prognozavimo modelius, kurie padeda politikos formuotojams ir įmonėms priimti pagrįstus sprendimus. Taigi, LSTM modelis yra ne tik technologinė naujovė, bet ir priemonė, didinanti ekonomikos prognozavimo tikslumą ir patikimumą.

Sklendžių rekurentinių vienetų modelis (GRU) taip pat priklauso RNN modelių kategorijai. Jis buvo sukurtas ir tobulinamas dėl jau minėtų įprasto RNN modelio trūkumų. Tokių kaip tam tikrų verčių nykimas laiko atžvilgiu. GRU modeliai, iš struktūros pusės, yra panašūs į LSTM ar RNN modelius (žr. 4 pav.). Neuronai GRU modeliuose susideda iš atminties ląstelių, kurios atlieka tam tikrą, į sklendę panašų, procesą (žr. 8 pav.). Ląstelės susideda iš dviejų pagrindinių dalių - atstatymo vartų r_t (angl. *reset gate*) ir atnaujinimo vartų u_t (angl. *update gate*) (Shen ir kt., 2018).



8 pav. Grafinė GRU modelio ląstelės struktūra (Shen ir kt., 2018)

Abu vartai reprezentuoja du elementus: x_t – yra įvesties duomenys laiko periodu, o h_{t-1} praeito laikotarpio atminties ląstelės vertė. Tam, kad duomenys būtų tinkamai atrinkti, vartams yra naudojama skirtinga filtravimo funkcija. Atstatymo vartai yra naudojami tam, kad įvertinti ar praeito laikotarpio atminties ląstelės duomenys yra svarbūs įvesties duomenims, jeigu ne, r_t gali būti atidaromi, tam, kad h_{t-1} nedarytų įtakos x_t . Atnaujinimo vartai veikia panašiu principu, tik atvirkščiai, vietoje to, kad vertinti h_{t-1} , jie vertina ar įvesties duomenys x_t turėtų būti naudojami. Jei ne, vartai yra atidaromi ir naudojami tik praeito laikotarpio ląstelės duomenys, tokiu būdu išsprendžiant gradiento nykimo problemą.

Apibendrinant galima teigti, kad GRU modelis tapo svarbia ekonominio prognozavimo priemone dėl savo veiksmingo požiūrio į laiko eilučių duomenų modeliavimą. Sukurtas kaip supaprastintas ilgosios trumpalaikės atminties (LSTM) modelio variantas, GRU supaprastina tinklo architektūrą, sumažindamas vartų skaičių, sutrumpina mokymosi laiką. Nors modelio tikslumas kai kuriais atvejais ir sumažėja, sutaupytas mokymosi laikas gali būti paskirtas geresnių modelio parametrų radimui. Geriau optimizuotas GRU modelis gali suteikti tikslesnius rezultatus nei prastai optimizuotas LSTM modelis. Taip pat kaip ir LSTM, taip ir GRU modelis yra ypač populiarus ekonominių rodiklių prognozavime, dėl savo galimybių įvertinti ilgalaikes priklausomybes, puikiai veikti tiek su vienmačiais, tiek su daugiamačiais duomenimis ir mažos prognozių paklaidos. Dėl šių priežasčių GRU modelis yra patrauklus pasirinkimas ekonomistams ir finansų analitikams.

2.5. Bendriniai mašininio mokymosi parametrai

Mašininio mokymosi modeliai yra sukurti remiantis matematiniais smegenų veikimo modeliais. Tipiškas neuroninis tinklas susidaro iš įvesties duomenų, vieno ar daugiau paslėpto sluoksnio ir išvesties duomenų. Šių tinklų efektyvumui ir tikslumui įtakos turi daug veiksnių. Duomenų savybės (kintamųjų atranka, pirminis duomenų apdorojimas, diferencijavimas), modelio architektūros parametrai (paslėptų sluoksnių skaičius, aktyvacijos funkcijų pasirinkimas), mokymosi ypatumai (optimizavimo algoritmai, partijos dydis (angl. *batch size*), epochų skaičius (angl. *epochs*) - visi šie veiksmai daro didžiulę įtaką modelio tikslumui (Ngoc, Dai ir Phuc, 2021). Dėl to verta giliau panagrinėti kiekvieną iš šių parametrų:

1. **Istorinių duomenų imtis, naudojama kiekvienai prognozuojamai vertei (angl. *lookback*):** tai yra parametras, kuriuo nustatoma, kiek iš tiesų norima naudoti ankstesnių verčių, norint prognozuoti ateinantį laikotarpį. Šis parametras įprastai naudojamas laiko eilučių prognozavime. Taip pat juo galima įtraukti ir tam tikrą su laiko eilute susijusį sezoniškumą, pvz. valandiniams duomenims galima naudoti 24, savaitiniams 7, mėnesiniams periodams 12 (Ngoc, Dai ir Phuc, 2021).

2. **Paslėptas dydis (angl. *hidden size*)** - yra neuronų skaičius naudojamas neuronų sluoksniuose. Remiantis Ngoc'u, Dai ir Phuc'u (2021), per mažas paslėptas dydis gali lemti didelę paklaidą, o per didelis dydis gali lemti perteklinį pritaikymą (angl. *overfit*) mokymosi duomenų imčiai. Taip pat gali būti naudojamas ir skirtingas sluoksnių skaičius, kuris taip pat gali daryti įtaką geriausiam paslėpto dydžio nustatymui.
3. **Paslėpti sluoksniai (angl. *hidden layers*)** – tai tarpiniai sluoksniai tarp įvesties ir išvesties, kurie apdoroja duomenis iš ankstesnio sluoksnio ir perduoda juos kitam sluoksniui. Daugiau sluoksnių lemia kompleksiškesnę modelio struktūrą. Priklausomai nuo naudojamų duomenų ir norimos išspręsti problemos, daugiau sluoksnių gali išžvelgti gilesnes ir sudėtingesnes sąsajas tarp kintamųjų, taip pagerinant prognozių tikslumą. Kitais atvejais sudėtingesnis modelis gali privesti prie perteklinio pritaikymo. (Karsoliya, 2012)
4. **Partijos dydis (angl. *batch size*)** - tai duomenų imties, apdorotų prieš atnaujinant modelio vidinius parametrus, skaičius. Pvz., jeigu laiko eilučių mokymuisi yra 1000, o partijos dydis yra 100, modelis mokymuisi iš pradžių paims tik pirmus 100 pavyzdžių. Po to imami antrieji 100 pavyzdžių, pagal kuriuos modelis mokosi iš naujo ir remiantis gautais rezultatais atnaujinama buvusius parametrus. Šis procesas kartojasi, kol yra įvertintos visos 1000 eilučių. Didesnis partijos dydžio nustatymas lemia greitesnį mokymosi procesą, mažinant reikalingus kompiuterinius pajėgumus, tačiau daugeliu atveju prarandant modelio tikslumą. Mažinant partijos dydį, modelio rezultatai gerėja, tačiau ir didėja reikalingi kompiuteriniai pajėgumai (Ngoc, Dai ir Phuc, 2021).
5. **Epochos (angl. *epochs*)** - tai yra skaičius, kuris nustato kiek kartų modelis turi praeiti visą duomenų imtį. Epochos susideda iš vieno ar daugiau partijos dydžių. Šio parametro nustatymas lemia geresnį modelio treniravimosi duomenų išmokimą, tačiau smarkiai padidina reikalingus kompiuterinius pajėgumus. Taip pat verta paminėti, kad tiek epochų didinimas, tiek partijos dydžio mažinimas gerina prognozių rezultatus tik iki perteklinio pritaikymo. Vėliau modelio rezultatai gerėja tik treniruojamiems duomenims, o testuojami duomenys bei reali prognozė dažnai atvejais būna visiškai netiksli (Ngoc, Dai ir Phuc, 2021).
6. **Optimizacijos algoritmai (angl. *optimizer algorithms*)** - tai algoritmai arba metodai, naudojami neuroninio tinklo atributams, pavyzdžiui, svoriams ir mokymosi greičiui, keisti, siekiant sumažinti modelio prognozavimo paklaidą. Tai atliekama taikant procesą, kurio tikslas - rasti svorių rinkinį, užtikrinantį geriausius prognozavimo rezultatus. Įprasti optimizavimo įrankiai neuroninių tinklų mokymui yra stochastinis gradientinis nuolydis (angl. *Stochastic Gradient Descent*, SGD), vidutinės kvadratinės reikšmės skleidimas (angl. *root mean square propagation*, RMSProp), adaptyvusis gradiento algoritmas (angl. *adaptive gradient algorithm*, Adagrad), adaptyvioji delta (angl. *adaptive delta*, Adadelta), adaptyvus momento įvertinimas (angl. *adaptive moment estimation*, Adam), adaptyvusis momentas su maksimumu (angl. *adaptive moment with maximum*, Adamax), Nesterov pagreitinintas adaptyvusis momento įvertinimas (angl. *Nesterov-accelerated adaptive moment estimation*, Nadam) (Ngoc, Dai ir Phuc, 2021).
7. **Diferencijavimas:** nors mašininio mokymosi modeliai ir gali susitvarkyti su duomenimis, kurie yra mažai apdirbti, nekeičiant mastelio ar nediferencijuojant jų dėl stacionarumo, nagrinėjant laiko eilučių duomenis, kartais vis tiek yra naudinga juos diferencijuoti.

Duomenys gali būti įprastai ir sezoniškai diferencijuoti siekiant palengvinti mašininio mokymosi procesą. To pasekoje kai kuriais atvejais yra gaunami tikslesni rezultatai. (Bouktif, Fiaz, Ouni ir Serhani, 2018)

Kaip matoma, galimų parametrų imtis yra didelė. Todėl ranka eksperimentuoti, bandant surasti geriausius rezultatus, gali būti labai varginanti ir laikui imli patirtis. Tam, kad paspartinti šiuos procesus, mašininio mokymosi tyrimuose yra naudojami keletas įrankių. Remiantis Petro‘u Liashchynskiyi‘u ir Pavlo‘u Liashchynskiyi‘u (2019), šiam procesui yra naudojami tokie įrankiai kaip tinklo paieška (angl. *grid search*), atsitiktinė paieška (angl. *random search*) ir genetinis algoritmas (angl. *genetic algorithm*).

- **Tinklo paieška** - tai tradicinis parametrų optimizavimo būdas, kuriuo atliekama išsami paieška apibrėžtame parametrų lauke. Kadangi kai kurie parametrai skaičių atžvilgiu gali būti begaliniai, naudojant šį įrankį yra svarbu nustatyti tam tikras parametro ribas. Nors yra išbandomos visos nustatytos parametrų variacijos, kartais tai nėra ypač efektyvu, kadangi reikšmės yra parenkamos nepriklausomai viena nuo kitos. Geriausi parametrai gali būti atrasti labai anksti, tačiau iki kol nesibaigs visa tinklo paieška, nebus aišku ar jie iš tiesų yra geriausi rasti parametrai (Liashchynskiyi, P. ir Liashchynskiyi, P., 2019).
- **Atsitiktinė paieška:** skirtingai nei tinklinė paieška, atsitiktinė paieška nevertina visų galimų derinių, bet juos pasirenka atsitiktinai. Atsitiktinė paieška gali greičiau gauti pakankamai gerus rezultatus nei tinklo paieška, ypač kai nedaug parametrų daro didelę įtaką algoritmo rezultatams. Tačiau parametrai visada yra parenkami atsitiktinai, dėl to niekada negalima garantuoti, jog tai yra tinkamiausi parametrai naudojamam modeliui (Liashchynskiyi, P. ir Liashchynskiyi, P. 2019).
- **Genetinis algoritmas** - tai evoliucinis metodas, naudojamas parametrų optimizavimui spręsti. Šis būdas primena biologinę evoliuciją, kai pagal gautus rezultatus yra parenkami ir derinami parametrai. Genetinis algoritmas pradamas nuo atsitiktinai sukurtos chromosomų populiacijos. Tada jis atlieka atranką ir rekombinaciją, remdamasis kiekvienos chromosomos tinkamumu (angl. *score*). Tėvų genetinę medžiagą rekombinuojama, kad būtų sukurtos naujos kartos chromosomos. Šis procesas kartojamas tol, kol pasiekiamas tam tikras sustabdymo kriterijus (Liashchynskiyi, P. ir Liashchynskiyi, P. 2019). Iš esmės yra parenkami atsitiktiniai parametrai, apmokius modelį įvertinami rezultatai ir bandoma nauja kombinacija parametrų, jei tikslumas didėja, dalis rastų parametrų paliekami tokie patys arba keičiami nedaug ir procesas kartojamas. Nors šis parametrų radimo būdas yra labiau pažengęs, remiantis Petro‘u Liashchynskiyi‘u ir Pavlo‘u Liashchynskiyi‘u (2019), laiko atžvilgiu jis yra vis tiek ganėtinai lėtas. Tyrimo metu genetinis algoritmas užtruko beveik tiek pat laiko kiek tinklo paieška.

Ngoc‘as, Dai ir Phuc‘as (2021) savo tyrime tai pat pateikė pavyzdį kaip gali atrodyti parametrų derinimo imtis (žr. 3 lentelę). Pasirinkus šią imtį, remiantis geriausiais RMSE, MAPE ir MAE rodikliais buvo atrinkti optimaliausi, geriausius rezultatus suteikiantys parametrai. Šiam parametrų radimui autoriai pasitelkė tinklo paieškos metodą. Nors šis metodas nėra optimaliausias laiko atžvilgiu, priešingai nei kiti metodai, jis garantuoja geriausius rezultatus nagrinėjamai imčiai.

3 lentelė. Tinklo paieškos parametrų derinimo imtis (Ngoc, Dai ir Phuc, 2021)

Stebiniai	Jungtinių valstijų oro linijų keleivių duomenys
Stebinių vėlinimas	[1, 12]
Paslėptas dydis	[50, 100]
EPOCHOS	[10, 100]
Partijos dydis	[1, 10]
Diferencijavimas	[0, 12]
Optimizacijos algoritmas	[SGD, RMSprop, Adagrad, Adadelta, Adam, Adamax, Nadam]
Galimas stebinių kiekis	224

Remiantis išanalizuotais šaltiniais galima teigti, kad parametrai yra labai svarbus mašininio mokymosi aspektas. Nuo jų parinkimo priklauso visas modelio tikslumas. Atsižvelgiant į tai, jog šių parametrų pasirinkimas gali būti labai didelis, laiko atžvilgiu yra naudinga pasitelkti tokius įrankius kaip tinklo, atsitiktinę ar net genetinių algoritmų paiešką. Taip pat yra svarbu nustatyti realistišką nagrinėjamų parametrų kiekį, kadangi per didelė imtis gali lemti parametrų paieškas, kurios trunka metus ar net ilgiau.

Atlikus literatūros analizę ir aptarus „OMX Vilnius“ gražos indeksą buvo pastebėta, kad tyrimų, naudojančių šį indeksą, nėra daug. Įvertinus šaltinius pastebėta, kad šis indeksas dažniais atvejais yra naudojamas kaip bendras rodiklis, vertinantis Lietuvos akcijų rinką. Nors moksliniuose straipsniuose ir buvo įvertintas šio indekso sąryšis su įvairiais makroekonominiais rodikliais, šaltinių, tiriančių šio indekso koreliacijas su įvairiu biržose prekiaujamu turtu, nebuvo rasta. Taip pat, atsižvelgiant į mašininio mokymosi pusę, buvo rastas tik vienas 2012 metų Džikevičiaus ir Stabužytės tyrimas naudojantis šį indeksą. Įvertinus įvairius prognozavimo modelius, išaiškinus jų veikimo principus ir atradus mašininio mokymosi modelių pranašumą prognozuojant kitus rinkos rodiklius, buvo atrinkti LSTM ir GRU modeliai. Įvertinus tai, kad nebuvo rasta jokių mokslinių tyrimų, naudojančių LSTM ir GRU modelius „OMX Vilnius“ gražos indekso prognozavimui, ir remiantis tuo, kad nebuvo rasta šaltinių, kurie nagrinėtų šio indekso sąryšį su kitu biržose prekiaujamu turtu, buvo pasirinkta tiriamojame dalyje į šias sritis ir orientuotis.

3. Gilaus mašininio mokymosi ir klasikinių modelių, prognozuojančių „OMX Vilnius“ gražos indeksą, metodika

Siekiant užtikrinti šio tyrimo tikslumą, teisingumą ir sąžiningumą, bus pateikta išsami ir sisteminga darbo metodika. Joje bus išsamiai aprašyti duomenų atrankos ir parengimo procesai. Tyrime bus taikomi įvairūs klasikiniai ir mašininio mokymosi modeliai, prognozuojant „OMX Vilnius“ gražos indeksą. Nors ši metodika ir bus orientuota į „OMX Vilnius“ gražos indeksą, tai nereiškia, kad pagal ją galima bus prognozuoti tik šį indeksą. Šios metodologijos tikslas yra aiškiai ir tiksliai pateikti visus svarbiausius tyrimo eigos žingsnius.

Atsižvelgiant į tyrimo eigą, planuojama darbo eiga prasideda nuo „OMX Vilnius“ gražos indekso istorinių duomenų gavimo ir jų paruošimo, užtikrinant, kad jie būtų tinkami analizei. Apdirbus duomenis, pirmiausia bus įvertinti klasikiniai prognozavimo metodai. Šiai analizei planuojama naudoti ARIMA ir ARDL modelius, kurių tinkamiausia struktūra bus atrasta remiantis „Python“ „auto.arima“ biblioteka, ir „Eviews“ ARDL radimo įrankiu. Atlikus prognozes, pasitelkiant klasikinius prognozavimo metodus bus paruošti LSTM ir GRU gilaus mašininio mokymosi modeliai, naudojantis „Python“ su „Pytorch“, „Numpy“, „Pandas“ ir „Sklearn“ bibliotekomis. Tyrimo metu taip pat bus atrinkti tiksliausi parametrai, pasitelkiant tinklo paieška tam, kad gauti tiksliausius pasiekiamus rezultatus. Galiausiai, modeliams įvertinti bus atlikta lyginamoji analizė, daugiausia dėmesio skiriant jų prognozavimo tikslumui „OMX Vilnius“ gražos indekso kontekste. Gavus rezultatus bus padarytos išvados ir sudaryta reali prognozė, remiantis geriausiu sukurtu modeliu. Taigi tam, kad tyrimas būtų skaidrus, tikslus ir atkartojamas, tyrimo metodika bus pateikta detaliau.

Pirmiausia, kadangi tyrimo metu yra norima atskleisti gilaus mašininio mokymosi modelio tikslumą lyginant su klasikiais prognozavimo metodais tiek dieniniu, tiek mėnesiniu lygiu, buvo atrinkti duomenys, kurie yra sekami kiekvieną dieną. Atsižvelgiant į pagrindinį kintamąjį, „OMX Vilnius“ gražos indekso duomenys buvo gauti iš „Nasdaq“ biržos. Kadangi šį indeksą apskaičiuoja pats „Nasdaq“, visa jo istorija nuo 2000 metų yra viešai prieinama jų tinklalapyje (Nasdaq, 2024). Atsižvelgiant į daugiamačius modelius, kaip ir minėta anksčiau, pirminis kriterijus buvo tai, jog duomenys turi būti sekami kiekvieną dieną. Dėl šios priežasties galimų kintamųjų imtis buvo ganėtinai apribota, kadangi kiekvieną dieną sekamos vertės yra tik biržose prekiaujamo turto. Kadangi šių turto įrankių pasirinkimas yra vis tiek labai didelis, tyrimui buvo pasirinkta orientotis tik į pačius didžiausius pagal rinkos kapitaliaciją prekiaujamus vienetus. Taip pat, kadangi literatūros analizė parodė, jog didesnė duomenų imtis suteikia geresnius mašininio mokymosi rezultatus, buvo atrinkti tik 15 metų ir ilgiau gyvuojantys rinkos vienetai. Tyrimui iš „Yahoo Finance“ duomenų bazės buvo atrinktos didžiausios pagal rinkos kapitalą prekiaujamos akcijos Lietuvoje, Latvijoje, Estijoje, Euro Zonoje ir JAV. Tam, kad išvelgti ne tik akcijų įtaką „OMX Vilnius“ gražos indeksui, taip pat buvo atrinktos tyrimo metu daugiausiai prekiaujamos žaliavos, valiutų kursai bei indeksai. Galiausiai, tam, kad įvertinti infliacijos įtaką „OMX Vilnius“ gražos indeksui, į duomenų imtį buvo įtrauktos ir Euro zonos obligacijų 1, 3, 5, 10 ir 30 metų pajamingumo kreivės (angl. *Euro zone yield curve*) iš „Eurostat“ duomenų bazės (žr. 4 lentelę). Daugelis šių verčių yra apskaičiuojamos tik darbo dienomis. Todėl atsižvelgiant į tai, jog prognozėms duomenys turėtų būti unifikuoti, kintamųjų vertės buvo užpildytos naujausiomis žinomomis to periodo vertėmis.

4 lentelė. Tyrimo metu naudojamų kintamųjų sąrašas pagal kategoriją.

Lietuva	Latvija	Estija	Euro Zona	JAV	Žaliavos	Valiutos	Indeksai	Euro zonos obligacijų pajamingumo kreivės
Telia Lietuva	AS Latvijas Gaze	TKM Grupp AS	Novo Nordisk	Microsoft Corporation	Auksas	USD/EUR	S&P 500	1 metų pajamingumas
AB Šiaulių Bankas	AS Amber Latvijas balzams	AS Merko Ehitus	AstraZeneca PLC	Apple Inc.	Varis	USD/GBP	DAX	3 metų pajamingumas
APB Apranga	SAF Tehnika A/S	AS Silvano Fashion Group	Equinor ASA	NVIDIA Corporation	Žalia nafta	USD/JPY	CAC40	5 metų pajamingumas
Grigeo AB	AS Latvijas Juras medicinas centrs	Nordic Fibreboard AS	Investor AB	Alphabet Inc.	Natūralios dujos	USD/CHF	-	10 metų pajamingumas
Invalda INVL	AS Siguldas Ciltslietu Un Maksligas Apsklosanas Stacija	-	AB Volvo (publ)	Amazon.com, Inc.	Kava	-	-	30 metų pajamingumas

Kadangi duomenų imtis yra skirtinga, kai kurie įrankiai yra naujesni arba tik vėliau buvo pradėti sekti „Yahoo Finance“. Dėl šios priežasties, tam, kad sulygtinti laiko imtį, duomenys buvo imti nuo 30/09/2004 iki 31/03/2024, kadangi šiuo laikotarpiu galima buvo išgauti didžiausią kiekį bendrų kintamųjų.

Žinoma, kaip ir visiems klasikiniams, taip ir mašiniams prognozavimo modeliams, svarbu yra įvertinti ir tinkamai paruošti duomenis. Nors kartais duomenų paruošimo metodika tarp modelių gali skirtis, yra tam tikri standartai, kurie tarp skirtingų modelių yra universalūs. Taigi, atrinkus duomenis prognozavimo metodams, pirmiausia reikia išsiaiškinti, ar visi turimi duomenys yra tinkami. Tam yra atliekama preliminarinė stebinių analizė. Remiantis Balaboniene, Bliekiene ir Stundžiene (2013), svarbiausi yra šie uždaviniai:

1. Pagrindinių skaitinių charakteristikų apskaičiavimas;
2. Grafinė kintamųjų analizė;
3. Duomenų unifikavimas;
4. Išskirčių nustatymas;
5. Duomenų normalumo patikrinimas;
6. Koreliacinės matricos suformavimas;

Atsižvelgiant į duomenų unifikavimą, be jau minėto laiko skalės suvienodinimo ir trūkstančių verčių užpildymo, duomenis taip pat reikėtų suvienodinti mastelio atžvilgiu. Seabe'as, Moutsinga ir Pindza (2023), siekdami sušvelninti nevienodą mastelį tarp kintamųjų, pasitelkė Min-Max metodą, kuris, remiantis autoriaus žodžiais, naujausiais tyrimais buvo pademonstruotas kaip efektyvus įrankis,

didinantis modelio efektyvumą. Remiantis Brownlee'o (2020) Min-Max metodu, apskaičiuotos vertės - tai kai kintamųjų minimumas yra lygus nuliui, o kintamųjų maksimumas lygus vienetui. Šis kintamųjų mastelio suvienodinimas gali būti apskaičiuojamas pagal (3.1) formulę:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

Įvertinus ir tinkamai paruošus duomenis, jeigu tyrimo metu yra naudojama daugiau nei vienas kintamasis, yra svarbu įvertinti ar naudojami kintamieji yra reikšmingi ir tinkami tiek klasikiniais, tiek mašininio mokymosi prognozavimo modeliams. Tam, kad atrinkti tik svarbius kintamuosius, galima atlikti koreliacijos matricos analizę tarp kintamųjų. Seabe'as, Moutsinga ir Pindza (2023) nagrinėjant koreliaciją tarp skirtingų kriptovaliutų pabrėžia, kad koeficientai virš |0.5| jau skaitomi, kad yra reikšmingai susiję ir gali būti naudojami tolimesniems tyrimams. Tačiau remiantis Balabonienė, Bliekiene ir Stundžiene (2013) (žr. 5 lentelę), silpna teigiama arba neigiama koreliacija yra jau nuo |0.3|. Tad jeigu yra sunku rasti smarkiau koreliuojančius rodiklius, retais atvejais galima naudoti kintamuosius ir nuo |0.3| koreliacijos.

5 lentelė. „Pearson“ koreliacijos reikšmių interpretacijos (Balabonienė, Bliekiene ir Stundžiene, 2013)

r reikšmė	Interpretacija
Nuo 0,9 iki 1,0 (nuo -0,9 iki -1,0)	Labai stipri teigiama (neigiama) tiesinė koreliacija
Nuo 0,7 iki 0,9 (nuo -0,7 iki -0,9)	Stipri teigiama (neigiama) tiesinė koreliacija
Nuo 0,5 iki 0,5 (nuo -0,5 iki -0,5)	Vidutinė teigiama (neigiama) tiesinė koreliacija
Nuo 0,3 iki 0,3 (nuo -0,3 iki -0,3)	Silpna teigiama (neigiama) tiesinė koreliacija
Nuo -0,3 iki 0,3	Labai silpna teigiama (neigiama) tiesinė koreliacija

Tam, kad gauti šiuos koreliacijos įverčius, galima naudoti „Pearson“ koreliacijos koeficiento (3.2) formulę (Balabonienė, Bliekiene ir Stundžiene, 2013):

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x * S_y} \quad (3.2)$$

čia x ir y – yra kintamieji, tarp kurių yra vertinama „Pearson“ koreliacija;

S_x ir S_y – yra kintamųjų standartiniai nuokrypiai.

„Granger“ priežastingumas yra statistinė sąvoka, naudojama siekiant nustatyti, ar viena laiko eilutė gali prognozuoti kitą. Ši koncepcija, pavadinta ekonomisto Clive W. J. Granger vardu, sukėlė revoliuciją laiko eilučių duomenų priežastinių ryšių analizėje. „Granger“ priežastinio ryšio apibrėžimas yra grindžiamas trimis prielaidomis (Maziarz, 2015):

1. Praeitis ir dabartis gali nulemti ateitį, bet ateitis negali nulemti praeities;
2. Visos bet kuriuo metu turimos žinios (žymimos Ω_n) turėtų apimti tik unikalią informaciją. Jei kintamasis, tarkime, Z_n , yra visiškai nuspėjamas pagal kitus kintamuosius dėl fiksuoto ryšio, tuomet Z_n nesuteikia jokios naujos informacijos ir turėtų būti neįtrauktas į Ω_n ;
3. Visi priežastiniai ryšiai išlieka pastovios krypties visą laiką.

Apibendrinus, „Granger“ teigė, kad jei tikimybė, jog tam tikra būsimos eilutės X reikšmė pasikeis įtraukus informaciją apie dabartinę eilutę Y pasikeičia, tuomet galima teigti, kad Y sukelia X . Šiam apibrėžimui reikia, kad statistiškai reikšmingai skirtingi modelių, į kuriuos įtraukta informacija apie dabartinę eilutę Y , ir modelių, į kuriuos neįtraukta informacija apie dabartinę eilutę Y , prognozavimo tikslumas. „Granger“ priešastingumas gali būti apskaičiuojamas pagal (3.3) formulę (Maziarz, 2015):

Y_n lemia Y_{n+1} jei:

$$P(X_{n+1} \in A | \Omega_n) \neq P(X_{n+1} \in A | \Omega_n - Y_n) \quad (3.3)$$

čia X_t ir Y_t – kintamieji, kurie, kaip įtariama, yra susiję priežastiniu ryšiu;

Ω_n – yra visos visatoje turimos žinios laiko momentu t .

Taigi, norint atlikti tikslingą tyrimą yra svarbu atkreipti dėmesį ir į duomenų apdirbimo procesus. Todėl prieš naudojant klasikinius ar mašininio mokymosi metodus daugiamačiams duomenims, reikia atlikti bent jau „Granger“ priešastingumo ir „Pearson“ koreliacijos analizę, o duomenis laiko ir mastelio atžvilgiu suvienodinti.

Atsižvelgiant į naudojamus prognozavimo modelius, tam, kad būtų visapusiškai įvertinta mašininio mokymosi grįstų modelių nauda, kaip ir minėta, bus vertinami dieninių ir mėnesinių, vienmačių ir daugiamačių duomenų imčių modeliai. Todėl kaip matyti (žr. 6 lentelę), tyrimo metu planuojama įvertinti 12 skirtingų prognozavimo modelių. Taip pat visi jie bus lyginami su vienu bendru standartu – naiviu prognozavimo metodu (vertinimo metodas, kai praėjusio laikotarpio faktiniai duomenys naudojami kaip šio laikotarpio prognozė, jų nekoreguojant ir nebandant nustatyti priežastinių veiksnių, angl. *naive*). Naivusis metodas iš esmės neprognozuoja, tačiau jis naudojamas kaip geras vertinimo standartas, kadangi jei modelis nesugeba aplenksti net naivaus prognozavimo modelio rezultatų, galima teigti, kad rezultatai nėra labai naudingi. Taip pat verta paminėti, kad duomenys tyrimo metu tiek klasikiniams, tiek mėnesiniams modeliams bus padalinti į 90% treniravimo (30/09/2004 - 18/04/2022) ir 10% testavimo (18/04/2022 - 31/03/2024) grupes, tam, kad gautus rezultatus būtų galima įvertinti vienodomis sąlygomis.

6 lentelė. Tyrimo metu naudojami prognozavimo modeliai

Bendras vertinimo standartas	Dieniniai duomenys		Mėnesiniai duomenys	
	Vienmačiai	Daugiamačiai	Vienmačiai	Daugiamačiai
Naive	ARIMA	ARDL	ARIMA	ARDL
	LSTM	LSTM	LSTM	LSTM
	GRU	GRU	GRU	GRU

Atsižvelgiant į modelių pritaikymą, bus naudojami keli įrankiai. Kaip ir minėta, ARIMA modeliui bus pasitelkiamas „Python“ su „auto.arima“ biblioteka, kuris remiantis geriausiu gautu AIC kriterijumi atras tinkamiausią šio ARIMA modelio struktūrą. Atsižvelgiant į nustatytus ieškomus parametrus, dieniniams duomenims vėlinimų periodas bus nustatytas nuo 0 iki 7, kol mėnesiniams nuo 0 iki 12. Iš sezoniskumo pusės, dieniniams duomenims bus nustatytas 7 periodų sezoniskumas, kol mėnesiniams – 12. Šios imtys buvo pasirinktos atsižvelgiant į logišką sezoniskumą ir turimus kompiuterinius pajėgumus.

ARDL modeliui bus naudojami keli įrankiai, pirmiausiai, pasitelkiant „Eviews“ bus įkeliami treniravimui naudojamų 90% visų duomenų ir bus atliekama geriausio ARDL modelio paieška,

pasitelkiant AIC kriterijų. Kadangi studentiška Eviews versija turi apribojimų, radus geriausią modelį tolimesnė analizė ir prognozavimas bus atliekamas pasitelkiant „Python“ su „Statsmodel“ biblioteką. Geriausias rastas modelis bus taip pat testuojamas su paskutiniais 10% duomenų, kurių rezultatai tyrimo metu bus lyginami su kitais modeliais.

LSTM ir GRU modeliai bus kuriami per „Python“ programą. Mašininiam mokymuisi bus naudojamas „Pytorch“ dėl savo pažangių modelio koregavimo galimybių. Tyrimui atlikti mašininio mokymosi kodas bus kuriamas remiantis Tam'o (2023) sukurtu LSTM modeliu, skirtu prognozuoti oro linijų keleivių skaičių. Modelis bus tik pakoreguotas veikti su daugiamačiais duomenimis ir remiantis Maity'u (2021) bus pridėtas minėtas Min-Max mastelio keitimo metodas, kad modelis geriau įsisavintų pateiktus duomenis. Taip pat, kadangi Tam'o (2023) sukurtas kodas buvo pritaikytas tik LSTM skaičiavimui, jis bus pakoreguotas, kad veiktų ir su GRU modelio architektūra.

Žinoma, kaip ir minėta literatūros analizėje, svarbi tyrimo dalis yra gilaus mašininio mokymosi modelių parametrų nustatymas, nuo kurių rezultatai gali skirtis labai smarkiai. Tam, kad gauti gerus rezultatus, svarbu parinkti kiekvienam modeliui tinkamus parametrus. Kadangi jų imtis gali būti begalinė, bus pasirinkta naudoti ganėtinai standartinį įrankį, naudojamą šiai problemai spręsti – tinklo paiešką. Šis įrankis pagal nustatytus parametrus, pvz. kaip mažiausias MAE, testuoja visas įmanomas nustatytų parametrų variacijas ir randa tinkamiausius atitinkamos problemos sprendimui. Tačiau naudojant tinklo paiešką yra ir tam tikrų trūkumų Nors šis įrankis ir įvertina visas pateiktas parametrų variacijas, šis procesas yra labai technologiškai intensyvus ir gali būti labai imlus laikui. Dėl šios priežasties parametrų paieška gali trukti nuo 1 valandos iki kelių dienų, savaitių ar net metų. Todėl renkantis parametrų paieškos variacijas buvo pasirinkta naudoti standartinės variacijas, naudojamas kituose moksliniuose darbuose. Kaip matoma (žr. 7 lentelę), paslėptas dydis, paslėptų sluoksnių kiekis, partijos dydis, ir atsitiktinės praretinimo transformacijos parametrai visais atvejais buvo palikti vienodi, kadangi šios variacijos yra ganėtinai standartinės ir nėra labai jautrios duomenų imčiai ar struktūrai. Atsižvelgiant į epochas ir „Lookback“, šios variacijos buvo koreguotos atsižvelgiant į visą duomenų imtį ir kompiuterinius pajėgumus. Kaip pavyzdys, kadangi vienmačių duomenų imtis yra mažesnė nei daugiamačių, vienmačių duomenų „Lookback“ galima nustatyti šiek tiek didesnę, per daug nepailginus paieškos laiko. Panašiai ir su epochų pasirinkimu, kadangi vienmačių duomenų yra kur kas daugiau, epochas tenka nustatyti mažesnes, kadangi modelio treniravimas tampa per daug imlus laikui. Mėnesiniams duomenims epochas galima laikyti kur kas didesnes, smarkiai nepailginant treniravimosi laiko. Apibendrinus, kaip galima matyti iš 7 lentelės, net ir naudojant reliatyviai nedidelį galimų parametrų kiekį, testuojamų modelių kiekis auga labai greitai. Naudojant šią parametrų imtį, tinklo paieškos metu būtų įvertinti net 3 456 modeliai.

7 lentelė. Tyrimo metu naudojama tinklo paieškos parametrų derinimo imtis

Tinklo paieška	Vienmačiai duomenys		Daugiamačiai duomenys	
	LSTM ir GRU dieniniai	LSTM ir GRU mėnesiniai	LSTM ir GRU dieniniai	LSTM ir GRU mėnesiniai
Paslėptas dydis	[50, 200, 500]	[50, 200, 500]	[50, 200, 500]	[50, 200, 500]
Paslėptų sluoksnių kiekis	[1, 2]	[1, 2]	[1, 2]	[1, 2]
EPOCHS	[200, 500, 1000]	[1000, 2000, 5000]	[200, 500, 1000]	[1000, 2000, 5000]
Partijos dydis	[32, 64, 128]	[32, 64, 128]	[32, 64, 128]	[32, 64, 128]
„Lookback“	[5, 15, 30, 60]	[2, 6, 9, 12]	[5, 15, 30, 60]	[2, 6, 9, 12]
Atsitiktinė praretinimo transformacija	[0.0, 0.1]	[0.0, 0.1]	[0.0, 0.1]	[0.0, 0.1]
Galimų modelių kiekis	432 LSTM 432 GRU	432 LSTM 432 GRU	432 LSTM 432 GRU	432 LSTM 432 GRU

Tiek tradiciniams, tiek mašininio mokymusi pagrįstiems prognozavimo metodams yra svarbu tinkamai įvertinti modelių tikslumą. Jis gali būti apskaičiuojamas įprastais metodais. Remiantis Omar'u, Huang'u, Salameh'u, Khurram'u ir Fareed'u (2022), modelio našumą galima apskaičiuoti pasitelkiant, MAE, MSE, RMSE ir MAPE. Šios paklaidos gali būti apskaičiuojamos pagal (3.4), (3.5), (3.6) ir (3.7) formules:

$$MAE = \frac{\sum |Y_A - f|}{m} \quad (3.4)$$

$$MSE = \frac{\sum (Y_A - f)^2}{m} \quad (3.5)$$

$$RMSE = \sqrt{\frac{\sum (Y_A - f)^2}{m}} \quad (3.6)$$

$$MAPE = \frac{1}{m} \sum \left| \frac{Y_A - f}{Y_A} \right| \quad (3.7)$$

čia Y_A – prognozuojama duomenų vertė;

f – tikra duomenų vertė;

m – stebėjimų skaičius bandomajame duomenų rinkinyje.

Apibendrinus metodologinę dalį, atrinkus minėtus duomenis, bus atliekama stebinių analizė. Kintamieji bus suvienodinti laiko ir mastelio atžvilgiu, o trūkstamos vertės užpildytos naudojantis vėliausiomis žinomomis vertėmis. Vėliau, iš daugiamačių duomenų perspektyvos, bus atlikta „Granger“ priežastingumo ir „Pearson“ koreliacijos analizė, taip atrenkant tik reikšmingus kintamuosius. Sutvarkius duomenis ir atrinkus tik svarbius kintamuosius, tyrimo metu bus apskaičiuojami ir vertinami naivaus, ARIMA, ARDL, LSTM ir GRU modeliai, prognozuojant „OMX Vilnius“ gražos indeksą. Geriausi ARIMA ir ARDL parametrai bus parinkti remiantis pirmais 90% duomenų, o gautų modelių tikslumas bus vertinimas su likusiais 10%. Lygiai taip pat bus apmokomi mašininio mokymosi modeliai, 90% bus naudojama modelio mokymui, o likę 10% prognozavimui. Pagrindinės naudojamos programos bus „Eviews“ ir „Python“ su „Pytorch“, „Statsmodel“, „auto.arima“, „Pandas“ ir „Numpy“ bibliotekomis. Mašininio mokymosi modelių parametrų radimui

bus naudojama tinklo paieška, atsižvelgiant į kompiuterinius pajėgumus ir laiko imlumą. Galiausiai, visų modelių prognozės bus palygintos su faktiniais, išgaunant MSE, RMSE, MAE ir MAPE rodiklius, pagal kuriuos ir bus lyginami modelių tikslumai. Gauti rezultatai padės nustatyti, ar mašininio mokymosi modeliai turi pranašumą „OMX Vilnius“ gražos indekso prognozavimui.

4. Gilaus mašininio mokymosi ir klasikinių modelių lyginamoji analizė, prognozuojant „OMX Vilnius“ gražos indeksą

Tyrimui atlikti, kaip minėta metodologijoje, buvo atrinkti didžiausią rinkos kapitalizaciją savo kategorijose turintis turtas. Surinkus dieninius duomenis, jais remiantis buvo sukurta mėnesinių duomenų imtis. Atsižvelgiant į atrinktus duomenis, buvo apskaičiuotos pagrindinės skaitinės charakteristikos ir atlikta grafinė kintamųjų analizė. Taip pat, kaip ir minėta anksčiau, duomenys buvo unifikuoti. Atsižvelgiant į išskirčių nustatymą, buvo aptiktos išskirtys, tačiau, kadangi tai yra biržoje prekiaujamos vertės, būtų jas neteisinga pašalinti iš duomenų imties. Kadangi tai yra ekstremalūs judėjimai, jie įvyko kaip faktas, pagal kuriuos investuotojai priėmė vėlesnius sprendimus. Dėl šios priežasties ir remiantis Nevasalmi'o (2020) rekomendacija išskirtis naudoti modelio apmokyme, buvo nuspręsta jas palikti. Atsižvelgiant į duomenų normalumą, deja, bet dauguma kintamųjų nėra pasiskirstę pagal normalųjį skirstinį, o bandant juos pertvarkyti pasitelkus logaritmavimą ar kitas funkcijas, rezultatai nesikeičia. Galutinė atrinktų duomenų imtis susidarė iš dieninių ir mėnesinių 30/09/2004 - 31/03/2024 „OMX Vilnius“ gražos indekso ir 41 kintamojo (atvaizduoto 4 lentelėje) duomenų.

4.1. „Granger“ priežastingumo, „Pearson“ koreliacijos ir geriausių mašininio mokymosi modelių parametrų analizė

Prieš pradėdant atlikti įvairių prognozavimo modelių analizę, iš pradžių buvo įvertinti kintamieji, kurie bus naudojami daugiamačiuose modeliuose. Tam buvo atlikti „Granger“ priežastingumo ir „Pearson“ koreliacijos tyrimai tiek dieniniams, tiek mėnesiniams duomenims (žr. 1 ir 2 priedus). Atsižvelgiant į dieninius duomenis, iš šio tyrimo galima pastebėti, jog visgi didžioji dalis nagrinėjamų kintamųjų rodo esamą „Granger“ priežastingumą. Todėl galima būtų daryti išvadą, kad yra galimas ryšis tarp ganėtinai didelio kiekio rinkose prekiaujamo turto dieninių duomenų atžvilgiu. Iš bendros dieninių duomenų imties „Granger“ priežastingumas netenkinio tik „Telia Lietuva“, „AB Šiaulių Bankas“, „Grigeo AB“, „AS Latvijas Gaze“, „AS Amber Latvijas balzams“ ir Euro zonos obligacijų 3, 5 ir 10 metų pajamingumo kreivių. Įvertinus ir koreliacijos koeficientus, kurie bus aptarti vėliau, buvo atrinkti 3 indeksai, kurie dieninių „OMX Vilnius“ gražos indekso verčių prognozavime yra reikšminiai, vertinant 1-10 vėlavimų intervalą (žr. 8 lentelę). Kaip matyti iš gautų rezultatų, „Granger“ priežastingumas tarp tokių tarptautinių indeksų kaip „DAX“ ar „CAC40“ yra matomas visais nagrinėtais laikotarpiais. Tačiau atsižvelgiant į „Invalda“, yra matomas šioks toks vėlavimas, remiantis šiais rezultatais panašu, kad turi praeiti bene savaitė, iki kol yra pastebimas „OMX Vilnius“ gražos indekso judėjimas.

8 lentelė. Reikšmingų dieninių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu.

„Granger“ priežastingas	(F-Statistika /P-Tikimybė)									
	1 vėl.	2 vėl.	3 vėl.	4 vėl.	5 vėl.	6 vėl.	7 vėl.	8 vėl.	9 vėl.	10 vėl.
Invalda	3.203	5.233	5.755	6.990	9.191	9.271	10.383	17.702	19.033	20.515
INVL	/0.074	/0.073	/0.124	/0.137	/0.102	/0.159	/0.168	/0.024*	/0.025*	/0.025*
DAX	27.695	28.953	51.746	52.852	54.501	58.264	70.900	73.274	73.659	73.931
	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*
CAC40	45.794	47.947	60.397	61.398	65.765	70.676	78.501	81.100	82.000	81.682
	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*	/0*

Tačiau, kitoks vaizdas yra matomas nagrinėjant mėnesinius duomenis (žr. 2 priedą). Įvertinus „Granger“ priežastingumą, yra matoma, kad didelis kiekis kintamųjų iš tiesų netenkina „Granger“ P-tikimybės. Iš 41 kintamojo, „Granger“ priežastingumas buvo matomas tik tarp „AB Šiaulių Bankas“, „Grigeo AB“, „Invalda INVL“, „Nordic Fibreboard AS“, „Investor AB“, „Apple Inc.“, „Crude Oil“, „USD/EUR“ ir „CAC40“. Apibendrinus, atlikus „Granger“ priežastingumo tyrimą, dieninių duomenų imtis sumažėjo nuo 41 iki 33, o mėnesinių duomenų sumažėjo nuo 41 iki 9 kintamųjų. Tai rodo, kad nors dieniniai duomenys galimai ir turi priklausomybę nuo daugelio didžiausių biržoje prekiaujamų įrankių judėjimų, ši priklausomybė žymiai sumažėja nagrinėjant didesnes laiko žymas. Įvertinus mėnesinę „Pearson“ koreliaciją, kuri bus aptarta vėliau, taip pat buvo atrinkti 3 reikšmingi kintamieji, 1-10 vėlavimų intervale (žr. 9 lentelę). Kaip matoma, 2 iš 3 kintamųjų pasiliko tie patys – „Invalda INVL“ ir „CAC40“. Atsižvelgiant į tai, jog mėnesinių kintamųjų, kurie tenkina „Granger“ priežastingumą, yra žymiai mažiau, galima daryti išvadą, kad šie du rodikliai, priežastingumo verte, yra išties susiję su „OMX Vilnius“ gražos indeksu. Įdomu tai, kad vėlintų laikotarpių P-tikimybės tarp dieninių ir mėnesinių duomenų taip pat skiriasi. Mėnesinių duomenų atžvilgiu „Invalda“ visais laikais tenkina „Granger“ priežastingumą, kol „CAC40“ yra reikšmingas tik vienu laikotarpiu atgal. Taip pat iš naujų reikšmingų kintamųjų prisideda ir „AB Šiaulių bankas“, kadangi mėnesinių duomenų analizėje šis kintamasis pradėjo tenkinti „Granger“ priežastingumo kriterijų.

9 lentelė. Reikšmingų mėnesinių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu.

„Granger“ priežastingas	(F-Statistika /P-Tikimybė)									
	1 vėl.	2 vėl.	3 vėl.	4 vėl.	5 vėl.	6 vėl.	7 vėl.	8 vėl.	9 vėl.	10 vėl.
AB Šiaulių Bankas	9.514 /0.002*	9.636 /0.008*	8.991 /0.029*	8.795 /0.066	10.200 /0.070	10.826 /0.094	10.802 /0.148	10.249 /0.248	12.680 /0.178	14.441 /0.154
Invalda INVL	5.075 /0.024*	7.483 /0.024*	9.426 /0.024*	12.620 /0.013*	16.145 /0.006*	15.876 /0.014*	16.477 /0.021*	16.944 /0.031*	19.052 /0.025*	23.746 /0.008*
CAC40	4.850 /0.028*	4.901 /0.086	4.313 /0.230	5.723 /0.221	5.552 /0.352	5.756 /0.451	9.626 /0.211	9.457 /0.305	9.498 /0.393	8.974 /0.535

Įvertinus „Granger“ priežastingumą, kitas žingsnis kintamųjų analizėje įvertinti ir aptarti jau minėtus „Pearson“ koreliacijos koeficientus. Kadangi koreliacija yra svarbi tik tarp „OMX Vilnius“ gražos indekso ir kitų kintamųjų, tik ji ir buvo vertinama. Žinoma, buvo pažvelgta ir į tarpusavio kintamųjų koreliaciją, tačiau stipriai tarpusavyje koreliuojančių kintamųjų aptikta nebuvo. Nagrinėjant dieninius duomenis (žr. 3 priedą) buvo pastebėta, jog tarp „OMX Vilnius“ gražos indekso ir visų kitų kintamųjų nebuvo pastebėta net vidutinė koreliacija (virš |0.5|). Tačiau tarp kelių kintamųjų buvo pastebėta silpna koreliacija (tarp |0.3| ir |0.5|). Didžiausia koreliacija pasižymėjo „Telia Lietuva“ su 0.45, „AB Šiaulių bankas“ su 0.44, „Invalda INVL“ su 0.32, „DAX“ su 0.31 ir „CAC40“ su 0.31. Deja, nors „Telia Lietuva“ ir „AB Šiaulių bankas“ turėjo didžiausią dieninę koreliaciją, dėl netenkinto „Granger“ testo šie duomenys turėjo būti neįtraukti į tolimesnius skaičiavimus. Taip pat verta paminėti, kad nei vienas iš reikšmingą koreliaciją turinčių kintamųjų neturėjo neigiamos koreliacijos, tai reiškia, jog visi šie kintamieji koreliuoja, kuomet jų vertė juda ta pačia kryptimi. Taip pat yra įdomu tai, jog nebuvo pastebėta koreliacija tarp Euro zonos obligacijų pajamingumo, kuris reprezentuoja infliacijos lygį. Į tai atsižvelgiant galima teigti, kad „OMX Vilnius“ gražos indekso kainų svyravimai nepriklauso arba labai nežymiai priklauso nuo infliacijos ir palūkanų normų. Taip pat verta paminėti, kad dieniniai „OMX Vilnius“ gražos indekso svyravimai labai silpnai koreliuoja su valiutų kursų judėjimais, natūraliomis dujomis, nafta ar auksu. Galiausiai, remiantis koreliacijos matrica galima teigti, kad „OMX Vilnius“ gražos indeksas padieniu ganėtinai koreliuoja su kitais tarptautiniais

indeksais, tokiais kaip „S&P 500“, „DAX“, ir „CAC40“. Taigi remiantis „Granger“ priežastingumu ir koreliacijos matrica buvo atrinkti rodikliai, tinkantys tolimesnei analizei ir modelių pritaikymui (žr. 10 lentelę). Nors šių kintamųjų koreliacija yra ganėtinai silpna, ji vis tiek egzistuoja. O atsižvelgiant į tai, kad čia yra dieniniai indekso judėjimai, kurie priklauso nuo skirtingų akcijų verčių judėjimų bei investuotojų nuomonės ir noro pirkti ar parduoti šias akcijas, tikėtis rasti didelę koreliaciją turinčius kintamuosius nėra labai realistiška.

10 lentelė. Reikšmingų dieninių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu

„Pearson“ koreliacijos koeficientas	„OMX Vilnius“ gražos indeksas
„OMX Vilnius“ gražos indeksas	1
Invalida INVL	0.3219
DAX	0.3142
CAC40	0.3126

Analizuojant mėnesinių „OMX Vilnius“ gražos indekso kintamųjų koreliacijos koeficientus buvo pastebėta tam tikrų skirtumų (žr. 4 priedą). Pirmiausia, absoliutus kintamųjų koreliacijos vidurkis tarp mėnesinių duomenų yra 7 procentiniais punktais didesnis (dieninių vidurkis – 0.21, mėnesinių vidurkis – 0.28). Atmetus kintamuosius, kurie netenkina „Granger“ priežastingumo, skirtumas yra net 19 procentinių punktų (dieninių vidurkis – 0.32, mėnesinių vidurkis – 0.51). Tai parodo, jog didesnės laiko žymos duomenys yra kur kas labiau koreliuojantys su „OMX Vilnius“ gražos indeksu. Taip pat verta paminėti, kad priešingai nei dieniniai duomenys, mėnesiniai kintamieji siekia vidutinę koreliaciją (tarp |0.5| ir |0.7|) (žr. 11 lentelę). Atsižvelgiant į tai, kad „AB Šiaulių Bankas“, „Invalida INVL“ ir „CAC40“ tenkina ir „Granger“ priežastingumą, galima daryti išvadą, jog šie kintamieji yra pakankamai reikšmingi naudoti tiek klasikinio, tiek gilaus mašininio mokymosi modelių prognozėse.

11 lentelė. Reikšmingų mėnesinių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu

„Pearson“ koreliacijos koeficientas	„OMX Vilnius“ gražos indeksas
„OMX Vilnius“ gražos indeksas	1
AB Šiaulių Bankas	0.598
Invalida INVL	0.500
CAC40	0.443

Apibendrinant, iš nagrinėto 41 kintamojo reikšmingi atrinkti kintamieji buvo tik 4: dieniniams – „Invalida“, „DAX“ ir „CAC40“; mėnesiniams „AB Šiaulių Bankas“, „Invalida“ ir „CAC40“. Visi kiti 37 kintamieji, remiantis naudotais atrinkimo procesais, buvo nereikšmingi. Dėl šių rezultatų, galima daryti išvadą, kad „OMX Vilnius“ gražos indeksas yra mažai įtakojamas didelę kapitalizaciją turinčių akcijų, žaliavų, valiutų kursų ar net Euro zonos obligacijų pajamingumo. Vieninteliai reikšmingi kintamieji, kurie daro įtaką šiam indeksui, yra Lietuvos įmonės ir keli tarptautiniai panašaus pobūdžio indeksai.

Kitas svarbus žingsnis prieš giliu mašininio mokymosi grįstą prognozavimo modelių vertinimą yra tinkamiausių parametrų parinkimas. Kaip minėta metodologijoje, šiam tyrimui buvo naudojamas tinklo paieškos metodas (žr. 7 lentelę). Iš viso buvo atliktos 8 paieškos, kiekvienam modeliui įvertinant 96 parametrų variacijas. Kadangi paieška ir taip užtruko ganėtinai ilgai (~50 valandų),

parametrų imtis būtent tinklo paieškai buvo sumažinta. Dėl reikalingų didelių kompiuterinių pajėgumų, iš parametrų variacijos imties buvo pašalinti šie parametrai: dieniniams duomenims (epochos – 1000, partijos dydis – 32), mėnesiniams (epochos – 5000, partijos dydis – 32). Vėliau atradus tiksliausią modelį, jis buvo pakoreguotas, ranka pakeičiant epochos ir partijos dydžio vertes į minėtas. Po korekcijos, modelio rezultatai buvo iš naujo įvertinami ir tam tikrais atvejais, kaip matyti 12 lentelėje, nauji parametrai, sugebėjo pasiekti geresnių rezultatų. Tinklo paieškos metu ir rankinio testavimo metu buvo pastebėtos kelios išvalgos apie „OMX Vilnius“ gražos indekso LSTM ir GRU modelių parametrų tinkamumą. Pirmiausia, atsižvelgiant į paslėptą dydį, buvo pastebėta, jog rezultatai LSTM atveju yra pasiekiami geresni, kuomet nustatytas parametras yra didesnis. Atliekant rankinius ir tinklo paieškos tyrimus buvo pastebėta, kad didinant šį parametras ir toliau, galima išgauti geresnius rezultatus, tačiau viršijus 200 rezultatų gerėjimas pradeda eksponentiškai mažėti, kol treniravo laikas vis didėja. Taip pat didesnė nei 200 parametro vertė veda prie perteklinio modelio pritaikymo (angl. *overfit*). Kitas pastebėjimas buvo, jog didesnis paslėptų sluoksnių kiekis, dažniais atvejais, taip pat privedavo prie perteklinio pritaikymo. Atsižvelgiant į partijos dydį ir epochas šių parametrų radimas yra ganėtinai priklausomas nuo kitų nustatytų parametrų. Buvo rasta, kad mažinant partijos dydį rezultatai gerėja iki kol modelis tampa pertekliniai pritaikytas, taip pat šis procesas labai išaugina mokymosi laiką. Epochos, tuo tarpu, veikia panašiai, tik rezultatai gerėja šį parametras didinant taip pat iki tam tikros ribos, kol modelis tampa pertekliniai pritaikytas. Abejais atvejais ar partijos dydžio mažinimas, ar epochų didinimas, šie procesai smarkiai didina mašininio mokymosi treniravimo laiką. Atsižvelgiant į „Lookback“, tinklo paieškos rezultatai buvo ganėtinai nepastovūs, vienais atvejais didesnė duomenų istorija lemdavo tikslesnius rezultatus, kitais atvejais ne. Atsitiktinė praretinimo transformacija naudojama kaip apsauga nuo perteklinio apmokymo, kuomet modelis susideda iš dviejų ar daugiau paslėptų sluoksnių. Kadangi tik vienas modelis turėjo 2 paslėptus sluoksnius, šis parametras galėjo būti pritaikytas tik mėnesiniam GRU modeliui. Tačiau atsižvelgiant į tai, kad jį pritaikius rezultatai nepagerėjo, galima teigti, kad mėnesinis GRU modelis nebuvo toks sudėtingas, kad pasiektų perteklinį apmokymą.

12 lentelė. Geriausi rasti tinklo paieškos parametrai mašininio mokymosi modeliams

Tinklo paieška	Vienmačiai duomenys				Daugiamačiai duomenys			
	Dieniniai		Mėnesiniai		Dieniniai		Mėnesiniai	
	LSTM	GRU	LSTM	GRU	LSTM	GRU	LSTM	GRU
Paslėptas dydis	200	50	200	200	200	200	200	50
Paslėptų sluoksnių kiekis	1	1	1	1	1	1	1	2
Epochos	200	500	2000	2000	200	200	1000	1000
Partijos dydis	32	64	128	128	64	32	32	128
„Lookback“	60	30	6	6	30	30	9	9
Atsitiktinė praretinimo transformacija	0	0	0	0	0	0	0	0

Taigi, nors naudojant šiuos parametrus rezultatai buvo gauti ganėtinai neblogi, jie jokiais būdais nėra galutiniai. Modelių tikslumas dar tikrai gali būti didinamas naudojant platesnę tinklo paiešką ar pasitelkiant kitas parametrų radimo priemones, tokias kaip atsitiktinė paieška, genetinis algoritmas ar kiti pažangesni metodai.

4.2. Modelių vertinimas remiantis vienmačiais duomenimis

Atlikus priklausomų ir nepriklausomų kintamųjų analizę, atrinkus tik reikšmingus kintamuosius bei atradus tinkamiausius parametrus, naudojamus mašiniam mokymuisi, buvo pradėtas 14 modelių tyrimas. Tyrimas buvo padalintas į 4 dalis pagal duomenų imtis – vienmačiai dieniniai, vienmačiai mėnesiniai, daugiamačiai dieniniai ir daugiamačiai mėnesiniai. Visų gautų modelių rezultatai buvo lyginami su paskutiniais 10% duomenų.

13 lentelė. Naiviuoju metodu apskaičiuotų prognozių tikslumas

Rodiklis	Dieninis tikslumas	Mėnesinis tikslumas
MSE	7.05	388.34
RMSE	2.66	19.71
MAE	1.64	16.72
MAPE	0.1725%	1.7697%

Atliekant tyrimą pirmasis modelis baziniam visų modelių vertinimui buvo atliktas pasitelkiant naivųjį (angl. *naive*) metodą. Šis modelis buvo naudojamas kaip lyginamasis standartas tam, kad įvertinti, kaip tiksliai gali prognozuoti tiek tradiciniai, tiek mašininio mokymusi grįsti prognozavimo modeliai. Jeigu modelio rezultatai bus prastesni nei 13 lentelėje pateikti duomenys, vadinasi modelis nėra pakankamai patikimas, kad jį būtų galima naudoti. Kaip matyti iš gautų duomenų, dieninių rezultatų paklaida yra mažesnė nei mėnesinių. Žinoma, taip yra dėl to, kad „OMX Vilnius“ gražos indekso vertė per mėnesį pajuda kur kas daugiau nei per dieną.

Pereinant prie dieninių duomenų vienmačių modelių vertinimo, pirmasis modelis, kuris buvo vertinimas, yra ARIMA. Kaip ir minėta metodologijoje, geriausio modelio radimui buvo pasitelkta „Python“ „auto.arima“ biblioteka. Šio įrankio dėka, pagal geriausią AIC kriterijų buvo atrinktas tiksliausias ARIMA modelis. Kaip matyti 9 paveiksle tiksliausias gautas modelis yra SARIMAX(3,1,1)(2,0,0)[7]. Remiantis gautu modeliu matyti, jog visi koeficientai, išskyrus „intercept“ (konstanta), yra statistiškai reikšmingi. Modelis taip pat tenkina paklaidų autokoreliacijos (Ljung-Box p-tikimybė > 0.05) ir heteroskedastijos hipotezes (p-tikimybė < 0.05). Tačiau nors šios dvi hipotezės yra tenkinamos, paklaidos nėra pasiskirsčiusios pagal normalųjį skirstinį (Jarque-Bera p-tikimybė < 0.05), o jų vidurkis nėra lygus nuliui (0.068). Apibendrinus, yra tenkinamos tik 2/4 hipotezių, tai rodo, kad modelis nėra visiškai tinkamas prognozavimui. Tačiau tokie rezultatai buvo tikėtini, kadangi šio tyrimo metu yra prognozuojamas „OMX Vilnius“ gražos indeksas, kuris yra apskaičiuojamas remiantis biržoje parduojamų akcijų judėjimais. Dėl šios priežasties yra mažai tikėtina, kad ganėtinai paprastas vienmatis ARIMA modelis sugebėtų išvelgti visas judėjimo priklausomybes patikimam rodiklio prognozavimui. Prognozuoti akcijų biržos judėjimus yra be galo sudėtinga užduotis, kainų judėjimas priklauso nuo labai didelio kiekio kintamųjų, įskaitant ir investuotojų nuomones ir sprendimus pirkti ar parduoti. Dėl šių priežasčių tyrimo metu buvo labiau orientuojamasi į prognozės tikslumą, o ne hipotezių tenkinimą. Atsižvelgiant į tai, jog šis modelis sugebėjo pasiekti geriausius rezultatus pagal AIC kriterijų, jis ir bus naudojamas tolimesniame tyrime.

```

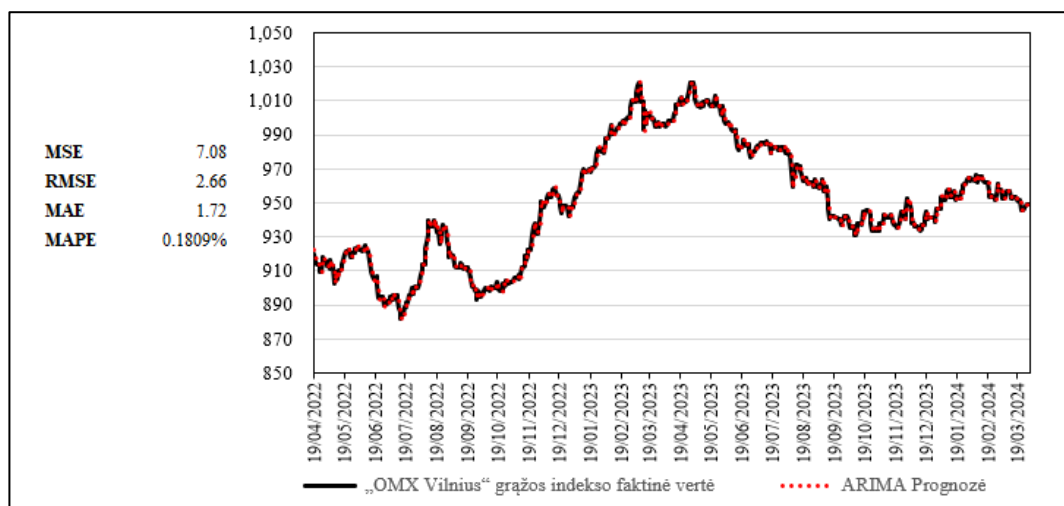
Best model: ARIMA(3,1,1)(2,0,0)[7] intercept
Total fit time: 233.826 seconds

SARIMAX Results
=====
Dep. Variable: y No. Observations: 6410
Model: SARIMAX(3, 1, 1)x(2, 0, [], 7) Log Likelihood: -17125.361
Date: Sun, 21 Apr 2024 AIC: 34266.722
Time: 21:06:52 BIC: 34320.846
Sample: 09-30-2004 HQIC: 34285.455
- 04-18-2022
Covariance Type: opg
=====
coef std err z P>|z| [0.025 0.975]
-----
intercept 0.0094 0.007 1.342 0.180 -0.004 0.023
ar.L1 0.8980 0.021 42.813 0.000 0.857 0.939
ar.L2 -0.0295 0.008 -3.497 0.000 -0.046 -0.013
ar.L3 0.0452 0.009 5.176 0.000 0.028 0.062
ma.L1 -0.8639 0.021 -42.001 0.000 -0.904 -0.824
ar.S.L7 -0.0321 0.006 -5.288 0.000 -0.044 -0.020
ar.S.L14 0.0408 0.007 6.198 0.000 0.028 0.054
sigma2 12.2594 0.057 216.008 0.000 12.148 12.371
=====
Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 583659.67
Prob(Q): 0.98 Prob(JB): 0.00
Heteroskedasticity (H): 1.23 Skew: -1.95
Prob(H) (two-sided): 0.00 Kurtosis: 49.59
=====

```

9 pav. Tiksliausias dieninių duomenų pagal AIC kriterijų išgautas modelis, naudojant „auto.arima“

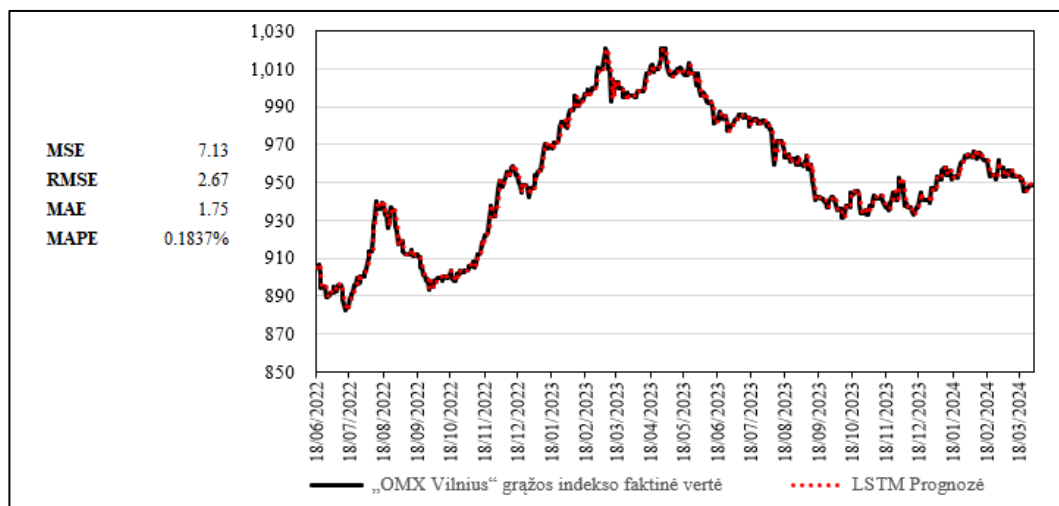
Remiantis geriausiu rastu ARIMA modeliu – SARIMAX(3,1,1)(2,0,0)[7], paskutiniams 10% dieninių duomenų buvo atlikta prognozė (žr. 10 pav.). Kaip matyti MAPE rezultatai lyginant su naiviu metodu nėra tokie įspūdingi. Šiuo atveju modelis pasirodė prasčiau nei naivusis prognozavimo būdas (naivus – 0.1725%, SARIMAX – 0.1809%). Nors vizualiai pažvelgus į grafiką nėra matoma ryškių nukrypimų, jeigu modelis nesugeba aplenksti net naivaus prognozavimo būdo, nėra stipraus pagrindo juo remtis. Apibendrinus galima teigti, jog naudojant dieninius duomenis SARIMAX modelis nesugeba tiksliai ir patikimai prognozuoti „OMX Vilnius“ gražos indekso vertės.



10 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį SARIMAX(3,1,1)(2,0,0)[7] modelį

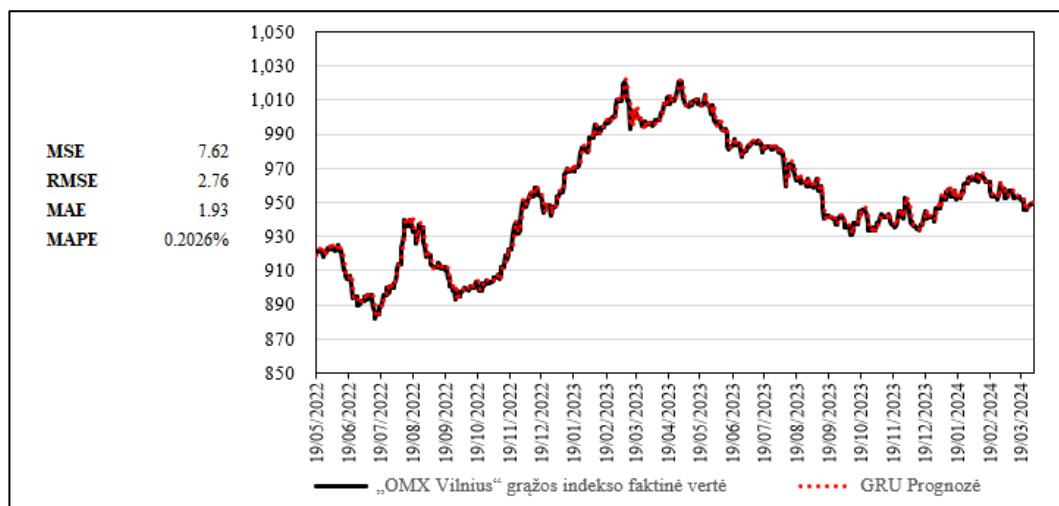
Pirmojo mašininiu mokymusi grįšto prognozavimo rezultatai yra matomi 11 paveiksle. Nors vizualiai nagrinėjant gautus rezultatus ir nėra matyti akivaizdžių kainos nukrypimų, gauti rezultatai rodo, jog modelis, taip pat kaip SARIMAX(3,1,1)(2,0,0)7, vis dėlto nesugebėjo aplenksti naivaus modelio tikslumo. Žinoma, šio mašininio mokymosi rezultatai labai priklauso nuo naudojamų parametru.

Atsižvelgiant į tai, kad tinklo paieškos metu parametų imtis nebuvo ypač didelė, yra tikimybė, kad gauti rezultatai nėra patys tiksliausi ir koreguojant parametrus jie dar gali gerėti. Tačiau šiuo metu, atlikus tyrimą su šia parametų imtimi galima teigti, kad LSTM modelis tinkamai prognozuoti „OMX Vilnius“ gražos indeksą nesugeba. Įvertinus tai, jog jau 2 modeliai nesugeba tinkamai prognozuoti dienišės „OMX Vilnius“ gražos indekso vertės, galima daryti išvadą, kad galbūt šio indekso negalima patikimai prognozuoti, panaudojus tik istorinius šio indekso judėjimus.



11 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį LSTM modelį

Antrasis naudojamas mašininis modelis dienišės „OMX Vilnius“ gražos indekso duomenų prognozei atlikti buvo GRU modelis. Kaip matome iš rezultatų (žr. 12 pav.), nors čia irgi nėra matomų akivaizdžių nukrypimų, šis modelis pasirodė dar prasčiau nei buvęs mašininio mokymosi modelis LSTM. Iš savo funkcionalumo pusės GRU modelis yra supaprastinta LSTM modelio atšaka. Dėl šios priežasties, taip pat remiantis nagrinėta literatūra, buvo tikimasi, kad šio modelio rezultatai bus šiek tiek prastesni nei LSTM. Tačiau nebuvo tikėtasi, jog abu šie mašininio mokymosi modeliai pasirodys prasčiau nei SARIMAX modelis ir prasčiau nei naivusis modelis. Todėl atsižvelgiant į šiuos gautus rezultatus yra keletas galimų paaiškinimų, kodėl taip galėjo nutikti. Pirma, tai tinklo paieškos parametų imtis buvo per maža, todėl modelis gavosi pernelyg paprastas, kad sugebėtų išvelgti gilesnes istorinių duomenų sąsajas. Antra priežastis gali būti ta, jog patikimai prognozuoti „OMX Vilnius“ gražos indeksą naudojant tik istorines šio rodiklio vertes yra neįmanoma. Gali būti, kad šio indekso vertės judėjimas padieniui yra tiesiog per daug nuspėjamas, kad būtų įmanoma išvesti patikimą prognozavimo modelį.



12 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį GRU modelį

Apibendrinant galima daryti išvadą, kad nors SARIMAX(3,1,1)(2,0,0)[7] modelis pasiekė geriausių rezultatų pagal AIC kriterijų, jo prognozavimo tikslumas „OMX Vilnius“ gražos indeksui buvo netenkinamas. Net lyginant su paprastesniu – naiviuoju metodu, ARIMA grįžtas modelis buvo mažiau tikslus. Kiti bandymai taikyti mašininio mokymosi modelius, įskaitant LSTM ir GRU, nepagerino prognozavimo tikslumo, o jų rezultatai buvo dar prastesni nei SARIMAX modelio. Šie rezultatai rodo, kad tiksliai prognozuoti dienes „OMX Vilnius“ gražos indekso vertes, remiantis vien istoriniais duomenimis, yra sudėtinga ir šiais laikais galbūt net neįmanoma. Galimos to priežastys – akcijų rinkos dinamikos sudėtingumas, nuspėjamumas ir taikomų modelių paprastumas dėl mažos parametrų imties. Norint patikimiau prognozuoti, gali prireikti toliau tyrinėti alternatyvias metodikas arba papildomus veiksnius, darančius įtaką rinkos judėjimams.

Nagrinėjant mėnesinius duomenis taip pat buvo atlikta „auto.arima“ paieška geriausių parametrų radimui (žr. 13 pav.). Kaip matyti, geriausias rastas modelis pagal AIC kriterijų šiuo atveju buvo SARIMAX(1,1,1)(0,1,1)[12]. Remiantis gautais rezultatais galima teigti, jog modelio visi koeficientai yra statistiškai reikšmingi, kadangi p-tikimybė < 0.05. Ljung-Box testas patvirtina, kad tarp paklaidų nėra autokoreliacijos. Atsižvelgiant į Jarque-Bera testą matoma, jog paklaidos nėra pasiskirsčiusios pagal normalųjį skirstinį. Heteroskedastijos testo p-tikimybė > 0.05, o tai parodo, kad tarp likučių yra homoskedastija, nors dėl šios priežasties viena iš prognozavimo patikimumo hipotezių nėra tenkinama, verta paminėti jog p-tikimybė labai nedaug viršija alpha. Galiausiai, apskaičiuotas paklaidų vidurkis yra 1.02. Apibendrinus, yra tenkinama tik 1 iš 4 prielaidų, tai rodo, jog modelis neapima visų mėnesinių „OMX Vilnius“ gražos indekso priklausomybių. Žinoma, taip pat kaip ir su dieniniais duomenimis, atsižvelgiant į tai, kad „OMX Vilnius“ gražos indeksas reprezentuoja akcijų biržų judėjimus, būtų optimistiška tikėtis, kad visos paklaidų hipotezės bus tenkinamos. Nepaisant to, „auto.arima“ atrinko tiksliausią modelį pagal AIC kriterijų, todėl tolimesniam tyrimui šis modelis ir bus naudojamas.


```

Best model: ARIMA(1,1,1)(0,1,1)[12]
Total fit time: 10.391 seconds

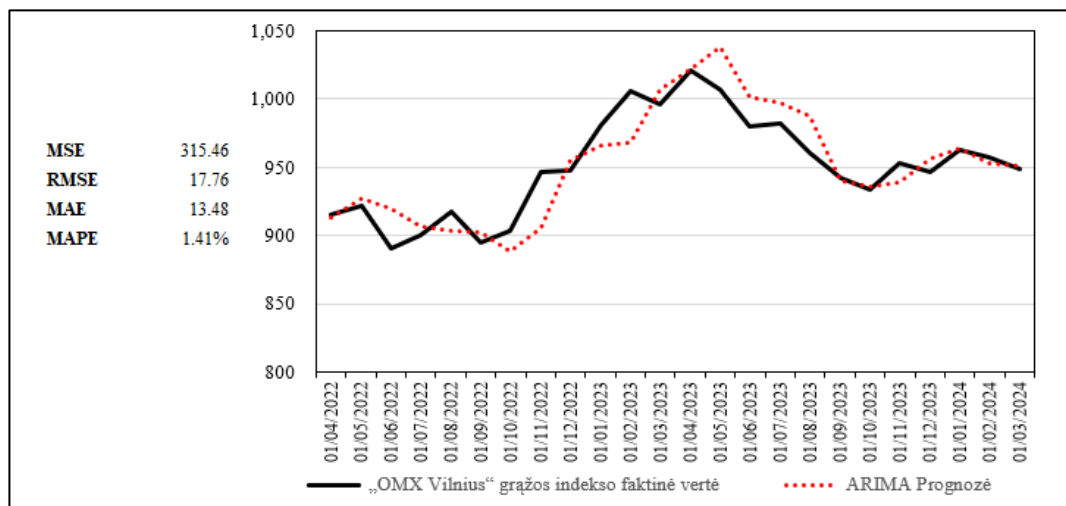
SARIMAX Results
=====
Dep. Variable: y No. Observations: 211
Model: SARIMAX(1, 1, 1)x(0, 1, 1, 12) Log Likelihood: -937.226
Date: Sun, 21 Apr 2024 AIC: 1882.452
Time: 20:04:17 BIC: 1895.605
Sample: 09-30-2004 HQIC: 1887.776
- 03-31-2022

Covariance Type: opg
=====
coef std err z P>|z| [0.025 0.975]
-----
ar.L1 0.8071 0.183 4.420 0.000 0.449 1.165
ma.L1 -0.7076 0.202 -3.510 0.000 -1.103 -0.313
ma.S.L12 -0.8420 0.090 -9.384 0.000 -1.018 -0.666
sigma2 702.5226 49.286 14.254 0.000 605.924 799.121
=====
Ljung-Box (L1) (Q): 0.02 Jarque-Bera (JB): 160.49
Prob(Q): 0.88 Prob(JB): 0.00
Heteroskedasticity (H): 0.64 Skew: -0.66
Prob(H) (two-sided): 0.07 Kurtosis: 7.21
=====

```

13 pav. Tiksliausias dieninių duomenų pagal AIC kriterijų išgautas modelis, naudojant „auto.arima“

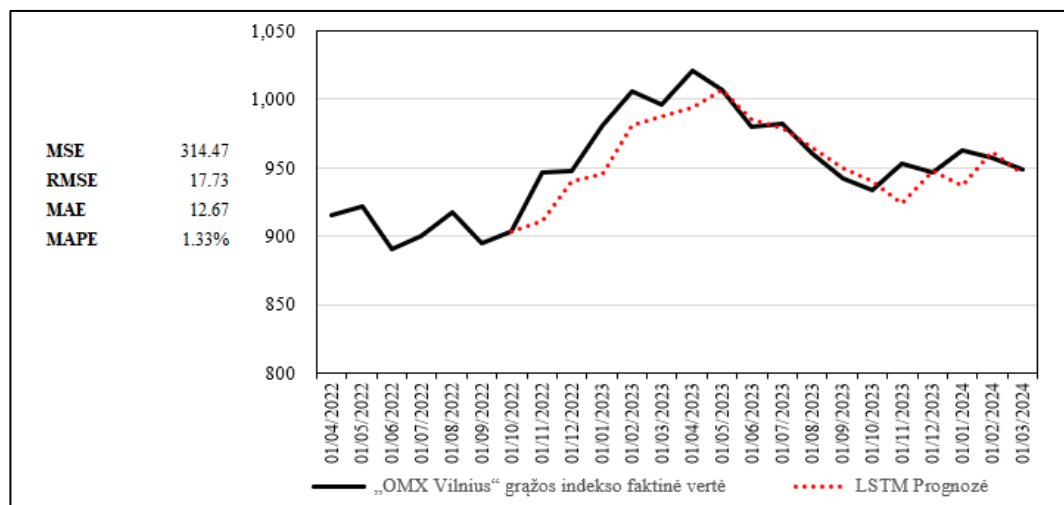
Remiantis geriausiu rastu ARIMA modeliu – SARIMAX(1,1,1)(0,1,1)[12], paskutiniams 10% mėnesinių duomenų buvo atlikta prognozė (žr. 14 pav.). Kaip matyti, MAPE rezultatai lyginant su naiviu metodu yra net 0.36 procentiniais punktais geresni. Tai rodo, kad nors šis modelis netenkina 3 iš 4 paklaidų prielaidų, jo prognozės tikslumas yra visgi reikšmingas. Tą galima patvirtinti ir pažvelgus į linijinę diagramą, kurioje matyti, jog rezultatai iš tiesų yra gana neblogi, vietomis nuspėjant „OMX Vilnius“ gražos indekso vertes beveik idealiai.



14 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant vienmatį SARIMAX(1,1,1)(0,1,1)[7] modelį

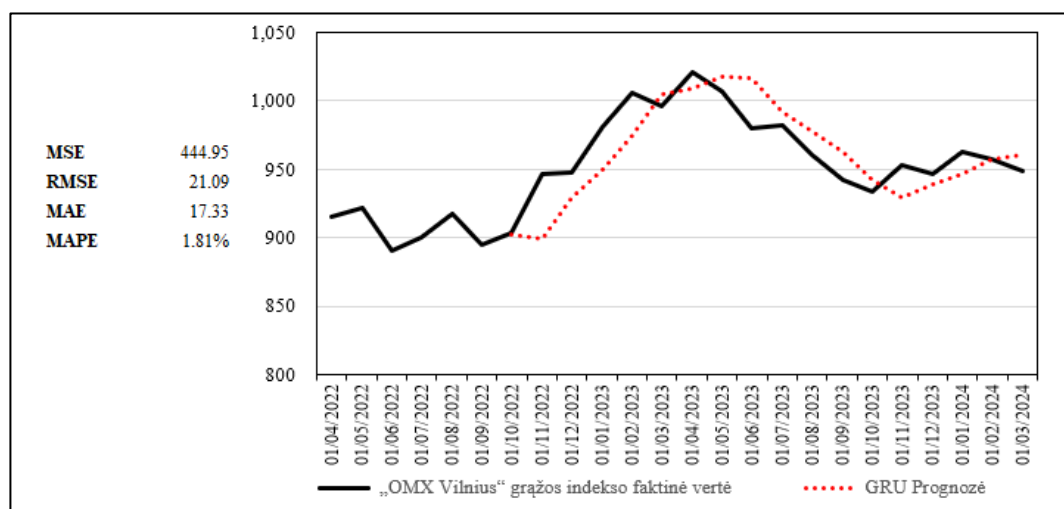
Toliau buvo pritaikytas LSTM modelis (žr. 15 pav.). Kaip matyti iš gautų rezultatų, šis modelis pasirodė dar geriau nei prieš tai nagrinėtas SARIMAX modelis. Taip pat verta pabrėžti, kad priešingai nei su dieniniais duomenimis, šis modelis iš tiesų sugebėjo aplenksti savo rezultatais tiek SARIMAX, tiek naivųjį metodą. Remiantis šiais rezultatais galima teigti, jog prognozuoti mėnesines „OMX Vilnius“ gražos indekso vertes, remiantis istoriniais judėjimais, galima kur kas tiksliau nei dienes. Panašu, kad priešingai nei dieniniai vertės judėjimai, kurie galimai didžiąja dalimi yra nenuspėjami, mėnesiniai svyravimai yra lengviau nuspėjami. Todėl remiantis šiais rezultatais galima būtų teigti, kad

mėnesines prognozes įmanoma atlikti tiksliau nei dienines. Taip pat panagrinėjus linijinę diagramą galima matyti ganėtinai nemažai verčių, kurios yra net labai artimos realioms vėrtėms, vietomis beveik idealios.



15 pav. Mėnesinė „OMX Vilnius“ gražos indekso vėrtės prognozė naudojant vienmatį LSTM modelį

Paskutinis vienmačių „OMX Vilnius“ gražos indekso duomenų nagrinėtas modelis buvo GRU. Kaip matyti (žr. 16 pav.), šio modelio rezultatai nebuvo ypač įspūdingi. Iš visų nagrinėtų mėnesinių vienmačių modelių, GRU pasirodė prasčiausiai. Gauti rezultatai buvo prastesni nei naiviu metodu gauta prognozė, tai rodo, jog šiuo modeliu visgi nereikėtų pasitikėti prognozuojant. Taip pat vizualiai nagrinėjant linijinę diagramą yra matomas akivaizdus modelio vėrtės atsilikimas, savo judėjimu ir rezultatais GRU beveik primena naivųjį metodą. Remiantis šiais rezultatais galima spėti, kad arba naudojama parametų imtis yra netinkama, arba GRU modelio architektūra yra per daug paprasta, norint patikimai prognozuoti „OMX Vilnius“ gražos indekso vėrtę, remiantis istoriniais vėrtės judėjimais.



16 pav. Mėnesinė „OMX Vilnius“ gražos indekso vėrtės prognozė naudojant vienmatį GRU modelį

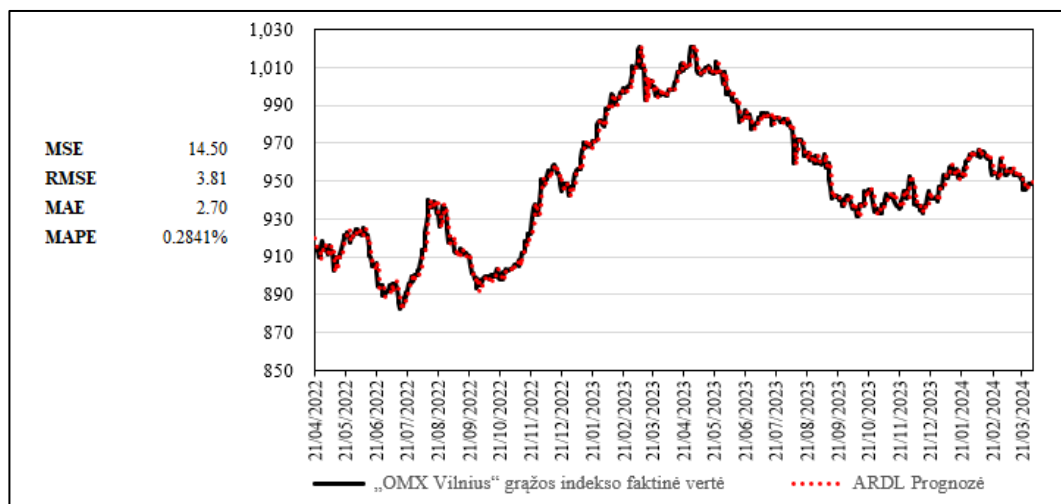
Apibendrinant, mėnesinė analizė naudojant „auto.arima“ parodė, kad pagal AIC kriterijų tiksliausias ARIMA modelis yra SARIMAX(1,1,1,1)(0,1,1)[12]. Nepaisant to, kad jis neatitiko trijų iš keturių paklaidų prielaidų, jo prognozavimo tikslumas reikšmingai pranoko naivųjį metodą, o tai rodo jo potencialų patikimumą prognozuojant mėnesines „OMX Vilnius“ gražos indekso vėrtes. Vėliau

pritaikius LSTM modelį, tikslumas, palyginti su SARIMAX ir naiviuoju metodais, dar labiau padidėjo, tai rodo, kad mėnesiniai indekso svyravimai yra labiau nuspėjami, palyginti su dieniniais svyravimais. GRU modelis pasižymėjo prasčiausiais rezultatais iš visų mėnesinių modelių, nesugebėdamas pranokti net naiviosios prognozės, o tai rodo galimus modelio architektūros ar parametru parinkimo trūkumus. Apibendrinus, šie rezultatai rodo, kad nors mėnesinės „OMX Vilnius“ gražos indekso prognozės gali būti patikimesnės už dienes, norint tiksliai prognozuoti, vis dar labai svarbu atidžiai apsvarstyti modelio parinkimą ir parametru optimizavimą.

4.3. Modelių vertinimas remiantis daugiamėčiais duomenimis

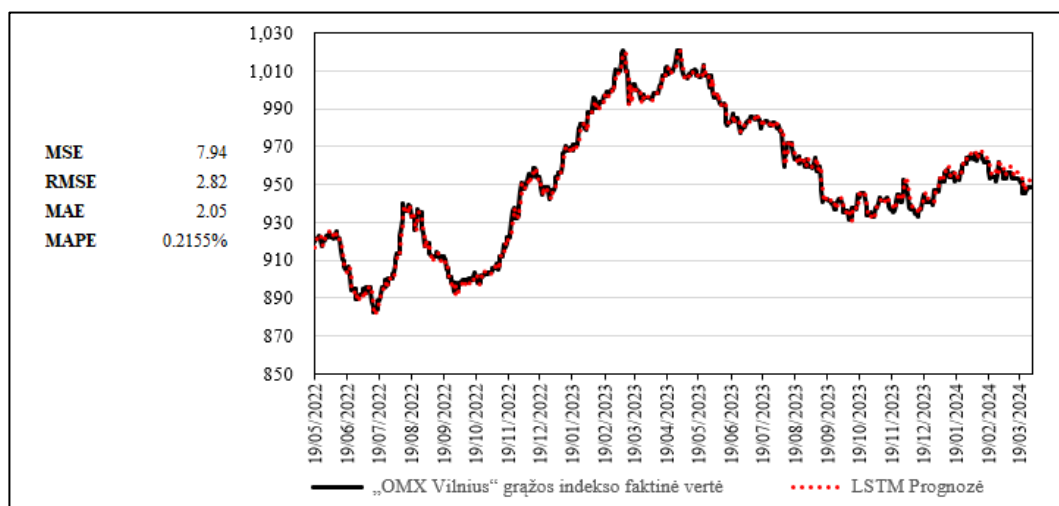
Išnagrinėjus įvairius vienmačius modelius, toliau buvo pereita prie daugiamėčių. Nors istoriniai „OMX Vilnius“ gražos indekso judėjimai turi naudingos informacijos nuspėjant ateities vertes, iš ekonominės pusės yra suprantama, kad indekso kaina nepriklauso tik nuo istorinių judėjimų. Tai yra indeksas, kuris seka įvairių Vilniaus biržoje prekiaujamų akcijų judėjimus. Šio indekso vertė priklauso nuo akcijų judėjimo, kurios greičiausiai priklauso nuo įvairių makro ir mikro ekonominių judėjimų – nuo kitų biržose prekiaujamų kainų judėjimo ar net nuo individualių žmonių nuomos ir noro įsigyti ar parduoti kažkurią iš Vilniaus biržoje prekiaujamų akcijų. Apibendrinus, kintamųjų kiekis, kurie gali daryti įtaką „OMX Vilnius“ gražos indekso judėjimui, gali būti neapskaičiuojamas, dėl šios priežasties yra būtina atlikti ne tik vienmačius prognozavimo modelius, tačiau ir daugiamėčius.

Pereinant prie dieninių duomenų daugiamėčių modelių, kaip minėta, buvo pasirinkta naudoti ARDL modelį. Taigi, naudojant 90% duomenų, remiantis geriausiu AIC kriterijumi „Eviews“ atrado šią dieninių kintamųjų struktūrą – „OMX Vilnius“ gražos indeksas 12; „Invalda“ 13; „DAX“ 8; „CAC40“ 4 arba dar kitaip – ARDL(12,13,8,4). Išgavus šį modelį, naudojant „Python“ įrankį buvo atliktos 1 dienos į priekį prognozės visam testavimo laikotarpiui. Kaip matyti 17 paveiksle, nors duomenys vizualiai atrodo neblogai, gauti rezultatai yra ganėtinai prasti, palyginus su anksčiau apskaičiuotais vienmačiais duomenimis. Atsižvelgiant į gautus rezultatus pastebėta, kad nors ir duomenys atitiko „Granger“ ir „Pearson“ koreliacijos kriterijus, visgi jie nepagerino situacijos taip, kaip tikėtasi. Todėl remiantis gautais rezultatais galima teigti, kad duomenys buvo labiau iškraipyti nei patikslinti. Atsižvelgiant į tai, jog gauti rezultatai yra prastesni nei visi kiti vienmačiai modeliai, įskaitant ir naivųjį, galima daryti išvadą, kad prognozuoti „OMX Vilnius“ gražos indeksą, pasitelkiant šiuos kintamuosius ir būtent šiuos parametrus, nereikėtų.



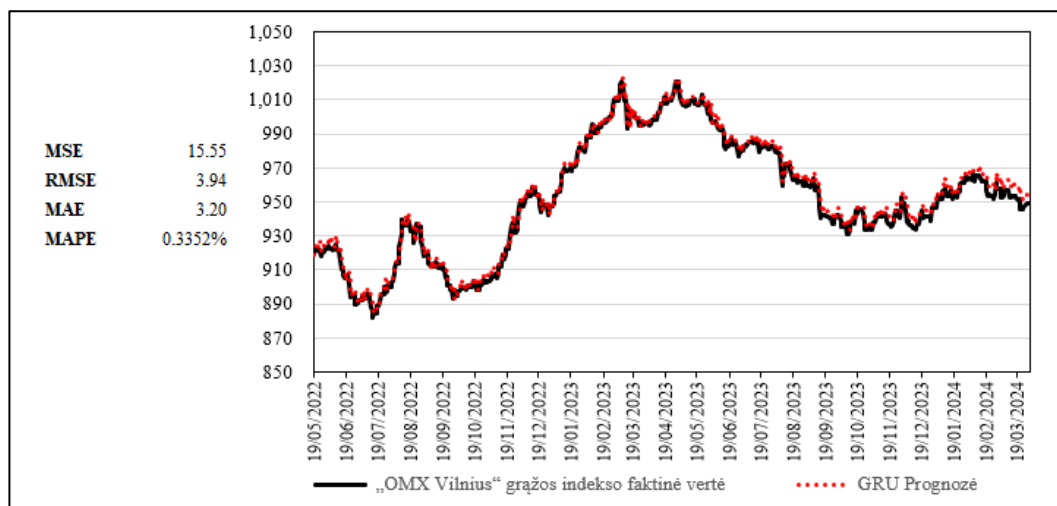
17 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamačią ARDL(12,13,8,4) modelį

Įvertinus ARDL modelį ir nepasiekus norimų rezultatų buvo pereita prie LSTM modelio pritaikymo. Deja, kaip matyti iš rezultatų (žr. 18 pav.), nors LSTM pasirodė geriau nei ARDL, lyginant su vienmačiais duomenų modeliais šie rezultatai yra vis tiek prasti. Todėl remiantis gautais duomenimis galima teigti, kad arba kintamieji įtraukti į lygtį nėra naudingi ir tik iškraipo duomenis, arba mašininio mokymosi parametrų imtis buvo per maža, kad modelis sugebėtų įsisavinti gilesnes sąsajas tarp kintamųjų. Galimi šios problemos sprendimo būdai būtų praplėsti kintamųjų paiešką, nustatyti griežtesnes kintamųjų atrinkimo ribas, praplėsti modelio parametrų paiešką.



18 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamačią LSTM modelį

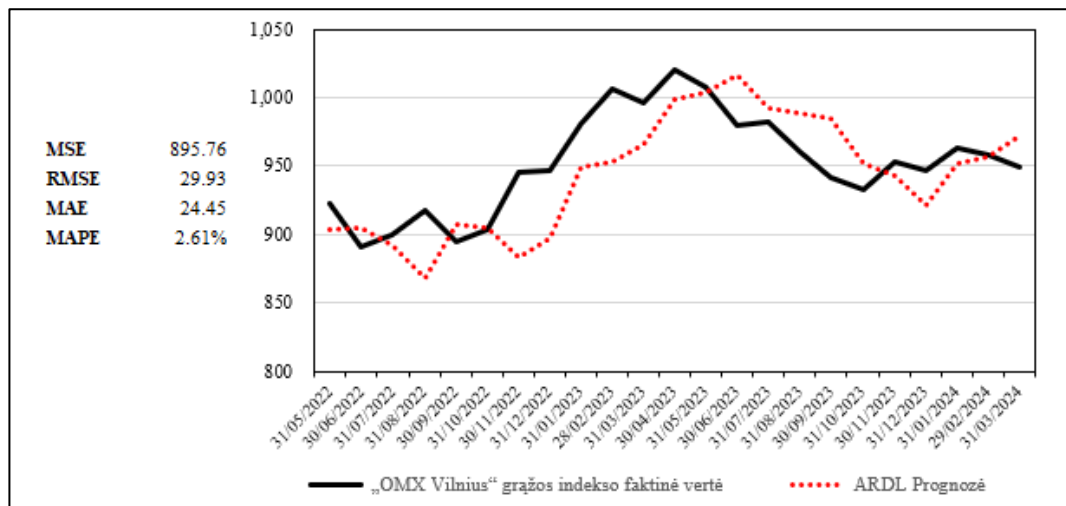
Vertinant GRU modelio rezultatus, kaip matyti 19 paveiksle, jie yra dar prastesni. Vertinant visus dieninių duomenų modelius, GRU modelis suteikė prasčiausius rezultatus. Net vizualiai vertinant linijinę diagramą yra matomi akivaizdūs nukrypimai, ypač prognozuojant paskutines vertes. Verta paminėti, kad visų modelių prognozės buvo padalintos į keturias dalis, tam, kad įvertinti, kaip modelis tiksliai prognozuoja laiko atžvilgiu. Tai buvo padaryta tam, kad įvertinti ar su laiku modelio rezultatai prastėja, dėl to, kad modelis nebūna atnaujinamas. Nors prieš tai vertintų modelių rezultatai nieko panašaus neindikavo, atsižvelgiant į šį modelį yra matoma modelio degradavimo tendencija. Kol pirmus 3 ketvirčius prognozuojamų duomenų modelio MAPE rezultatai buvo ~0.30%, paskutiniame laikotarpyje šis rodiklis padidėjo net iki ~0.45%. Atsižvelgiant į šiuos rezultatus iš dalies galima daryti prielaidą, kad laikui bėgant, dėl vienokių ar kitokių priežasčių šio modelio rezultatai palaipsniui prastėja. Žinoma, vienas iš paprastų sprendimo būdų būtų dažnesnis modelio atnaujinimas. Atsižvelgiant į gautus rezultatus galima prieiti prie panašios išvados kaip ir su LSTM ir ARDL modeliais – arba kintamieji per daug iškraipo rezultatus, arba išmokti kintamųjų ryšiai yra aktualūs tik treniravimosi metu, dėl to reikėtų koreguoti modelio parametrus.



19 pav. Dieninė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamačią GRU modelį

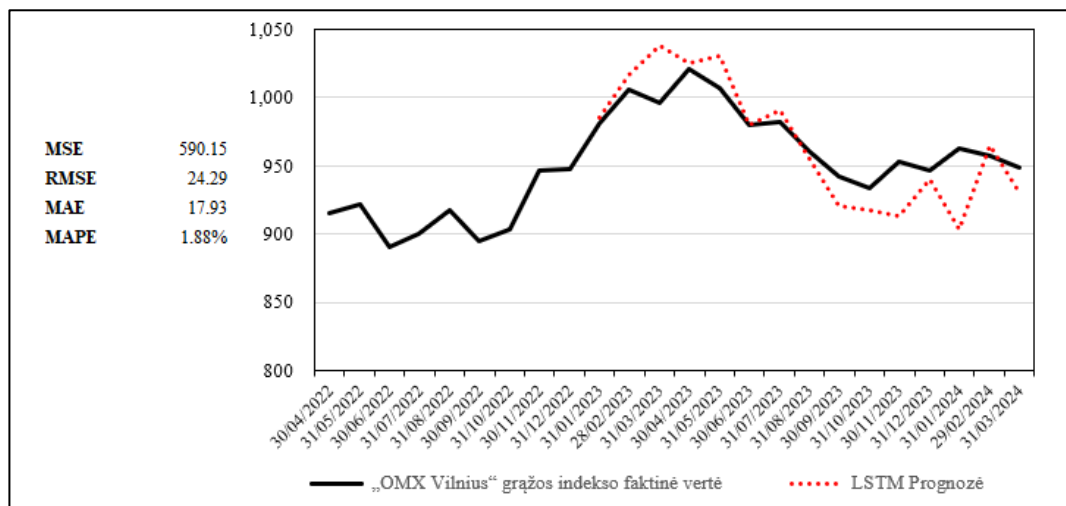
Apibendrinant galima teigti, kad daugiamačiai modeliai, kurie į „OMX Vilnius“ gražos indekso prognozavimą įtraukė „Invalda“, „DAX“ ir „CAC40“ nepasiteisino. Jie ne tik nepadidino prognozavimo tikslumo, o dar ir pasiekė prastesnius rezultatus, palyginti su vienmačiais metodais. ARDL modelis savo rezultatais pasirodė prasčiausiai, o vėliau taikyti LSTM ir GRU modeliai taip pat nepagerino vienmačių modelių rezultatų. Tai rodo arba įtrauktų kintamųjų neveiksmingumą, arba parametrų imčių ribotumą sudėtingiems kintamųjų tarpusavio ryšiams įvertinti.

Atsižvelgiant į mėnesinius duomenis, taip pat buvo atliktas ARDL modelis. Šio modelio parametrai tai pat buvo parinkti pasitelkiant geriausius AIC kriterijaus rezultatus naudojant „Eviews programą“. Remiantis gautais rezultatais tiksliausia pagal AIC kriterijų ARDL modelio struktūra yra „OMX Vilnius“ gražos indeksas 12; „AB Šiaulių bankas“ 1; „Invalda“ 6; „CAC40“ 2 arba dar kitaip – ARDL(12,1,6,2). Taigi, gavus šias vertes, pasitelkus Python buvo sudaryta prognozė 10% procentų naujausių mėnesinių duomenų. Ir kaip matyti iš 20 paveikslo, rezultatai yra labai prasti. Vien vizualiai pažvelgus į linijinę diagramą yra matoma, kad modelis prognozuoja prastai. Daugumoje laiko žymų prognozė yra vėluojanti. Atsižvelgiant į MAPE bei kitus tikslumą vertinančius rodiklius matoma, jog šis daugiamačis modelis taip pat pasirodė prasčiau nei vienmačių duomenų modeliai. Nors mėnesiniams duomenims prognozuoti buvo atrinkti šiek tiek kitokie kintamieji, kurie, lyginant su dieniniais duomenimis, turėjo netgi stipresnę koreliacinę ryšį, rezultatai rodo, jog visgi šie kintamieji vis tiek iškraipo prognozę. Todėl šie gauti rezultatai taip pat prisideda prie jau minėtų išvadų, kad arba kintamieji yra netinkami, arba šis modelis yra per prastas išvelgti sudėtingesnes judėjimo tendencijas.



20 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį ARDL(12,1,6,2) modelį

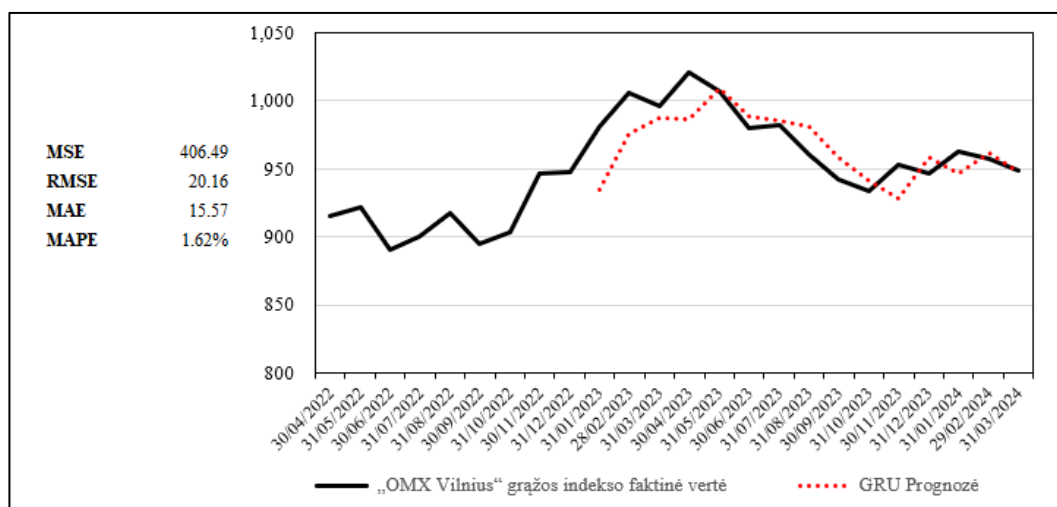
Prognozuojant mėnesines daugiamates „OMX Vilnius“ gražos indekso vertes pasitelkiant LSTM modelį, rezultatai yra kur kas geresni nei naudojant ARDL modelį (žr. 21 pav.). Nors ir buvo aplenkta ARDL modelis, ir vizualiai rezultatai atrodo arčiau tiesios, visgi šis modelis taip pat pasirodė prasčiau nei vienmačių modelių prognozės. Dėl šios priežasties prognozuoti pasitelkiant šį modelį nereikėtų. Gauti rezultatai taip pat sustiprina jau anksčiau diskutuotą išvadą, kad nauji kintamieji iškraipo rezultatus. Atsižvelgiant į tai, kad jau keli modeliai turi tokius pačius rezultatus, galima daryti išvadą, kad didžiaja dalimi dėl prastėjančių rezultatų yra kalti kintamieji, o ne modelio struktūra.



21 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį LSTM modelį

Galiausiai buvo atliktas paskutinis mašininis mokymusi paremtas prognozavimo modelis. Kaip matome iš 22 paveikslo, priešingai nei visi kiti nagrinėti daugiamatiai modeliai, ši GRU architektūra sugebėjo aplenkti naiviu metodu apskaičiuotos prognozės rezultatus. Tyrimo metu buvo rasti tik 3 modeliai, sugebantys aplenkti naivųjį metodą. Tai įrodo, kad yra įmanoma išgauti ganėtinai patikimas „OMX Vilniaus“ gražos indekso prognozes, pasitelkiant daugiamatę modelio struktūrą. Nors rezultatai buvo vis tiek prastesni nei ARIMA ar LSTM, šis modelis yra 3 pagal tikslumą. Taip pat atsižvelgiant į tai, kad nebuvo įvertinta labai didelė parametrų imtis, šiuos rezultatus gal net būtų

galima pagerinti. Apibendrinus galima teigti, kad naudojant nepriklausomus kintamuosius prognozuoti reliatyviai tiksliai „OMX Vilnius“ gražos indeksą yra įmanoma, tačiau rezultatai, bent jau šio tyrimo atveju, nebuvo geresni nei vienmačių duomenų modelių.



22 pav. Mėnesinė „OMX Vilnius“ gražos indekso vertės prognozė naudojant daugiamatį LSTM modelį

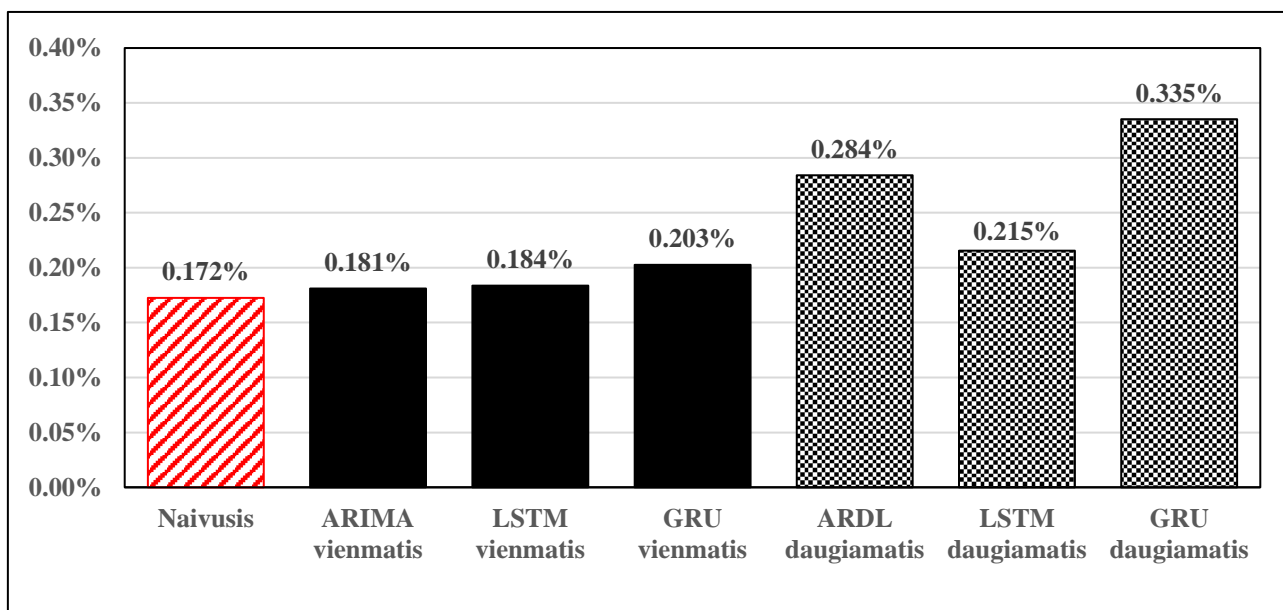
Apibendrinant, ARDL modelis, panašiai kaip ir su dieninėmis prognozėmis, nebuvo tikslus. Lyginant visus mėnesinius modelius, šio tyrimo metu ARDL pasirodė prasčiausiai. Pereinant prie kitų modelių, LSTM rezultatai buvo šiek tiek geresni, bet vis tiek prastesni, palyginti su vienmačiais metodais, tai patvirtina idėją, kad naujų kintamųjų įtraukimas tik sumažina prognozės tikslumą. Iš visų nagrinėtų daugiamatinių modelių tik GRU buvo geresnis už naivųjį metodą. Nors ir nepralenkdamas vienmačių SARIMAX ar LSTM modelių tikslumo, GRU modelis užėmė trečią vietą pagal tikslumą, įrodydamas, kad „OMX Vilnius“ gražos indekso prognozavimui galima naudoti daugiamatį modelį. Net ir įtraukus šiek tiek kitokius kintamuosius, turinčius stipresnes koreliacijas, daugiamatiai modeliai nepagerino prognozavimo tikslumo, lyginant su vienmačiais, o tai rodo arba netinkamą kintamųjų atranką, arba modelių silpnumą įvertinant sudėtingus judėjimus.

4.4. Modelių palyginimas ir tiksliausio modelio pritaikymas

Šio darbo metu, siekiant efektyviai prognozuoti „OMX Vilnius“ gražos indeksą, buvo atlikti keli tyrimai. Pirmiausia buvo išaiškinti kurie kintamieji tenkina „Granger“ priežastingumą, ir kurie iš šių kintamųjų koreliuoja tiek su dieniniais, tiek su mėnesiniais „OMX Vilnius“ gražos indekso duomenimis. Taip pat tyrimo metu remiantis AIC rezultatais buvo rasti tiksliausi ARIMA ir ARDL modeliai turimai duomenų imčiai. Tyrimo metu, pasitelkus tinklo paieškos įrankį, kiekvienam mašininio mokymosi modeliui taip pat buvo rasti geriausias rezultatus turimai imčiai išgaunantys parametrai. Galiausiai, atlikus prognozes tiek su dieniniais, tiek su mėnesiniais duomenimis, buvo gauti įvairių modelių rezultatai, pagal kuriuos galima nustatyti, kuris modelis yra tinkamas „OMX Vilnius“ gražos indekso prognozavimui, o kuris ne.

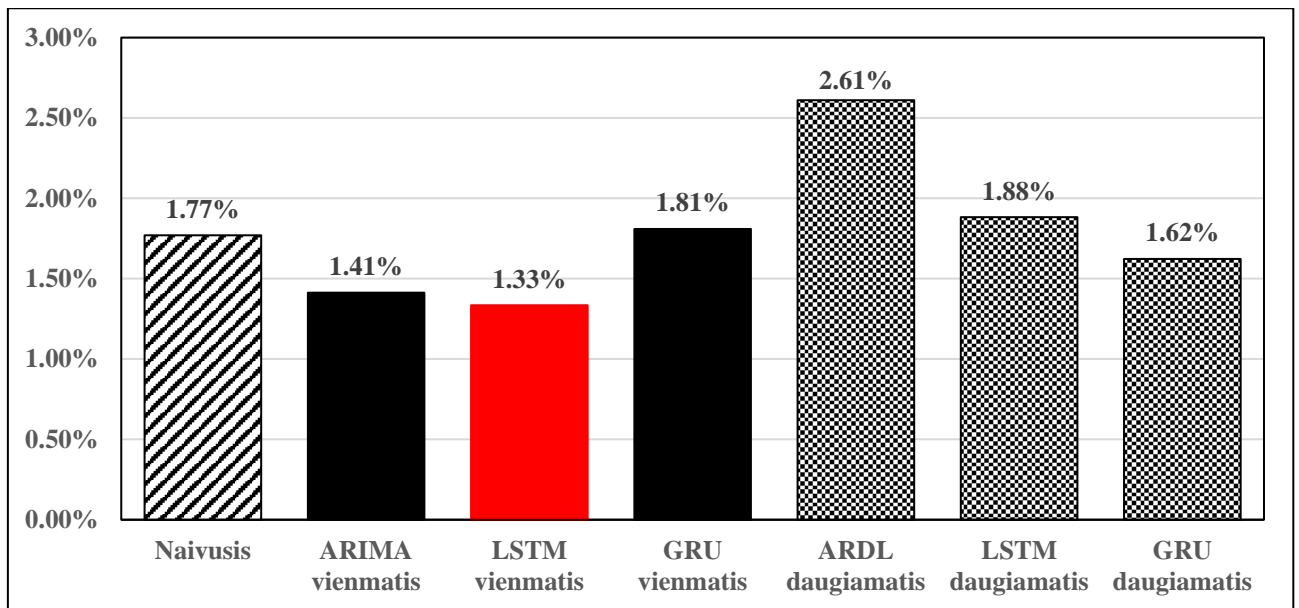
Atlikus dieninių duomenų prognozes gauti rezultatai nebuvo tokie, kokių tikėtasi. Pirmiausia, vienmačių duomenų modeliai nesugebėjo aplenksti įprasto naivaus modelio (žr. 23 pav.). Tai veda prie jau minėtos išvados, kad dieninė „OMX Vilnius“ gražos indekso vertė didžiąja dalimi juda nesiremiant praeities judėjimo tempais. Nepaisant to, daugiamatiai modeliai, nors turėdami daugiau informacijos, sugebėjo pasirodyti dar prasčiau nei vienmačiai. Todėl įvertinus šiuos modelius galimos tik 2 minėtos išvados, arba dieninių duomenų negalima prognozuoti pasitelkiant tiek istorinius „OMX

Vilnius“ gražos indekso judėjimo tempus, tiek naudotus kintamuosius, arba naudoti modeliai buvo per daug paprasti, kad sugebėtų tinkamai, ir tinkamais laiko momentais įvertinti visus nepriklausomų ir priklausomų kintamųjų istorinius judėjimus. Atsižvelgiant į tai, jog tyrimo metu tinklo paieškoje buvo įvertinti dieniniai vienmačiai ir daugiamačiai 864 LSTM ir 864 GRU modeliai, galima prieiti prie išvados, kad remiantis šiais gautais rezultatais, dieniniai „OMX Vilnius“ gražos indekso judėjimai yra per daug nenuspėjami, kad galima būtų išgauti tikslų prognozavimo modelį. Žinoma, nors ši tyrimo dalis ir nepasiekė norimų rezultatų, tai nebūtinai reiškia, kad jų nėra įmanoma pasiekti. Kitokie mašininio mokymosi modeliai, jų parametrai ar kitokia kintamųjų imtis ateities tyrimuose gali pagerinti šiuos rezultatus.



23 pav. Prognozavimo modelių MAPE rodiklio palyginimas prognozuojant dienes „OMX Vilnius“ gražos indekso vertes.

Atsižvelgiant į mėnesinių duomenų prognozes (žr. 24 pav.), rezultatai yra geresni. Pirmiausia matoma tai, kad tyrimo metu net 3 modeliai aplenkė naivaus prognozavimo metodo rezultatus, tai jau rodo, jog modeliai suteikia naudingas prognozes. Remiantis gautais rezultatais, viena iš svarbiausių išvalgų būtų ta, kad „OMX Vilnius“ gražos indekso mėnesines vertes galima ganėtinai tiksliai prognozuoti. Šie rezultatai yra priešingi nei gauti su dieninių duomenų prognozėmis. Remiantis šia išvalga yra tikėtina, kad kuo „OMX Vilnius“ laiko žyma yra mažesnė, tuo šio indekso judėjimai yra labiau nenuspėjami. Tuo tarpu kuo laiko žyma didesnė, tuo rezultatai geresni lyginant su naiviuoju metodu. Tam, kad pilnai įvertinti šį ryšį, ateityje būtų naudinga atlikti prognozes pasitelkiant ir ketvirtinius, ir metinius duomenis. Taip pat svarbi išvada yra ta, kad LSTM modelis sugebėjo aplenkėti klasikinio prognozavimo būdo rezultatus. O tai tik patvirtina nagrinėtą literatūrą ir įrodo, kad gilaus mašininio mokymosi LSTM modelis sugeba prognozuoti tiksliau nei klasikiniai metodai. Galiausiai, kaip ir su dieniniais duomenimis, taip ir su mėnesiniais, daugeliu atveju daugiamačiai modeliai pasirodė prasčiau nei vienmačiai. Todėl remiantis šiais gautais rezultatais galima daryti išvadą, kad tyrimo metu naudojami kintamieji, nors ir tenkino „Granger“ ir „Pearson“ koreliacijos tyrimus, vis dėlto yra nelabai tinkami „OMX Vilnius“ gražos indekso prognozavimui. Apibendrinant galima teigti, kad tyrimas išryškina geresnius mėnesinių duomenų prognozavimo rezultatus nei dieninių duomenų prognozavimas ir įrodo, kad gilaus mašininio mokymosi LSTM modelis sugeba prognozuoti tiksliau nei klasikiniai metodai.

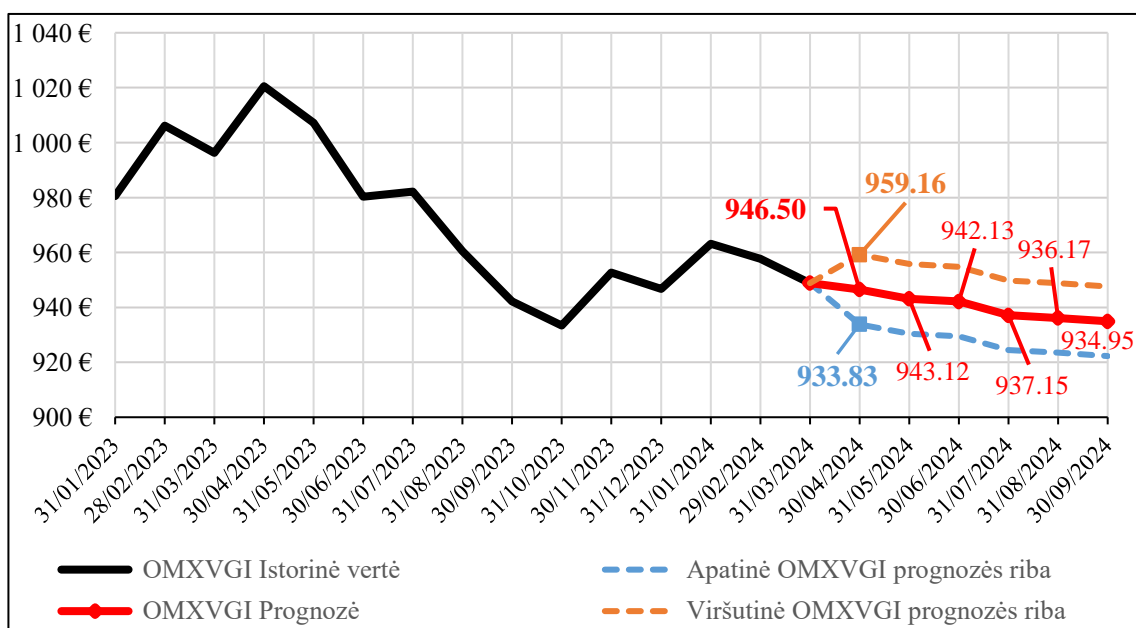


24 pav. Prognozavimo modelių MAPE rodiklio palyginimas prognozuojant mėnesines „OMX Vilnius“ gražos indekso vertes.

Įvertinus gautus rezultatus, taip pat yra svarbu aptarti tobulintinas vietas ateities tyrimams. Nors gauti rezultatai nėra blogi, autoriaus nuomone, juos dar tikrai galima gerinti pasiekiant dar tikslesnes „OMX Vilnius“ gražos indekso prognozes. Atsižvelgiant į dieninių duomenų prognozes, ateities tyrimams galima būtų pritaikyti pažangesnes mašininio mokymosi modelių parametrų paieškos priemones, tokias kaip genetiniai algoritmai, atsitiktinė paieška ar panašiai. Šie įrankiai savo greitesniais skaičiavimo algoritmais padėtų ištirti didesnę parametrų imtį, galbūt net rasti idealius modelio parametrus, maksimizuojančius tyrimo rezultatus. Taip pat šio tyrimo metu buvo koreguojami tik tam tikri mašininio mokymosi modelio parametrai, tačiau paiešką dar galima praplėsti įtraukiant skirtingus optimizacijos algoritmus bei jų mokymosi greitį, skirtingą neuronų tinklo struktūrą ar duomenų perdirbimo procesus. Visi šie veiksmai gali pagerinti modelio tikslumą. Atsižvelgiant į daugiamačius duomenis yra rekomenduojama atlikti platesnę kintamųjų atrinkimo analizę, galbūt atrenkant kintamuosius pagal griežtesnius parametrus. Tyrimo rezultatus taip pat galima būtų pagerinti įtraukus techninės analizės kintamuosius, tokius kaip santykinio stiprumo indeksą (angl. *relative strength index*, RSI), judėjimo vidurkius (angl. *moving averages*, MA), prekybos apimtį (angl. *trading volume*) ir t.t.. Taip pat šiais laikais yra vis dažniau pritaikomi kalbos modeliai, jų dėka tarp kintamųjų galima įtraukti naujienas iš socialinių ar žinių tinklalapių, tai galėtų padėti įvertinti bendrą nuomonę apie Vilniaus biržoje prekiaujamas akcijas, kas tiksliau padėtų prognozuoti „OMX Vilnius“ gražos indeksą. Nors atliktas tyrimas ir buvo platus, atsižvelgiant į tai, kad kintamųjų kiekis, darantis įtaką „OMX Vilnius“ gražos indeksui gali būti begalinis, šis ir ateities tyrimai dar galės būti daugybę kartų tobulinami.

Taigi, atlikus tiriamąją analizę, paskutinis svarbus žingsnis yra pritaikyti gautus rezultatus realioje situacijoje. Todėl, remiantis tiksliausiu gautu modeliu – mėnesiniu vienmačių duomenų LSTM modeliu, buvo sukurta prognozė ateinančiam pusmečiui (30/04/2024 – 30/09/2024) (žr. 25 pav.). Dar reiktų pabrėžti, kad atliekant prognozes LSTM ir kiti mašininio mokymosi modeliai prognozuoja ir geriausiai veikia nuspėjant tik 1 periodą į priekį. Tam, kad būtų galima padaryti pusmečio prognozę, tarp istorinių duomenų buvo įtraukti jau suprognozuoti periodai. Dėl šių priežasčių reikia suprasti, kad tiksliausiai modelis prognozuos tik balandžio mėnesį, po kurio modelį reiktų atnaujinti nauja

„OMX Vilnius“ gražos indekso vertė ir prognozuoti iš naujo tam, kad gauti tiksliausius rezultatus. Kaip matyti iš gautų rezultatų, remiantis šiuo modeliu yra prognozuojama, kad „OMX Vilnius“ gražos indekso vertė ateinančiais laikotarpiais palaipsniui kris. Taip pat remiantis testuojamų duomenų gautu standartiniu nuokrypiu buvo apskaičiuotos viršutinės ir apatinės nuokrypio vertės. Apibendrinus, gauti LSTM modelio rezultatai prognozuoja, jog balandžio mėn. gale OMX Vilniaus gražos indekso vertė bus ~946,50, atsižvelgiant į vidutinį standartinį testuojamų duomenų nuokrypį ši vertė gali svyruoti nuo 933,83 iki 959,16 EUR. Galiausiai, apibendrinant ateinančio pusmečio prognozę, remiantis gautais rezultatais yra prognozuojama, kad „OMX Vilnius“ gražos indekso vertė palaipsniui ir toliau kris.



25 pav. „OMX Vilnius“ gražos indekso mėnesinių duomenų prognozė 31/03/2024 - 30/09/2024 periodu naudojant LSTM modelį

Siekiant efektyviai prognozuoti „OMX Vilnius“ gražos indeksą, tyrimo metu buvo visapusiškai įvertintas tiek pats indeksas, tiek įvairūs kintamieji. Buvo atlikta preliminarinė stebinių analizė, duomenys unifikuoti, o reikšmingi kintamieji atrinkti pasitelkiant „Granger“ priežastingumą ir „Pearson“ koreliaciją. Naudojant „auto.arima“ ir „Eviews“ remiantis geriausiu AIC kriterijumi buvo atrastos geriausius rezultatus gaunanti ARIMA ir ARDL architektūros. Naudojant tinklo paiešką gilios mašininio mokymosi modeliams, buvo atrinkti geriausi modelio parametrai. Prognozuojant dienišius modelius tyrimas atskleidė tiek vienmačių, tiek daugiamačių modelių trūkumus, nesugebant pranokti naivaus metodo prognozavimo tikslumo. Remiantis šiais rezultatais galima teigti, kad arba tyrime naudojami modeliai yra per paprasti, tam, kad tiksliai išžvelgti subtilius indekso judėjimus, arba tikslus dienišius judėjimų prognozavimas yra labai sunkiai nuspėjamas, galbūt šiais laikais net nenuspėjamas. Iš kitos pusės, mėnesinės prognozės atskleidė daugiau perspektyvų – keli modeliai, ypač LSTM, pranoko naivųjį metodą, parodydamas savo pranašumą prieš klasikinius prognozavimo metodus. Šių rezultatų pokytis tarp dienišius ir mėnesinių prognozių tikslumo indikuoja, kad galimai ir toliau didinant laiko skalę į ketvirtines ar metines prognozes, tikslumas didėtų ir atvirksčiai. Galiausiai, remiantis tiksliausiu gautu vienmačiu mėnesiniu LSTM modeliu buvo atlikta reali pusmečio prognozė, kuri rodo, kad ateinančiu periodu „OMX Vilnius“ gražos indekso vertė palaipsniui kris. Atsižvelgiant į ateityje tobulintinas vietas, yra rekomenduojama taikyti pažangesnius parametrų optimizavimo metodus ir platesnę kintamųjų atrankos analizę, įskaitant techninius rodiklius ir naujienų analizę, pasitelkiant kalbos modelius. Apibendrinus, nors tyrimo metu ir

nepavyko išgauti dieninių modelių, kurie būtų tikslesni už naiviojo metodo prognozes, žvelgiant į mėnesinius modelius buvo rasti net 3, kurie gali būti naudojami „OMX Vilnius“ gražos indekso vertės prognozėms.

5. Išvados

1. Atliekant mokslinės problemos analizę, paaiškėjo, kad mašininis mokymasis tampa vis svarbesne šiuolaikinės analitikos technologija, kuri pastaraisiais metais ne tik labai išpopuliarėjo, tačiau ir tapo labiau prieinama dėl tobulėjančių technologijų. Nors mokslinių tyrimų šioje srityje daugėja, dėl didelės kintamųjų, modelių ir parametrų variacijos auga ir autorių rekomendacijų kiekis ateities tyrimams. Įvairūs šaltiniai rekomenduoja pritaikyti SVM, LSTM ir GRU modelius įvairiose pasaulio rinkose, įtraukti „bull“ ir „bear“ sertifikatus, panaudoti kalbos modelius socialinių tinklų sentimentų analizei, įvertinti skirtingos laiko žymos įtaką rezultatų tikslumui, galiausiai pritaikyti šiuos modernius neuronų tinklus ir pačiam „OMX Vilnius“ gražos indeksui prognozuoti. Atsižvelgiant į „OMX Vilnius“ gražos indeksą, jo reikšmingumas atsiranda iš to, kad tai yra bendra Vilniaus biržoje prekiaujamų akcijų reprezentacija. Šis indeksas parodo Lietuvoje viešai prekiaujamų įmonių vertės judėjimus, iš dalies apibendrinant Lietuvos akcijų ekonomikos pulsą. Dėl šios priežasties tikslios „OMX Vilnius“ gražos indekso prognozės yra labai svarbios priimant pagrįstus investicinius sprendimus ir strategiškai formuojant ekonominę politiką. Apibendrinus atlikta problemos analizė rodo, kad nors tyrimų kiekis šioje sferoje didėja, jis tikrai dar nėra pakankamas, kadangi dar yra daugybė ne iki galo ištirtų rinkų, modelių, jų architektūrų ar parametrų. Dėl šios priežasties verta ir toliau plėsti šią temą, pritaikant šias technologijas naujais būdais ir naujose rinkose.
2. Atliekant mokslinės literatūros analizę buvo gilintos žinios apie „OMX Vilnius“ gražos indeksą, apie klasikinius ir mašininio mokymosi grįstus prognozavimo modelius. Išanalizavus iš ko šis indeksas susideda ir kaip jis yra apskaičiuojamas, buvo įvertinta literatūra, naudojanti šį indeksą savo moksliniuose darbuose. „OMX Vilnius“ gražos indekso panaudojimas moksliniuose darbuose buvo ganėtinai platus, dažniausiai naudojamas kaip pagrindinis lyginamasis įvertinti bendrą Lietuvos akcijų rinkos būklę. Iš mašininio mokymosi pusės, atlikus literatūros analizę paaiškėjo, kad mašininis mokymasis susideda iš 2 pagrindinių dalių – duomenų parinkimo ir perdėbimo, bei mašininio mokymosi modelio parinkimo ir parametrų nustatymo. Įvertinus mašininio mokymosi tyrimus prognozuojant įvairias akcijų biržas buvo pastebėta, kad šie modeliai dažnai atvejais savo tikslumu aplenkia klasikinius prognozavimo metodus. Buvo pastebėta, kad LSTM modeliai ypač naudojami tokio pobūdžio tyrimams, o gauti rezultatai, dažnai atvejais, buvo tiksliausi. Atliekant tolimesnę analizę buvo įvertinti tiek klasikinių, tiek mašininio mokymosi modelių veikimo principai. Buvo pastebėta, kad RNN modeliai tyrimuose suteikia prasčiausius rezultatus dėl gradiento nykimo problemos. Dėl šios priežasties buvo sukurtos tokios atšakos kaip ilgalaikės-trumpalaikės atminties tinklų (LSTM) ir sklendžių rekurentinių vienetų GRU modeliai. Dėl šių modelių sudėtingesnės ląstelės struktūros, jie sugeba susitvarkyti su gradiento nykimo problema, kas padaro šiuos modelius labiau tinkamus laiko eilučių prognozavimui. Galiausiai buvo įvertinti gilaus mašininio mokymosi modelių parametrai, dėl kurių priklauso didžioji dalis prognozės tikslumo. Paaiškėjo, kad dėl galimos didelės parametrų variacijos yra rekomenduotina naudoti tokius paieškos ir optimizavimo įrankius, kaip tinklo paieška, atsitiktinė paieška ar genetiniai algoritmai. Apibendrinus, mašininio mokymosi algoritmai ir veikimo principai yra labai plati ir sudėtinga tema, tačiau remiantis nagrinėtais šaltiniais šie modeliai gali suteikti didesnio tikslumo prognozes nei klasikiniai prognozavimo modeliai. Atsižvelgiant į tai, kad tokie mašininio mokymosi modeliai kaip LSTM ir GRU dar nebuvo pritaikyti „OMX Vilnius“ gražos indeksui prognozuoti, yra

svarbu įvertinti šių modelių efektyvumą prognozuojant indeksą, kuris reprezentuoja Lietuvos akcijų rinką.

3. Atsižvelgiant į mokslinės literatūros analizę buvo sudaryta tyrimo metodologija. Tyrimui buvo naudojami 30/09/2004 – 31/03/2024 „OMX Vilnius“ gražos indekso ir 41 nepriklausomo kintamojo dieniniai ir mėnesiniai duomenys. Duomenų unifkavimui trūkstamos vertės buvo užpildytos senosiomis, o mastelio suvienodinimui buvo naudojamas Min-Max metodas. Galiausiai duomenys buvo padalinti į 90% treniravimosi ir 10% testavimo grupes. Tuomet kintamųjų vertinimui buvo pasirinkta naudoti „Granger“ priežastingumo ir „Pearson“ koreliacijos analizes, tam, kad įvertinti, kurie kintamieji yra reikšmingi. Tyrimo metu buvo pasirinkta naudoti ARIMA, ARDL, LSTM ir GRU modelius, kurių rezultatai buvo lyginami su naiviuoju metodu. Taip pat, tinkamiausių gilaus mašininio mokymosi parametrų radimui buvo naudojama tinklo paieška su nustatyta paieškos imtimi. Apibendrinus, tyrimo metu buvo tiriamas dieninių ir mėnesinių bei vienmačių ir daugiamačių duomenų modelių tikslumas.
4. Pirmiausia, atsižvelgiant į duomenų imtį, atlikus „Granger“ priežastingumą ir „Pearson“ koreliaciją, buvo nustatyta, kad dieniniams duomenims reikšmingi kintamieji yra „Invalda INVL“, „DAX“, „CAC40“, o mėnesiniams – „AB Šiaulių Bankas“, „Invalda INVL“, „CAC40“. Atsižvelgiant į šiuos rezultatus galima pastebėti, kad iš 41 vertinto kintamojo, tik 4 buvo reikšmingi „OMX Vilnius“ gražos indekso prognozavimui. Todėl galima daryti išvadą, kad didelę kapitalizaciją turinčios akcijos, žaliavos, valiutų kursai ar net Euro zonos obligacijų pajamingumai turi mažą įtaką „OMX Vilnius“ gražos indekso būsimai vertei. Vieninteliai reikšmingi kintamieji, kurie daro įtaką šiam indeksui, yra Lietuvos įmonių akcijos ir keli tarptautiniai indeksai. Turint visus duomenis buvo atlikta tinklo paieška, kurios dėka buvo rasti tiksliausi, iš duotos imties, modelių parametrai. (žr. 12 lentelę). Atlikus dieninių duomenų tyrimą paaiškėjo, kad tiek vienmačiai, tiek daugiamačiai modeliai nebuvo geresni net už naiviuoju metodu apskaičiuotas prognozes. Taip pat daugiamačiai modeliai, nors turėdami daugiau informacijos, sugebėjo pasirodyti dar prasčiau nei vienmačiai. Atsižvelgiant į tai, kad dieninių duomenų modelių prognozės buvo tokios prastos, galimos kelios išvados. Pirmoji, kad dieninių duomenų analizuojamame laikotarpyje neįmanoma tiksliai prognozuoti pasitelkiant tiek istorinius „OMX Vilnius“ gražos indekso judėjimus, tiek naudotus kintamuosius, kadangi dieniniai judėjimai galimai yra per daug nenuspėjami. Antroji, kad naudoti modeliai buvo per daug paprasti, jog sugebėtų tinkamai ir tinkamais laiko momentais įvertinti visus nepriklausomų ir priklausomų kintamųjų istorinius judėjimus tiksliai „OMX Vilnius“ gražos indekso prognozavimui. Atsižvelgiant į mėnesinių duomenų prognozes, rezultatai yra geresni. Tyrimo metu net 3 modeliai aplenkė naiviojo prognozavimo metodo rezultatus. Tai rodo, kad istoriniuose mėnesiniuose judėjimuose yra naudingesnės informacijos nei dieniniuose. Kita svarbi pasiekta išvada, kad remiantis tiek vienmačiais, tiek daugiamačiais duomenimis, mėnesiniai mašininio mokymosi modeliai sugeba prognozuoti tiksliau nei klasikiniai modeliai. Taip pat, kaip ir su dieniniais duomenimis, taip ir su mėnesiniais, daugiamačiai modeliai vidutiniškai pasirodė prasčiau nei vienmačiai. Tai veda prie išvados, kad nors tyrimo metu „Granger“ priežastingumas ir „Pearson“ koreliacija buvo tenkinamos, naudojami kintamieji vis dėlto yra nelabai tinkami „OMX Vilnius“ gražos indekso prognozavimui arba naudojami modeliai buvo per daug paprasti, kad sugebėtų išvelgti tendencijas, darančias įtaką indekso ateities vertėms. Apibendrinus, giliuoju mašininio mokymosi grįsti prognozavimo modeliai,

prognozuojant „OMX Vilnius“ gražos indeksą gali būti tikslesni nei klasikiniai metodai, kuomet yra naudojamos mėnesinės indekso vertės.

5. Galiausiai, remiantis tiksliausiu gautu modeliu, mėnesiniu vienmačių duomenų LSTM modeliu, buvo sukurta prognozė ateinančiam pusmečiui (30/04/2024 – 30/09/2024). Verta paminėti, kad mašininio mokymosi modeliai geriausiai veikia nuspėjant tik 1 periodą į priekį. Dėl šios priežasties tiksliausia gauta prognozė, paremta tik faktiniais duomenimis, yra balandžio mėn. galui, kol kiti būsimi periodai yra paremti ir buvusiomis prognozėmis. Remiantis tiksliausiu rastu gilaus mašininio mokymosi LSTM modelio rezultatais prognozuojama, kad balandžio pabaigoje „OMX Vilnius“ gražos indekso vertė bus ~946,50 EUR, taip pat, atsižvelgiant į vidutinį standartinį nuokrypį ši vertė gali varijuoti nuo 933,83 EUR iki 959,16 EUR. Apibendrinus, ilgajame laikotarpyje yra prognozuojama, kad „OMX Vilnius“ gražos indekso vertė palaipsniui ir toliau kris. Šis indeksas yra svarbus investuotojams ir politikams, nes yra regiono ekonominio stabilumo rodiklis. Ekonominiu požiūriu šis prognozuojamas sumažėjimas gali rodyti tam tikrą investuotojų suinteresuotumo mažėjimą ar neigiamų makroekonominių sąlygų numatymą. Politiniu požiūriu toks akcijų vertės kritimas gali skatinti ir tam tikrus fiskalinius pokyčius. Dėl šių priežasčių suinteresuotosioms šalims rekomenduojama atidžiai stebėti „OMX Vilnius“ gražos indekso tendencijas, nes jos gali turėti didelį ilgalaikį poveikį ekonominės strategijos ir regiono politikos formavimui.

Literatūros sąrašas

1. Aleksaitė, E., (2021). Rinkos modelių tinkamumo akcijų grąžai prognozuoti ir nuosavo kapitalo kainai nustatyti vertinimas Lietuvos akcijų rinkoje, Vytauto Didžiojo universitetas. <https://doi.org/10.7220/2538-6778.2021.24>
2. Balabonienė, I., Blikiėnė, R., ir Stundžienė, A. (2013). Ekonometrija. Praktinis regresijos ir laiko eilučių modelių taikymas. KTU leidykla „Technologija“. <https://doi.org/10.5755/e01.9786090210185>
3. Baranauskas, S. (2010). Portfolio formation and management according to macroeconomic indicators influence on OMXV, *Business: Theory and Practice* 11(3): 286–293.
4. Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(7), 1636. <https://doi.org/10.3390/en11071636>
5. Brownlee, J., (2020) How to Use StandardScaler and MinMaxScaler Transforms in Python. *Machine Learning Mastery*. [žiūrėta 2024-03-28]. Prieiga per internetą: <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/>
6. Cerqueira, V., Torgo, L., ir Soares, C. (2022). A case study comparing machine learning with statistical methods for time series forecasting: Size matters. *Journal of Intelligent Information Systems*, 59(2), 415–433. <https://doi.org/10.1007/s10844-022-00713-9>
7. Chan, F., ir Mátyás, L. (Eds.). (2022). *Econometrics with machine learning* (Vol. 53). Springer International Publishing. <https://doi.org/10.1007/978-3-031-15149-1>
8. Dautel, A. J., Härdle, W. K., Lessmann, S., ir Seow, H.-V. (2020). Forex exchange rate forecasting using deep recurrent neural networks. *Digital Finance*, 2(1–2), 69–96. <https://doi.org/10.1007/s42521-020-00019-x>
9. Džikevičius, A., ir Stabužytė, N. (2012). Forecasting omx vilnius stock index – a neural network approach. *Verslas: Teorija Ir Praktika*, 13(4), 324–332. <https://doi.org/10.3846/btp.2012.34>
10. EViews (2017). *Autoregressive distributed lag (Ardl) estimation. Part 1—Theory*. [žiūrėta 2024-03-02]. Prieiga per internetą: <https://blog.eviews.com/2017/04/autoregressive-distributed-lag-ardl.html>
11. Ferreira, I. H., ir Medeiros, M. C. (2021). Modeling and forecasting intraday market returns: A machine learning approach. <https://doi.org/10.48550/ARXIV.2112.15108>
12. Hyndman, R. J., ir Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd edition).
13. Jaya, J. (2023). *Machine learning for financial market forecasting*. Master's thesis, Harvard University Division of Continuing Education. <https://dash.harvard.edu/handle/1/37375052>
14. Karpuškienė, V., ir Lastauskas, P. (2012). Ekonometrinis modeliavimas su EViews: praktinis gidas. Vilniaus Universitetas. Ekonomikos fakultetas. Versija 10-03.
15. Karsoliya, S. (2012). Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology*, Volume 3, Issue 6, ISSN: 2231-5381.
16. Kumbure, M. M., Lohrmann, C., Luukka, P., ir Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, 116659. <https://doi.org/10.1016/j.eswa.2022.116659>
17. Laplace, P. S. (1902). *A philosophical essay on probabilities*. New York : J. Wiley ; London : Chapman & Hall. <http://archive.org/details/philosophicaless00lapliala>

18. Li, Q., Kamaruddin, N., Yuhaniz, S. S., ir Al-Jaifi, H. A. A. (2024). Forecasting stock prices changes using long-short term memory neural network with symbolic genetic programming. *Scientific Reports*, 14(1), 422. <https://doi.org/10.1038/s41598-023-50783-0>
19. Liashchynskiy, P., & Liashchynskiy, P. (2019). Grid search, random search, genetic algorithm: A big comparison for nas. <https://doi.org/10.48550/ARXIV.1912.06059>
20. Madziwa, L., Pillalamarry, M., ir Chatterjee, S. (2022). Gold price forecasting using multivariate stochastic model. *Resources Policy*, 76, 102544. <https://doi.org/10.1016/j.resourpol.2021.102544>
21. Maity, R. (2021). Min max scaler. *Medium*. <https://medium.com/@ranjitmaity95/min-max-scaler-b2411ab3136d>
22. Maziarz, M. (2015). A review of the Granger-causality fallacy. *The Journal of Philosophical Economics : Reflections on Economic and Social Issues*, VIII.(2), 86–105. Prieiga per internetą: <https://hrcak.srce.hr/155919>
23. Nasdaq (n.d.a). *Overview for omxvgi* [žiūrėta 2024-03-10]. Prieiga per internetą: <https://indexes.nasdaqomx.com/Index/Overview/OMXVGI>
24. NasdaqBaltic (2016a). *Nasdaq OMX Baltic equity indexes*. [žiūrėta 2024-03-10]. Prieiga per internetą: https://nasdaqbaltic.com/wp-content/uploads/page/about-indexes/Methodology_OMXBALTIC.pdf
25. NasdaqBaltic (2016b). *Nasdaq OMX all-share, benchmark and sector indexes*. [žiūrėta 2024-03-10]. Prieiga per internetą: <https://nasdaqbaltic.com/wp-content/uploads/page/about-indexes/Rules-of-the-Construction-and-Maintenance-of-the-Nasdaq-indexes.pdf>
26. NasdaqBaltic (2024). *Baltic market indexes*. [žiūrėta 2024-04-03]. Prieiga per internetą: <https://nasdaqbaltic.com/statistics/en/charts>
27. NasdaqBaltic (n.d.). *Apie indeksus*. [žiūrėta 2024-03-10]. Prieiga per internetą: <https://nasdaqbaltic.com/lt/rinkos-informacija/apie-indeksus/>
28. NasdaqOMXNordic (2024). *OMXVGI, OMX VILNIUS_GI, (LT0000999963)*. [žiūrėta 2024-04-10]. Prieiga per internetą: https://www.nasdaqomxnordic.com/index/index_info?Instrument=LT0000999963
29. Nevasalmi, L. (2020). Forecasting multinomial stock returns using machine learning methods. *The Journal of Finance and Data Science*, 6, 86–106. <https://doi.org/10.1016/j.jfds.2020.09.001>
30. Ngoc, T. T., Dai, L. V., ir Phuc, D. T. (2021). Grid search of multilayer perceptron based on the walk-forward validation methodology. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(2), 1742. <https://doi.org/10.11591/ijece.v11i2.pp1742-1751>
31. Olah, C. (2015). Understanding LSTM Networks, Colah's blog [žiūrėta 2024-03-28]. Prieiga per internetą: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
32. Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., ir Fareed, M. (2022). Stock market forecasting using the random forest and deep neural network models before and during the covid-19 period. *Frontiers in Environmental Science*, 10, 917047. <https://doi.org/10.3389/fenvs.2022.917047>
33. Pilinkus, D. (2009). Stock market and macroeconomic variables: evidences from Lithuania, *Economics & Management* 14: 884–891.
34. S&P Dow Jones Indices. (2024). *S&P U.S. Indices Methodology*. [žiūrėta 2024-04-05]. Prieiga per internetą: <https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf>

35. Saud, A. S., ir Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia Computer Science*, 167, 788–798. <https://doi.org/10.1016/j.procs.2020.03.419>
36. Seabe, P. L., Moutsinga, C. R. B., ir Pindza, E. (2023). Forecasting cryptocurrency prices using lstm, gru, and bi-directional lstm: A deep learning approach. *Fractal and Fractional*, 7(2), 203. <https://doi.org/10.3390/fractalfract7020203>
37. Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia Computer Science*, 131, 895–903. <https://doi.org/10.1016/j.procs.2018.04.298>
38. Shmueli, G., ir Lichtendahl, K. C. (2018). *Practical time series forecasting with R: A hands-on guide (Second edition.)*. Axelrod Schnall Publishers.
39. Tam, A., (2023). LSTM for Time Series Prediction in PyTorch. *Machine Learning Mastery*. <https://machinelearningmastery.com/lstm-for-time-series-prediction-in-pytorch/>
40. Teresienė, D. (2009). Lithuanian stock market analysis using a set of garch models. *Journal of Business Economics and Management*, 10(4), 349–360. <https://doi.org/10.3846/1611-1699.2009.10.349-360>
41. Tvaronavičienė, M.; Michailova, J. (2006). Factors affecting securities prices: theoretical versus practical approach, *Journal of Business Economics and Management* 7(4): 213–222.
42. Valakevicius, E., ir Vaznelyte, K. (2012). Development, implementation and evaluation of multistage investment strategies. *Engineering Economics*, 23(2), 144–153. <https://doi.org/10.5755/j01.ee.23.2.1542>
43. Xiao, R., Feng, Y., Yan, L., ir Ma, Y. (2022). Predict stock prices with ARIMA and LSTM. <https://doi.org/10.48550/ARXIV.2209.02407>
44. Zou, J., Zhao, Q., Jiao, Y., Cao, H., Liu, Y., Yan, Q., Abbasnejad, E., Liu, L., ir Shi, J. Q. (2022). Stock market prediction via deep learning techniques: A survey. <https://doi.org/10.48550/ARXIV.2212.12717>

Priedai

1 priedas. Dieninių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu

„Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu	(F-Statistika/P-Tikimybė)									
	1 vėl.	2 vėl.	3 vėl.	4 vėl.	5 vėl.	6 vėl.	7 vėl.	8 vėl.	9 vėl.	10 vėl.
Telia Lietuva	0.0509 / 0.8215	0.053 / 0.9738	2.1102 / 0.5498	2.1801 / 0.7027	2.3984 / 0.7917	2.5673 / 0.8609	4.3265 / 0.7415	4.7634 / 0.7825	13.0255 / 0.1615	13.9451 / 0.1755
AB Siaulių Bankas	1.2437 / 0.2648	1.8133 / 0.4039	2.984 / 0.3941	7.8464 / 0.0974	8.1096 / 0.1503	9.5107 / 0.1468	9.6839 / 0.2072	13.9015 / 0.0844	14.0687 / 0.1199	15.4393 / 0.1168
APB Apranga	4.4733 / 0.0344*	4.427 / 0.1093	4.4043 / 0.221	12.3238 / 0.0151*	13.5751 / 0.0185*	13.2849 / 0.0387*	13.285 / 0.0655	15.0245 / 0.0587	16.9568 / 0.0494*	18.0619 / 0.0539
Grigeo AB	0.0234 / 0.8785	0.1689 / 0.919	0.5911 / 0.8985	0.5371 / 0.9698	0.6136 / 0.9874	0.6593 / 0.9953	1.0327 / 0.9943	3.0851 / 0.9289	8.0271 / 0.5314	7.9828 / 0.6305
Invalda INVL	3.2027 / 0.0735	5.2325 / 0.0731	5.7554 / 0.1241	6.989 / 0.1365	9.1914 / 0.1017	9.2708 / 0.1589	10.3829 / 0.1679	17.702 / 0.0236*	19.0329 / 0.0249*	20.5145 / 0.0247*
AS Latvijas Gaze	0.4736 / 0.4913	0.583 / 0.7471	5.8341 / 0.12	6.3429 / 0.175	6.9291 / 0.226	7.8619 / 0.2484	9.8173 / 0.1992	9.5774 / 0.2959	9.6595 / 0.3788	12.6359 / 0.2447
AS Amber Latvijas balzams	0.0238 / 0.8775	2.3311 / 0.3117	3.9788 / 0.2638	5.5069 / 0.2391	8.2934 / 0.1408	12.3424 / 0.0548	12.5351 / 0.0843	13.0122 / 0.1114	13.4936 / 0.1415	14.2278 / 0.1629
SAF Tehnika A/S	1.0088 / 0.3152	1.8767 / 0.3913	15.347 / 0.0015*	18.0756 / 0.0012*	28.0452 / 0*	27.5989 / 0.0001*	29.9773 / 0.0001*	34.1497 / 0*	35.2765 / 0.0001*	35.1283 / 0.0001*
AS Latvijas Juras medicinas centrs	0.0006 / 0.981	1.0662 / 0.5868	1.1359 / 0.7684	2.9935 / 0.5589	3.1572 / 0.6758	3.5156 / 0.7419	17.2331 / 0.016*	18.5235 / 0.0176*	19.2397 / 0.0232*	19.22 / 0.0376*
AS Siguldas Ciltslietu Un Makslīgas Apsklosanas Stacija	5.7614 / 0.0164*	6.1358 / 0.0465*	6.1486 / 0.1046	6.4972 / 0.165	8.7103 / 0.1212	8.9634 / 0.1756	9.1531 / 0.2419	10.3051 / 0.2443	10.6252 / 0.3023	15.6292 / 0.1108
TKM Grupp AS	1.183 / 0.2768	5.2167 / 0.0737	10.0717 / 0.018*	10.1574 / 0.0379*	18.9468 / 0.002*	20.8035 / 0.002*	22.2251 / 0.0023*	22.9263 / 0.0035*	23.0777 / 0.006*	24.845 / 0.0056*
AS Merko Ehitus	0.7003 / 0.4027	6.0506 / 0.0485*	8.9651 / 0.0298*	9.9696 / 0.0409*	10.3605 / 0.0656	10.2757 / 0.1135	11.6411 / 0.113	11.3591 / 0.1822	12.4639 / 0.1884	15.5396 / 0.1136
AS Silvano Fashion Group	21.6367 / 0*	31.0706 / 0*	38.713 / 0*	62.299 / 0*	61.6019 / 0*	63.9889 / 0*	80.5669 / 0*	82.6579 / 0*	93.3088 / 0*	95.2931 / 0*
Nordic Fibreboard AS	0.9866 / 0.3206	3.1769 / 0.2042	5.2368 / 0.1553	5.191 / 0.2683	5.2457 / 0.3866	13.0612 / 0.0421*	19.4897 / 0.0068*	20.1511 / 0.0098*	20.332 / 0.016*	22.3651 / 0.0133*
Novo Nordisk	15.817 / 0.0001*	16.4065 / 0.0003*	16.7823 / 0.0008*	18.7421 / 0.0009*	18.8599 / 0.002*	20.3902 / 0.0024*	20.5569 / 0.0045*	22.3251 / 0.0043*	22.4018 / 0.0077*	24.2131 / 0.0071*
AstraZeneca PLC	26.7875 / 0*	34.2666 / 0*	40.4253 / 0*	41.7278 / 0*	43.1615 / 0*	46.3406 / 0*	46.4132 / 0*	49.1275 / 0*	49.0758 / 0*	49.1388 / 0*
Equinor ASA	35.8962 / 0*	36.4519 / 0*	42.1934 / 0*	42.9307 / 0*	44.9494 / 0*	47.4492 / 0*	48.3774 / 0*	49.2173 / 0*	58.985 / 0*	67.0453 / 0*
Investor AB	13.9936 / 0.0002*	15.2209 / 0.0005*	46.0115 / 0*	47.6369 / 0*	49.0684 / 0*	51.1687 / 0*	62.5118 / 0*	65.6313 / 0*	66.0195 / 0*	68.9674 / 0*
AB Volvo (publ)	20.3502 / 0*	20.6433 / 0*	31.3414 / 0*	34.425 / 0*	34.9744 / 0*	35.9159 / 0*	45.7857 / 0*	50.1195 / 0*	50.1748 / 0*	50.4168 / 0*
Microsoft Corporation	58.2544 / 0*	66.5453 / 0*	71.1945 / 0*	71.6714 / 0*	70.6049 / 0*	71.8618 / 0*	77.5403 / 0*	80.3483 / 0*	81.4207 / 0*	88.8486 / 0*
Apple Inc.	46.3347 / 0*	50.1798 / 0*	55.0557 / 0*	55.6914 / 0*	54.8802 / 0*	55.7995 / 0*	59.2554 / 0*	61.0543 / 0*	61.1267 / 0*	62.8773 / 0*
NVIDIA Corporation	25.131 / 0*	27.5894 / 0*	30.1017 / 0*	30.3926 / 0*	30.3934 / 0*	33.2524 / 0*	33.3758 / 0*	34.8677 / 0*	35.4583 / 0*	35.8178 / 0.0001*
Alphabet Inc.	63.8879 / 0*	68.2344 / 0*	82.7938 / 0*	83.3922 / 0*	83.3707 / 0*	85.2157 / 0*	100.083 / 0*	103.9176 / 0*	104.6524 / 0*	107.1863 / 0*
Amazon.com, Inc.	46.1812 / 0*	46.9357 / 0*	56.5114 / 0*	58.2527 / 0*	57.844 / 0*	59.3258 / 0*	71.0219 / 0*	72.2291 / 0*	72.3025 / 0*	74.914 / 0*
Gold	3.0102 / 0.0827	2.9683 / 0.2267	4.3427 / 0.2268	5.1881 / 0.2685	14.8821 / 0.0109*	15.0018 / 0.0202*	15.0388 / 0.0355*	14.9026 / 0.0611	17.3386 / 0.0437*	19.1996 / 0.0378*
Copper	8.4101 / 0.0037*	8.9344 / 0.0115*	10.0325 / 0.0183*	10.3752 / 0.0346*	11.2465 / 0.0467*	19.1827 / 0.0039*	26.2533 / 0.0005*	26.832 / 0.0008*	27.2611 / 0.0013*	30.3842 / 0.0007*
Crude Oil	16.3886 / 0.0001*	18.274 / 0.0001*	18.1653 / 0.0004*	20.0673 / 0.0005*	19.5972 / 0.0015*	26.283 / 0.0002*	28.1625 / 0.0002*	28.7651 / 0.0003*	35.4585 / 0*	37.1403 / 0.0001*
Natural Gas	7.1228 / 0.0076*	7.1988 / 0.0273*	8.1881 / 0.0423*	8.4472 / 0.0765	13.092 / 0.0225*	18.0249 / 0.0062*	18.1671 / 0.0112*	19.157 / 0.014*	19.3399 / 0.0225*	19.3543 / 0.036*
Coffee	15.0109 / 0.0001*	15.2159 / 0.0005*	15.9591 / 0.0012*	16.2447 / 0.0027*	16.9117 / 0.0047*	17.18 / 0.0086*	17.1939 / 0.0162*	17.6416 / 0.0241*	17.6049 / 0.04*	18.6981 / 0.0443*
USD/EUR	7.0811 / 0.0078*	7.0495 / 0.0295*	7.5236 / 0.057	7.4125 / 0.1156	10.4839 / 0.0626	13.7042 / 0.0331*	18.5156 / 0.0098*	21.033 / 0.0071*	21.0893 / 0.0123*	21.1752 / 0.0199*
USD/GBP	2.3629 / 0.1243	2.2518 / 0.3244	2.7161 / 0.4375	3.4884 / 0.4796	3.5972 / 0.6087	3.6524 / 0.7236	10.1617 / 0.1796	9.0992 / 0.334	11.0068 / 0.2752	13.9153 / 0.1769
USD/JPY	42.5017 / 0*	43.6199 / 0*	47.3201 / 0*	53.0541 / 0*	61.3995 / 0*	61.9174 / 0*	62.0829 / 0*	64.0149 / 0*	62.5689 / 0*	64.7449 / 0*
USD/CHF	0.1744 / 0.6762	0.2148 / 0.8982	3.6422 / 0.3028	3.8067 / 0.4328	6.4498 / 0.2649	6.3125 / 0.3891	7.9798 / 0.3344	10.2565 / 0.2475	10.8819 / 0.2839	12.6786 / 0.2422
S&P 500	170.1283 / 0*	174.6704 / 0*	190.2114 / 0*	193.0182 / 0*	189.9547 / 0*	193.169 / 0*	210.4728 / 0*	214.5017 / 0*	221.4883 / 0*	232.4419 / 0*
DAX	27.6947 / 0*	28.9531 / 0*	51.746 / 0*	52.852 / 0*	54.5007 / 0*	58.2644 / 0*	70.9004 / 0*	73.2744 / 0*	73.6588 / 0*	73.9306 / 0*
CAC40	45.794 / 0*	47.9468 / 0*	60.397 / 0*	61.3984 / 0*	65.7645 / 0*	70.6762 / 0*	78.5095 / 0*	81.1002 / 0*	81.995 / 0*	81.6824 / 0*
Maturity: 1 year	0.3371 / 0.5615	7.7606 / 0.0206*	7.1765 / 0.0665	10.8571 / 0.0282*	11.1138 / 0.0492*	12.3951 / 0.0537	12.4966 / 0.0854	12.5738 / 0.1274	13.2253 / 0.1527	14.0351 / 0.1714
Maturity: 3 years	0.0025 / 0.9599	3.3324 / 0.189	2.7689 / 0.4286	6.6709 / 0.1543	7.2247 / 0.2045	11.8597 / 0.0652	12.4104 / 0.0878	12.951 / 0.1135	12.884 / 0.1679	14.036 / 0.1714
Maturity: 5 years	0.0012 / 0.9725	1.9655 / 0.3743	1.4273 / 0.6992	6.1832 / 0.1859	7.0286 / 0.2185	9.3364 / 0.1555	11.7995 / 0.1073	11.911 / 0.1552	11.8298 / 0.2231	12.7526 / 0.2378
Maturity: 10 years	0.4084 / 0.5228	2.1759 / 0.3369	1.488 / 0.685	5.9167 / 0.2055	7.0205 / 0.2191	6.9975 / 0.3211	11.0241 / 0.1376	10.7354 / 0.2171	11.3477 / 0.2526	11.927 / 0.29
Maturity: 30 years	0.7165 / 0.3973	2.4013 / 0.301	8.5432 / 0.036*	10.1663 / 0.0377*	11.3896 / 0.0442*	14.1601 / 0.0279*	14.8096 / 0.0385*	15.311 / 0.0534	16.3595 / 0.0597	16.6346 / 0.0828

2 priedas. Mėnesinių duomenų „Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu

„Granger“ priežastingumo tyrimas lyginant su „OMX Vilnius“ gražos indeksu	(F-Statistika/P-Tikimybė)									
	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.	1 vėl.
Telia Lietuva	1.0014 / 0.317	1.2829 / 0.5265	1.7879 / 0.6176	2.0032 / 0.7352	2.0539 / 0.8416	2.5309 / 0.865	2.8739 / 0.8964	3.0882 / 0.9287	4.1616 / 0.9005	4.9009 / 0.8977
AB Siaulių Bankas	9.5137 / 0.002*	9.636 / 0.0081*	8.991 / 0.0294*	8.795 / 0.0664	10.1996 / 0.0698	10.8257 / 0.0939	10.8016 / 0.1475	10.2488 / 0.248	12.6796 / 0.1776	14.4409 / 0.1538
APB Apranga	2.8093 / 0.0937	3.3323 / 0.189	3.0758 / 0.3801	3.0444 / 0.5504	5.8583 / 0.3203	5.5528 / 0.4751	9.0597 / 0.2484	9.8777 / 0.2737	11.7058 / 0.2304	15.177 / 0.1257
Grigeo AB	5.2027 / 0.0226*	5.2631 / 0.072	6.2456 / 0.1003	6.5044 / 0.1645	7.441 / 0.1899	8.4469 / 0.2072	8.2797 / 0.3086	8.1595 / 0.418	8.5566 / 0.4792	10.2583 / 0.4181
Invalda INVL	5.0754 / 0.0243*	7.4833 / 0.0237*	9.4261 / 0.0241*	12.615 / 0.0133*	16.1454 / 0.0064*	15.8755 / 0.0144*	16.4771 / 0.0211*	16.9444 / 0.0307*	19.0521 / 0.0248*	23.7459 / 0.0083*
AS Latvijas Gaze	0.1723 / 0.6781	0.166 / 0.9203	0.4231 / 0.9354	0.5309 / 0.9704	0.57 / 0.9893	0.5842 / 0.9967	0.6864 / 0.9984	1.585 / 0.9912	2.8141 / 0.9712	2.5072 / 0.9908
AS Amber Latvijas balzams	0.7925 / 0.3733	0.9035 / 0.6365	1.0783 / 0.7823	1.2902 / 0.863	2.1971 / 0.8213	2.3683 / 0.8829	3.8288 / 0.7993	4.3209 / 0.8271	4.4796 / 0.8771	5.0095 / 0.8905
SAF Tehnika A/S	0.4262 / 0.5139	0.6881 / 0.7089	4.9192 / 0.1778	4.9692 / 0.2905	5.9664 / 0.3095	5.9815 / 0.4253	7.3089 / 0.3974	8.3211 / 0.4027	8.4572 / 0.4888	10.8973 / 0.3656
AS Latvijas Juras medicinas centrs	0.5842 / 0.4447	4.4484 / 0.1082	4.7231 / 0.1932	7.9214 / 0.0945	10.0585 / 0.0736	10.317 / 0.1119	11.4942 / 0.1185	12.4099 / 0.1338	12.4389 / 0.1897	13.2056 / 0.2124
AS Siguldas Ciltslietu Un Makslīgas Apseklšanas Stacija	0.0012 / 0.972	0.04 / 0.9802	0.0667 / 0.9955	0.0948 / 0.9989	0.9316 / 0.9679	1.0339 / 0.9843	2.6953 / 0.9117	3.211 / 0.9204	4.1744 / 0.8996	5.1115 / 0.8836
TKM Grupp AS	0.8581 / 0.3543	0.9317 / 0.6276	2.4482 / 0.4847	3.9944 / 0.4068	7.1738 / 0.208	8.8933 / 0.1797	8.2733 / 0.3091	8.4202 / 0.3935	8.7634 / 0.4594	7.6016 / 0.6677
AS Merko Ehitus	0.0849 / 0.7708	0.4412 / 0.802	4.9587 / 0.1748	5.2698 / 0.2607	5.5374 / 0.3539	5.1995 / 0.5185	5.7568 / 0.5684	8.0377 / 0.4298	7.9425 / 0.5399	9.2874 / 0.5051
AS Silvano Fashion Group	0.0021 / 0.9636	0.044 / 0.9782	1.483 / 0.6862	1.8645 / 0.7607	2.0642 / 0.8402	4.744 / 0.577	6.7853 / 0.4516	7.3526 / 0.4991	8.0257 / 0.5316	8.4362 / 0.5863
Nordic Fibreboard AS	0.3088 / 0.5784	0.3537 / 0.8379	1.3231 / 0.7236	4.5527 / 0.3363	11.2355 / 0.0469*	13.4732 / 0.0361*	13.1514 / 0.0685	14.0831 / 0.0796	15.8716 / 0.0696	14.9178 / 0.1351
Novo Nordisk	1.3644 / 0.2428	2.4454 / 0.2944	2.1024 / 0.5514	2.4471 / 0.6541	2.5977 / 0.7617	6.3818 / 0.3818	7.4082 / 0.3876	8.1622 / 0.4178	10.8057 / 0.2893	11.5328 / 0.3175
AstraZeneca PLC	1.3203 / 0.2505	1.722 / 0.4227	2.1541 / 0.541	2.5198 / 0.6411	3.6946 / 0.5942	5.1774 / 0.5213	6.1425 / 0.5232	7.3291 / 0.5016	7.0226 / 0.6348	7.1022 / 0.7158
Equinor ASA	0.187 / 0.6654	0.404 / 0.8171	0.668 / 0.8807	1.5623 / 0.8156	2.8498 / 0.7231	3.7125 / 0.7155	3.9169 / 0.7893	5.9469 / 0.6532	6.0609 / 0.7338	9.5353 / 0.4822
Investor AB	4.8112 / 0.0283*	7.975 / 0.0185*	7.2268 / 0.065	7.2589 / 0.1228	7.41 / 0.1919	8.4969 / 0.2039	10.4877 / 0.1626	10.7399 / 0.2169	11.0049 / 0.2754	10.6345 / 0.3867
AB Volvo (publ)	0.4177 / 0.5181	0.7227 / 0.6967	1.1768 / 0.7586	4.3121 / 0.3654	4.2107 / 0.5195	7.4964 / 0.2774	8.2998 / 0.3069	8.0142 / 0.4321	8.5781 / 0.4771	8.308 / 0.5988
Microsoft Corporation	0.7372 / 0.3906	2.9732 / 0.2261	5.4202 / 0.1435	6.4382 / 0.1687	6.8346 / 0.2332	8.185 / 0.2249	7.881 / 0.3432	7.538 / 0.4798	7.2296 / 0.6132	6.5061 / 0.7711
Apple Inc.	3.5748 / 0.0587	19.1147 / 0.0001*	21.6358 / 0.0001*	21.6552 / 0.0002*	22.5404 / 0.0004*	25.0268 / 0.0003*	25.5185 / 0.0006*	24.5365 / 0.0019*	24.4565 / 0.0036*	25.1071 / 0.0051*
NVIDIA Corporation	0.1182 / 0.731	0.1309 / 0.9367	5.2034 / 0.1575	6.8352 / 0.1449	7.029 / 0.2185	8.7539 / 0.1879	11.0059 / 0.1384	11.0248 / 0.2003	11.2771 / 0.2572	10.0661 / 0.4347
Alphabet Inc.	0.0079 / 0.9293	1.2178 / 0.544	1.0288 / 0.7943	1.1497 / 0.8863	1.6765 / 0.8918	2.6104 / 0.8559	4.1558 / 0.7617	4.6777 / 0.7914	7.9102 / 0.5432	8.6945 / 0.5613
Amazon.com, Inc.	2.1669 / 0.141	4.4764 / 0.1067	5.9038 / 0.1164	6.0458 / 0.1958	6.567 / 0.2549	8.3415 / 0.2141	8.5729 / 0.2848	8.2367 / 0.4107	8.1526 / 0.5188	7.9527 / 0.6335
Gold	0.0902 / 0.764	0.4656 / 0.7923	0.3607 / 0.9482	1.3971 / 0.8447	1.4016 / 0.9241	1.2687 / 0.9734	1.4769 / 0.9831	1.5919 / 0.9911	5.979 / 0.742	6.3138 / 0.7882
Copper	1.8547 / 0.1732	4.9764 / 0.0831	4.7322 / 0.1925	4.969 / 0.2905	6.0298 / 0.3033	6.619 / 0.3575	8.4993 / 0.2906	10.1371 / 0.2555	10.88 / 0.284	11.5086 / 0.3193
Crude Oil	0.0172 / 0.8957	1.1761 / 0.5554	0.8205 / 0.8446	3.208 / 0.5236	10.816 / 0.0552	12.37 / 0.0542	11.5382 / 0.1168	14.4537 / 0.0707	21.731 / 0.0098*	19.4509 / 0.0349*
Natural Gas	0.4783 / 0.4892	2.749 / 0.253	2.8879 / 0.4092	2.9425 / 0.5675	8.5873 / 0.1267	8.0954 / 0.2312	8.9281 / 0.2579	9.9628 / 0.2676	10.4252 / 0.3172	14.7745 / 0.1405
Coffee	0.2387 / 0.6251	1.7863 / 0.4094	3.372 / 0.3377	3.8666 / 0.4244	3.9637 / 0.5547	5.7948 / 0.4466	6.3971 / 0.4942	6.9794 / 0.5389	7.3069 / 0.6052	7.84 / 0.6445
USD/EUR	3.9671 / 0.0464*	5.2982 / 0.0707	5.1284 / 0.1626	5.1682 / 0.2705	5.3168 / 0.3785	5.4666 / 0.4855	8.4311 / 0.2961	8.1372 / 0.4202	9.6931 / 0.3759	9.2697 / 0.5067
USD/GBP	2.184 / 0.1395	2.9922 / 0.224	3.3645 / 0.3388	3.3286 / 0.5044	3.4768 / 0.6269	4.1886 / 0.6512	5.16 / 0.6405	5.2052 / 0.7354	6.0279 / 0.7371	6.0958 / 0.8071
USD/JPY	0.5503 / 0.4582	1.2903 / 0.5246	2.1938 / 0.5332	2.4104 / 0.6607	2.5897 / 0.7629	2.8111 / 0.8322	4.3888 / 0.7341	4.019 / 0.8554	5.3127 / 0.8062	5.5885 / 0.8486
USD/CHF	3.309 / 0.0689	3.3767 / 0.1848	3.4713 / 0.3245	5.0901 / 0.2782	7.0767 / 0.215	7.9787 / 0.2397	9.3063 / 0.2314	9.2165 / 0.3244	14.8057 / 0.0964	13.8964 / 0.1778
S&P 500	3.1253 / 0.0771	4.0095 / 0.1347	4.4584 / 0.216	4.6408 / 0.3262	4.8581 / 0.4334	5.7785 / 0.4485	6.1406 / 0.5234	5.7568 / 0.6745	5.2691 / 0.8102	4.847 / 0.9012
DAX	2.4526 / 0.1173	2.5162 / 0.2842	2.177 / 0.5365	2.6253 / 0.6223	3.0577 / 0.6911	3.3309 / 0.7663	4.8118 / 0.6829	5.9819 / 0.6493	5.6471 / 0.7746	6.8006 / 0.7441
CAC40	4.8505 / 0.0276*	4.9009 / 0.0863	4.3128 / 0.2296	5.7233 / 0.2208	5.5516 / 0.3523	5.7561 / 0.4511	9.6255 / 0.2108	9.4567 / 0.3052	9.4987 / 0.3926	8.974 / 0.5346
Maturity: 1 year	0.0942 / 0.7589	1.0137 / 0.6024	0.5978 / 0.8969	1.4051 / 0.8433	1.4986 / 0.9132	3.023 / 0.806	3.0763 / 0.8779	3.688 / 0.8841	3.4543 / 0.9435	4.4193 / 0.9265
Maturity: 3 years	0.4748 / 0.4908	3.6461 / 0.1615	3.4301 / 0.3299	5.7188 / 0.2212	6.0594 / 0.3005	8.6464 / 0.1945	8.1196 / 0.3222	8.7943 / 0.3599	8.963 / 0.4407	11.6393 / 0.3099
Maturity: 5 years	1.0546 / 0.3045	3.424 / 0.1805	3.2145 / 0.3597	5.4137 / 0.2474	6.4841 / 0.2619	8.5096 / 0.2031	8.4623 / 0.2936	9.0789 / 0.3357	9.0934 / 0.4287	11.3169 / 0.3334
Maturity: 10 years	1.5423 / 0.2143	2.8057 / 0.2459	2.6361 / 0.4512	4.0347 / 0.4013	4.6977 / 0.4539	5.3087 / 0.5049	6.7193 / 0.4587	6.962 / 0.5407	6.7692 / 0.6611	9.7987 / 0.4583
Maturity: 30 years	3.0029 / 0.0831	3.1371 / 0.2083	3.0368 / 0.386	4.4951 / 0.3431	5.9971 / 0.3065	6.6488 / 0.3546	11.0295 / 0.1373	10.8119 / 0.2126	10.776 / 0.2914	13.9223 / 0.1766

3 priedas. Dieninių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu

Koreliacijos Matrica:	„OMX Vilnius“ gražos indeksas
OMXVGI	1*
Telia Lietuva	0.4463*
AB Siauliu Bankas	0.4387*
APB Apranga	0.2746
Grigeo AB	0.1776
Invalda INVL	0.3219*
AS Latvijas Gaze	0.1496
AS Amber Latvijas balzams	0.1989
SAF Tehnika A/S	0.1827
AS Latvijas Juras medicinas centrs	0.0029
AS Siguldas Ciltslietu Un Maksligas Apsklosanas Stacija	0.0159
TKM Grupp AS	0.2798
AS Merko Ehitus	0.2015
AS Silvano Fashion Group	0.2505
Nordic Fibreboard AS	0.1099
Novo Nordisk	0.0697
AstraZeneca PLC	0.1399
Equinor ASA	0.1197
Investor AB	0.2367
AB Volvo (publ)	0.216
Microsoft Corporation	0.1189
Apple Inc.	0.1118
NVIDIA Corporation	0.0382
Alphabet Inc.	0.1071
Amazon.com, Inc.	0.0833
Gold	0.0349
Copper	0.1495
Crude Oil	0.0883
Natural Gas	0.0354
Coffee	0.0752
USD/EUR	-0.0268
USD/GBP	-0.0713
USD/JPY	0.0055
USD/CHF	-0.0129
S&P 500	0.2091
DAX	0.3142*
CAC40	0.3126*
Maturity: 1 year	0.0656
Maturity: 3 years	0.0479
Maturity: 5 years	0.0381
Maturity: 10 years	0.0529
Maturity: 30 years	0.0437

4 priedas. Mėnesinių duomenų „Pearson“ koreliacijos tyrimas lyginant su „OMX Vilnius“ gražos indeksu

Koreliacijos Matrica:	„OMX Vilnius“ gražos indeksas
OMXVGI	1*
Telia Lietuva	0.463*
AB Siauliu Bankas	0.598*
APB Apranga	0.5051*
Grigeo AB	0.286
Invalda INVL	0.5008*
AS Latvijas Gaze	0.1475
AS Amber Latvijas balzams	0.274
SAF Tehnika A/S	0.2834
AS Latvijas Juras medicinas centrs	0.0686
AS Siguldas Ciltslietu Un Maksligas Apsklosanas Stacija	0.1247
TKM Grupp AS	0.4956*
AS Merko Ehitus	0.283*
AS Silvano Fashion Group	0.446*
Nordic Fibreboard AS	0.1715
Novo Nordisk	0.138
AstraZeneca PLC	0.1683
Equinor ASA	0.192
Investor AB	0.2811
AB Volvo (publ)	0.316*
Microsoft Corporation	0.2641
Apple Inc.	0.2476
NVIDIA Corporation	0.0697
Alphabet Inc.	0.2639
Amazon.com, Inc.	0.1758
Gold	0.139
Copper	0.3172*
Crude Oil	0.2539
Natural Gas	0.1397
Coffee	0.1268
USD/EUR	-0.1735
USD/GBP	-0.254
USD/JPY	0.0031
USD/CHF	-0.0891
S&P 500	0.4061*
DAX	0.4332*
CAC40	0.4432*
Maturity: 1 year	0.0714
Maturity: 3 years	0.0405
Maturity: 5 years	-0.0072
Maturity: 10 years	-0.0935
Maturity: 30 years	-0.0197