



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Akcijų Baltijos šalyse kintamumo ir jautrumo analizė

Baigiamasis magistro projektas

Indrė Jagminaitė
Projekto autorė

Prof. Dr. Robertas Alzbutas
Vadovas

Doc. Dr. Aura Drakšaitė
Vadovė

Kaunas, 2019



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Akcijų Baltijos šalyse kintamumo ir jautrumo analizė

Baigiamasis magistro projektas

Didžiųjų verslo duomenų analitika (6213AX001)

Indrė Jagminaitė

Projekto autorė

Prof. Dr. Robertas Alzbutas

Vadovas

Doc. Dr. Aura Drakšaitė

Vadovė

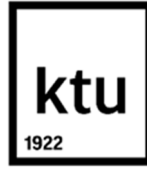
Doc. Dr. Vytautas Janilionis

Recenzentas

Doc. Dr. Ginta Railienė

Recenzentė

Kaunas, 2019



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas
Indrė Jagminaitė

Akcijų Baltijos šalyse kintamumo ir jautrumo analizė

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Indrės Jagminaitės, baigiamasis projektas tema „Akcijų Baltijos šalyse kintamumo ir jautrumo analizė“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

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

(vardą ir pavardę įrašyti ranka)

(parašas)

Jagminaitė, Indrė. Akcijų Baltijos šalyse kintamumo ir jautrumo analizė. Magistro baigiamasis projektas / vadovas prof. dr. Robertas Alzbutas, vadovė doc. dr. Aura Drakšaitė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Taikomosios matematikos kryptis, Didžiųjų verslo duomenų analitika.

Reikšminiai žodžiai: akcijos, jautrumas, kintamumas, prognozavimas, Baltijos šalys.

Kaunas, 2019. 80 p.

Santrauka

Pagrindinė atliekamo tyrimo paskirtis yra padėti investuotojams priimti kiek galima tikslesnius sprendimus, susijusius su akcijų pirkimu, pardavimu ar laikymu. Čia prognozuojamas trijų įmonių akcijų kainų kitimas viena diena į priekį, panaudojant uždarymo kainų istorinius duomenis (akcijų prekybos duomenys ir techniniai rodikliai), bei trijų Baltijos šalių akcijų rinkų indeksų kitimas mėnesiu į priekį, kai pasirenkami įvairūs makroekonominiai rodikliai. Prognozavimas įgyvendinamas, pritaikant mašininio mokymosi metodus ir atliekant klasifikavimą, kai tikslo stulpelių reikšmės yra tokios: „kaina/indeksas kils“ ir „kaina/indeksas nekils“. Prieš tai iškeliami prielaidai, jog kintamumo prognozavimas gali būti tikslesnis, jei atrenkami reikšmingi kintamieji ir tai daroma panaudojant įvairius jautrumo analizės metodus.

Kai prognozuojamas įmonių akcijų kainų kitimas, Lietuvos ir Estijos atvejais reikšmingų kintamųjų atrinkimas neturėjo įtakos, o Latvijos atveju buvo naudingas dalinės koreliacijos metodas, kadangi tada klasifikavimo tikslumas pagerėjo. Geriausių sudarytų modelių tikslumo (AUC) įverčiai (0,836, 0,824 ir 0,805) parodė, jog panaudojant vien tik istorinius akcijų kainos duomenis galima gauti gerus prognozavimo rezultatus viena diena į priekį, kai taikomas atraminių vektorių klasifikatorius.

Prognozuojant akcijų rinkų indeksų kitimą mėnesiu į priekį, sudarytų modelių tikslumo (AUC) reikšmės nebuvo aukštos nė vienos šalies atveju. Tačiau reikšmingų kintamųjų atrinkimas, pritaikant jautrumo analizės metodus, turėjo įtakos tam, jog klasifikavimo tikslumas pagerėdavo bent keliais procentais. Duomenų paslinkimas vienu ar dviem periodais (mėnesiais) į priekį (iškelta prielaida, jog makroekonominiai rodikliai gali ne iš karto veikti indeksų pokyčius) taip pat dažnai turėjo teigiamos reikšmės tam, kad buvo klasifikuojama tiksliau. Šio tyrimo metu geriausiai klasifikavo atsitiktinių miškų, atraminių vektorių ir sprendimų medžio metodai. Dažniausiai atrinkti reikšmingi makroekonominiai kintamieji pagal šalis buvo tokie: darbo užmokestis, gyventojų užimtumas, valdžios sektoriaus skola (Lietuva); darbo užmokestis, statybos sąnaudų elementų kainų indeksas (Latvija); pramonės produkcijos kainų indeksas, vieno ir trijų mėnesių EURIBOR palūkanų normos, bendrosios nacionalinės pajamos (Estija).

Tyrimas buvo pristatytas konferencijoje „Matematika ir matematikos dėstymas – 2019“.

Jagminaitė, Indrė. Analysis of the volatility and sensitivity of shares in the Baltic countries. Master's Final Degree Project / supervisor Prof. Dr. Robertas Alzbutas, supervisor Assoc. Prof. Dr. Aura Drakšaitė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics field, Business Big Data Analytics.

Keywords: stocks, sensitivity, volatility, prediction, the Baltic countries.

Kaunas, 2019. 80 pages.

Summary

The main purpose of the research is to help investors to make the most accurate decisions about buying, selling or holding shares. In this study it is predicted that the stock prices of the three companies will change one day ahead, using historical data on closing prices (stock trading data and technical indicators), and the three Baltic stock market indices will change one month ahead when various macroeconomic indicators are chosen. Predicting is implemented by adapting machine learning methods and using classification when the values of the target columns are as follows: "price / index will rise" and "price / index will not rise". This is preceded by the assumption that prediction of volatility can be more accurate if significant variables are selected and this is done using different sensitivity analysis methods.

When predicting stock price volatility of companies, in the cases of Lithuania and Estonia the selection of significant variables did not influence, and in the case of Latvia the partial correlation method was useful, as then the accuracy of classification was improved. AUC estimates (0.836, 0.824, and 0.805) of the best models showed that only historical stock price data can provide good predicting results one day ahead when using the Support Vector Machine algorithm.

Predicting the volatility of stock market indices one month ahead, the AUC values of models were not high for any country. However, the selection of significant variables by applying sensitivity analysis methods had an impact on the accuracy of classification by at least a few percent. The data drift in one or two periods (months) ahead (assuming that macroeconomic indicators may not immediately affect index changes) also often had positive impact on the accuracy of classification. In this study, the best classifiers are: Random Forests, Support Vector Machine and Decision Tree. The most significant macroeconomic variables selected by country were: wages, employment of population, general government debt (Lithuania); wages, construction cost element price index (Latvia); industrial output price index, one and three month EURIBOR interest rates, gross national income (Estonia).

Turinys

Lentelių sąrašas	7
Paveikslų sąrašas	8
Santrumpų ir terminų sąrašas	9
Įvadas.....	10
1. Literatūros apžvalga	11
1.1. Akcijos ir jų reikšmė vertybinių popierių kontekste	11
1.2. Jautrumo analizė akcijų kitimo tyrimuose	13
1.3. Veiksnių įtaka akcijų pokyčiams bei jų kategoriniai aspektai	14
1.4. Mašininio mokymosi metodų taikymas prognozuojant akcijų kitimą	21
1.5. Akcijų Baltijos šalyse kitimo tyrimų apžvalga	23
1.6. Darbo temos ir uždavinių pagrindimas	25
2. Tyrimų metodai ir priemonės	27
2.1. Tyrimo struktūra	27
2.2. Tyrimo metodai ir modelių vertinimo priemonės	28
2.2.1. Mašininio mokymosi metodai	28
2.2.2. Klasifikavimo modelių vertinimas ir atranka	32
2.2.3. Jautrumo analizės metodai.....	34
2.3. Programinė realizacija.....	35
3. Tyrimo rezultatai ir jų aptarimas	37
3.1. Duomenų apžvalga.....	37
3.1.1. Įmonių akcijų kainų kitimo prognozavimo metu naudojami duomenys	37
3.1.1.1. Lietuvos įmonė AB „Grigeo“	39
3.1.1.2. Latvijos įmonė AS „Grindeks“	40
3.1.1.3. Estijos įmonė AS „Tallinna Kaubamaja Grupp“	40
3.1.2. Baltijos šalių akcijų rinkų indeksų kitimo prognozavimo metu naudojami duomenys.....	42
3.2. Įmonių akcijų kainų kitimo prognozavimo diena į priekį rezultatai	44
3.2.1. Lietuvos įmonė AB „Grigeo“	44
3.2.2. Latvijos įmonė AS „Grindeks“	46
3.2.3. Estijos įmonė AS „Tallinna Kaubamaja Grupp“	48
3.3. Akcijų rinkų indeksų kitimo prognozavimo mėnesiu į priekį rezultatai.....	51
3.3.1. Lietuvos akcijų rinkos indeksas OMX Vilnius.....	51
3.3.2. Latvijos akcijų rinkos indeksas OMX Riga.....	54
3.3.3. Estijos akcijų rinkos indeksas OMX Tallinn	56
Išvados	60
Literatūros sąrašas	62
Informacinių šaltinių sąrašas	66
Priedai.....	67
1 priedas. Techninė analizė paremtos tyrimo dalies duomenų apdorojimo kodas	67
2 priedas. Įmonių akcijų kainų kitimo prognozavimo/klasifikavimo kodas	70
3 priedas. Fundamentaliąja analize paremtos tyrimo dalies duomenų apjungimo kodas.....	74
4 priedas. Fundamentaliąja analize paremtos tyrimo dalies duomenų apdorojimo kodas	75
5 priedas. Akcijų rinkų indeksų kitimo prognozavimo/klasifikavimo kodas.....	77

Lentelių sąrašas

1 lentelė. Akcijų ir makroekonominių rodiklių ryšį analizavusių tyrėjų darbų apibendrinimas.....	17
2 lentelė. Kitimo prognozavimo viena diena į priekį tikslai pagal Weng'ą [33]	22
3 lentelė. Akcijų kainų kitimo prognozavimą atlikusių tyrėjų darbų apibendrinimas	23
4 lentelė. Įvairių veiksnių poveikį Baltijos šalių akcijoms vertinusių tyrėjų darbų apibendrinimas	25
5 lentelė. Sumaišymų matrica, nurodant teigiamas ir neigiamas reikšmes iš viso [46].....	32
6 lentelė. Modelių vertinimo matai ir jų skaičiavimo formulės [46]	33
7 lentelė. Tyrime taikomi makroekonominiai rodikliai ir jų aprašymas	43
8 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės AB „Grigeo“ atvejį.....	44
9 lentelė. AB „Grigeo“ duomenų klasifikavimo rezultatai pagal tikslumo matą AUC	45
10 lentelė. AB „Grigeo“ akcijų kainos kitimo prognozavimo tikslumo matai. Slenkstis 0,5.....	46
11 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės AS „Grindeks“ atvejį.....	46
12 lentelė. AS „Grindeks“ duomenų klasifikavimo rezultatai pagal tikslumo matą AUC	47
13 lentelė. AS „Grindeks“ akcijų kainos kitimo prognozavimo tikslumo matai. Slenkstis 0,5.....	48
14 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės AS „Tallinna Kaubamaja Grupp“ atvejį...	49
15 lentelė. AS „Tallinna Kaubamaja Grupp“ duomenų klasifikavimo rezultatai pagal tikslumo matą AUC.....	49
16 lentelė. AS „Tallinna Kaubamaja Grupp“ akcijų kainos kitimo prognozavimo tikslumo matai.	50
17 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Vilnius indekso atvejį	52
18 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti.....	53
19 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu.....	53
20 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais	53
21 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Riga indekso atvejį.....	54
22 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti	55
23 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu.....	55
24 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais	55
25 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Tallinn indekso atvejį.....	57
26 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti.....	58
27 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu.....	58
28 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais.....	58

Paveikslų sąrašas

1 pav. Akcijų rinkų vertė pagal šalis pasaulyje 2017 metais	11
2 pav. Akcijų kainos kitimui įtaką darantys veiksniai	15
3 pav. Fundamentaliosios ir techninės analizės sudėtinės dalys	18
4 pav. Atliekamo tyrimo schema	27
5 pav. Neuroninio tinklo schema [46].....	30
6 pav. Atraminių vektorių metodo tiesinio klasių atskyrimo schema su skirtingomis paraštėmis [46]	31
7 pav. ROC kreivės pavyzdys [49]	33
8 pav. Įmonių akcijų kainų pokyčiai ilguoju laikotarpiu	37
9 pav. Įmonių akcijų kainų grafikai	38
10 pav. Bollinger'io juostos pagal įmonės AB „Grigeo“ akcijų uždarymo kainos duomenis.....	39
11 pav. Bollinger'io juostos pagal įmonės AS „Grindeks“ akcijų uždarymo kainos duomenis.....	40
12 pav. Bollinger'io juostos pagal įmonės AS „Tallinna Kaubamaja Grupp“ akcijų.....	40
13 pav. Baltijos šalių akcijų rinkų indeksų grafikai.....	42
14 pav. Geriausio modelio ROC kreivė pagal AB „Grigeo“ akcijų kainos duomenis	45
15 pav. Geriausi AUC įverčiai pagal AB „Grigeo“ akcijų kainos duomenis	45
16 pav. Geriausio modelio ROC kreivė pagal AS „Grindeks“ akcijų kainos duomenis	47
17 pav. Geriausi AUC įverčiai pagal AS „Grindeks“ akcijų kainos duomenis	48
18 pav. Geriausio modelio ROC kreivė pagal AS „Tallinna Kaubamaja Grupp“ akcijų kainos duomenis.....	50
19 pav. Geriausi AUC įverčiai pagal AS „Tallinna Kaubamaja Grupp“ akcijų kainos duomenis ..	50
20 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Vilnius indeksas.....	52
21 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Riga indeksas	55
22 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Tallinn indeksas	56

Santrumpų ir terminų sąrašas

Santrumpos:

BNP – bendrosios nacionalinės pajamos;

BVP – bendrasis vidaus produktas;

VP – vertybiniai popieriai.

Terminai:

Akcijų rinka – apibūdinama kaip „darinys“, kur bendradarbiauja akcijų leidėjai, pirkėjai, pardavėjai ir kiti suinteresuoti rinkos dalyviai. Tai – tik dalis finansų rinkos. Ji apima visus nuosavybės vertybinius popierius, nepriklausomai ar jie prekiaujami akcijų biržose, ar ne.

Kapitalizacija – kokio nors ūkio subjekto, prekės, gaminio, išteklių vertė, gaunama padauginus esamą rinkos kainą iš atitinkamo objekto vienetų skaičiaus.

Mokymasis „be mokytojo“ – terminas naudojamas duomenų tyrimuose, kai taikomi tokie metodai kaip klasterizavimas, susietumo taisyklės.

Mokymasis „su mokytoju“ – terminas naudojamas duomenų tyrimuose, kai taikomi tokie metodai kaip klasifikavimas, regresinė analizė.

Neapibrėžtumas – savybė, nusakanti analizuojamų parametru ar procesų netikslumą dėl žinių trūkumo ir (arba) atsitiktinių (stochastinių) procesų.

Normalusis skirstinys - vaizduoja vieną iš empiriškai patikrintų tiesų bendrosios realybės atžvilgiu ir jo padėtis gali būti nagrinėjama kaip viena iš fundamentaliųjų gamtos dėsnių.

Ivadas

Investavimas į vertybinius popierius, teigiamos grąžos generavimas žmones domina jau labai seniai. Bene populiariausios vertybinių popierių rūšys yra akcijos ir obligacijos. Obligacijos kai kam gali atrodyti esančios „nuobodokos“ ir ne itin pelningos, o akcijomis domisi plačioji visuomenės dalis, kadangi jų kainų ar indeksų, grąžų pokyčiai yra lengviau interpretuojami ir suprantami. Gal dėl to autoriai visame pasaulyje mieliau atlieka tyrimus, apibūdinančius šią vertybinių popierių rūšį, o ne obligacijas.

Pastebėta, jog tyrėjus visame pasaulyje, kai atliekama akcijų analizė, ypač dominantis objektas yra įvairūs veiksniai, darantys poveikį akcijų kainų ar akcijų rinkų indeksų kitimui. Iš to išplaukiantys prognozavimo tyrimai yra paremti kuo kokybiškesnių ir tikslesnių modelių sudarymo būdų paieškomis. Nors vieni autoriai bando prognozuoti tikslias reikšmes, kiti mano, jog to padaryti neįmanoma, todėl apsiriboja trumpo laikotarpio akcijų kainų ar akcijų rinkų indeksų kitimo spėjimais. Šiame darbe taip pat pasirenkamas pastarasis būdas.

Kai atliekami tokie tyrimai, vieni autoriai labiau kreipia dėmesį į kuo geresnių modelių parametru parinkimą, kiti labiau akcentuoja taikomus prognozavimo metodus, dar kiti labai daug reikšmės skiria duomenims, tačiau beveik nekreipia dėmesio į kintamųjų atrinkimą, nors dažnai teigiama, jog sudarant klasifikavimo/prognozavimo modelius yra svarbu atrinkti reikšmingus veiksnius tam, kad būtų galima pagerinti tikslumą ir pačių modelių kokybę. Būtent tai šiame darbe apibrėžiama kaip **problema**, į kurią verta įsigilinti labiau ir paieškoti sprendimo būdų.

Čia pagrindinis dėmesys bus skiriamas nustatyti, ar pritaikant skirtingus jautrumo analizės metodus ir tokiu būdu atrinkus reikšmingus kintamuosius, galima pagerinti klasifikavimo tikslumą, kai prognozuojamas atskirų įmonių akcijų kainų ir akcijų rinkų indeksų kitimas atitinkamai viena diena ir vienu mėnesiu į priekį. Naudojami Baltijos šalių akcijų rinkų duomenys, kadangi, kaip pastebėta, čia tokio pobūdžio tyrimai nelabai atlikti.

Darbo **tikslas** – ištirti jautrumo analizės metodų pritaikymą ir galimą netiesioginę įtaką akcijų Baltijos šalyse kitimo prognozavimo tikslumui.

Tikslui įgyvendinti iškeliami tokie **uždaviniai**:

- Apžvelgti Baltijos šalių akcijų rinką ir sudaryti informatyvias duomenų matricas;
- Išskirti prognozuojant galimai tinkamus jautrumo analizės ir mašininio mokymosi metodus;
- Taikant jautrumo analizės metodus, nustatyti kintamųjų reikšmingumą;
- Atlikti kitimo prognozavimą ir palyginti gautų modelių tikslumą;
- Nustatyti adekvačiausiai klasifikuojančius metodus.

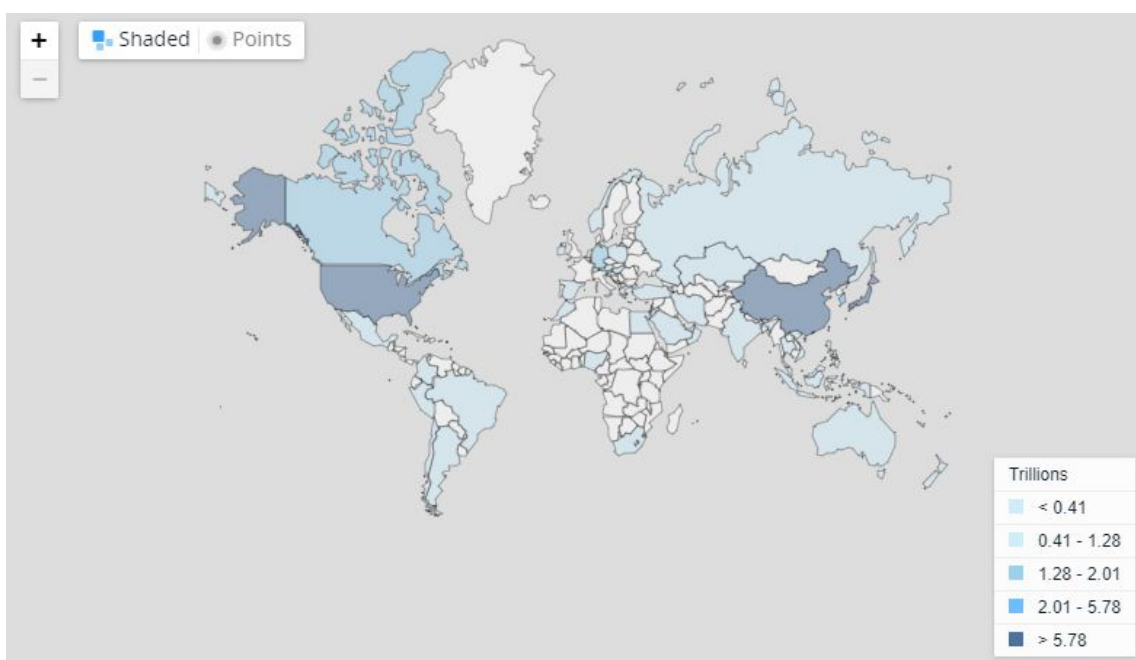
1. Literatūros apžvalga

Siekiant apibrėžti pasirinktos srities problemą bei išryškinti pagrindinius jos aspektus, kuriais remiantis bus atliekamas tyrimas, toliau apžvelgiama, kuo išsiskiria akcijos iš kitų vertybinių popierių; kodėl apskritai mokslininkai analizuoja įvairių veiksnių įtaką atskirų įmonių akcijų kainų ar akcijų rinkų indeksų kitimui; kokie tyrimai anksčiau atlikti analizuojama tema tiek pasaulyje, tiek Baltijos šalyse, ir kita.

1.1. Akcijos ir jų reikšmė vertybinių popierių kontekste

Vertybiniai popieriai yra skirstomi į *skolos* ir *nuosavybės* vertybinius popierius. Pagrindinės jų rūšys: akcijos, obligacijos, investiciniai fondai, išvestinės finansinės priemonės, indėlių sertifikatai, taupymo lankstinai, čekiai. Pagrindiniai skolos vertybiniai popieriai yra obligacijos, o nuosavybės – akcijos. Akcijos bei obligacijos ir yra tos vertybinių popierių rūšys, į kurias visame pasaulyje investuojama labiausiai. Pasaulio banko (angl. *The World Bank*) ir Tarptautinių atsiskaitymų banko (angl. *Bank of International Settlement*) duomenimis, 2017 metais viso pasaulio akcijų rinkų vertė siekė apie 77 trilijonus Jungtinių Amerikos Valstijų dolerių, o obligacijų rinkų – apie 100 trilijonų. Taigi, obligacijų rinkų vertė yra žymiai didesnė nei akcijų. Nors obligacijos iš pirmo žvilgsnio gali atrodyti kiek „nuobodokos“, tačiau, įsigilinus į jų specifiką labiau, tampa net „įdomesnės“ nei akcijos, kurios paprastam žmogui yra daug labiau suprantamos ir lengviau interpretuojamos. Nepaisant obligacijų pranašumo vertės prasme, tyrėjus ir investuotojus visame pasaulyje labiau domina akcijos ir tai, kas turi įtakos jų kainų kitimui, kaip sumažinti investavimo riziką, kaip gauti didesnę investicinę grąžą, kaip prognozuoti pokyčius ateityje, ir panašiai. Dėl jau minėtų priežasčių šiame darbe taip pat analizuojamos akcijos, jų kainų kitimas ir tai lemiantys veiksniai.

Remiantis Pasaulio banko sudarytu pasaulio žemėlapiu (žr. 1 pav.), kuriame šalys „nuspalvintos“ atitinkama spalva pagal generuojamą vertę, matoma, jog didžiausios pasaulyje yra Jungtinių Amerikos Valstijų, Kinijos ir Japonijos akcijų rinkos, o Europoje iš kitų šalių išsiskiria Vokietija.



1 pav. Akcijų rinkų vertė pagal šalis pasaulyje 2017 metais

Apžvelgus nemažai su akcijų kainų ir akcijų rinkų indeksų kitimo analize, prognozavimu susijusių tyrimų pastebėta, jog tyrėjai labai dažnai, o gal net ir dažniausiai, analizei renkasi ne kažkurių kompanijų akcijų kainas, o akcijų rinkų indeksus. Anot BalaGuru ir Kumar'o [1], akcijų rinkos indeksas sudaromas pasirinkus keletą akcijų, kurios atspindi visą rinką, tam tikrą sektorių ar rinkos segmentą. Tokio vertybinių popierių krepšelio kainų pokytis vertinamas pagal bazinį laikotarpį. Egzistuoja nuostata, jog skirtingoms akcijoms suteikiami tam tikri svoriai pagal jų svarbą ekonomikoje, o akcijų rinkos indeksas veikia kaip visos ekonomikos ar ūkio šakos, kurioje yra įmonės, darbą apibūdinantis indikatorius.

Faerber'is [2] savo knygoje teigia, jog akcijų rinkų indeksai vertinami kaip gyvybiškai svarbūs ženklai, rodantys akcijų rinkos kryptį. Šie indeksai daro įtaką ne tik atskirų investuotojų investavimo elgesiui, bet taip pat ir investavimo profesionalų (fondų valdytojų) sprendimams. Idealiu atveju indekso kainos pokytis yra proporcingas akcijų, kurios sudaro tą indeksą, kainos pokyčiui, bet dažnai taip nebūna [3]. Be viso to nurodoma, jog akcijų rinkos indeksai suteikia skirtingas akcijų rinkų matavimo galimybes. Juos galima panaudoti tokiais būdais:

- norint nustatyti, kaip apskritai veikia akcijų rinkos;
- taikyti kaip palyginamuosius standartus, siekiant apžvelgti individualių investicinių portfelių, investicinių fondų ar biržos prekybos fondų veikimą;
- kaip ateities tendencijų prognozavimo priemonės.

Šie indeksai suteikia galimybę įvertinti ir atskirų akcijų kainų kitimą, tačiau prieš imantis bet kokių veiksmų reikia suprasti ryšius tarp indeksų ir individualių akcijų [2].

Capital Market Information Center [3] papildoma teiginius apie akcijoms, sudarančioms indeksus, suteikiamus svorius ir išskiria tris pasaulyje egzistuojančias akcijų rinkos indeksų rūšis:

1. *Kainos svertinis indeksas*. Tai toks indeksas, kai kiekvienos akcijos kainai suteikiamas vienodas svoris. Jis apskaičiuojamas kaip indeksą sudarančių akcijų kainų aritmetinis vidurkis.
2. *Vertės svertinis indeksas*. Tai indeksas, kai kiekvienai jį sudarančiai akcijai suteikiamas svoris, atitinkantis tos akcijos vertę. Skaičiuojant šią vertę yra atsižvelgiama į bendrą akcijų rinkos kapitalizaciją. Dėl to didesniems indeksą sudarantiems komponentams suteikiamas didesnis svoris, o mažesniems – mažesnis. Toks indeksų kūrimo metodas pasaulyje yra taikomas plačiausiai.
3. *Lygiomis dalimis susvertas indeksas*. Čia kiekvienos akcijos pokyčiams suteikiamas toks pat svoris. Kartais tai vadinama *indeksu be svorio*. Sudarant tokį indeksą visos akcijos turi vienodą svorį, neatsižvelgiant į kainą ar rinkos vertę.

Apibendrinant galima teigti, kad nors obligacijų rinkų vertė pasaulyje yra didesnė nei akcijų, vis dėlto pastarosios, jų kainų ar indeksų, gražų pokyčiai yra lengviau interpretuojami ir suprantami. Taip pat akcijomis labiau domisi plačioji visuomenės dalis, todėl, galbūt dėl to, autoriai visame pasaulyje mieliau atlieka tyrimus apibūdinančius šią vertybinių popierių rūšį, o ne obligacijas. Pastebėta, jog daugelyje su akcijomis susijusių tyrimų renkamas analizuoti ne atskirų įmonių akcijų kainas, o indeksus. Jie informatyvesni, kai norima stebėti pokyčius visoje akcijų rinkoje, sektoriuje ar rinkos segmente. Taip pat jie nurodo akcijų rinkos kryptį. Vis dėlto šiame darbe analizuojamas tiek atskirų įmonių, veikiančių Baltijos šalyse, akcijų kainų kitimas, tiek šių šalių akcijų rinkų

indeksų kitimas, atliekant prognozavimą. Taip norima visapusiškai ištirti ir apibrėžti tam įtakos turinčius veiksnius.

Norint nustatyti, kokie veiksniai lemia akcijų kainų ar akcijų rinkų indeksų pokyčius, reikia pritaikyti tam tinkamus tyrimų metodus. Tada bus sužinoma, kaip jautriai akcijų kainos reaguoja į vienokius ar kitokius išorės faktorius. Kitame poskyryje ir apibrėžiama, kas yra jautrumo analizė, kokie metodai dažniausiai taikomi praktikoje, ir kaip tai galima panaudoti akcijų kainų kitimo tyrimuose.

1.2. Jautrumo analizė akcijų kitimo tyrimuose

Hofer'is [4] savo darbe nurodo, kad jautrumo analizė dažnai atliekama kartu su neapibrėžtumo analize, o pagrindinis jautrumo analizės tikslas yra apibrėžti veiksnius, lemiančius modelio rezultatų neapibrėžtumą. Kitaip tariant, jautrumo analizė parodo, kaip dėl modelio įvesties elementų reikšmių pasikeitimo keičiasi stebimojo dydžio (priklausomo kintamojo) reikšmės, ir kuris įvesties parametras ar parametrai yra svarbesni, turintys daugiau įtakos vienokiam ar kitokiam išvesties rezultatui. Dar kartais vartojamas terminas „neapibrėžtumo svarbos analizė“. Neapibrėžtumo analizė vykdoma *Monte Carlo* simuliacija. Kopustinskas, Alzbutas ir Augutis [5] papildo šias mintis teigdami, jog įprastas jautrumo analizės rezultatas yra svarbiausių (matematinio) modelio parametrų sąrašas, o parametrų svarba vertinama kiekybiniais dydžiais. Literatūroje išskiriamos dvi jautrumo analizės rūšys: *globali* ir *lokali* jautrumo analizės arba, kitaip tariant, *globalūs* ir *lokalūs* jautrumo analizės metodai.

Lokali jautrumo analizė. Kaip teigia Cacuci's [6], lokali jautrumo analizės tikslas yra ištirti kintamųjų įtaką modelio rezultatui pasirinkto taško aplinkoje, tačiau Kopustinskas, Alzbutas ir Augutis [5] nurodo, kad tai ir parametrų reikšmių tikimybinių skirstinių neįvertinimas yra šios analizės trūkumai. Pastarieji autoriai taip pat išskiria, kad vienas paprasčiausių tokios jautrumo analizės metodų yra taškų sklaidos diagrama.

Globali jautrumo analizė. Globalios jautrumo analizės tikslas yra ištirti kintamųjų įtaką modelio rezultatams visame nepriklausomų kintamųjų reikšmės kitimo intervale ir, nustatčius tam tikrus kritinius taškus, juos įvertinti pagal lokali jautrumo analizės principus [6]. Šios analizės metodai skirstomi į imties (koreliacinė analizė, standartizuotos tiesinės regresijos modelių sudarymas, Wilkso metodas) ir dispersijos išskaidymo (Sobol'io metodas, Furjė amplitudžių jautrumo nustatymo metodas) metodus [5].

Įdomų tyrimą šioje srityje atliko Antony's ir Jeevanand'as [7], kurie, remdamiesi kitais tyrimais, teigiančiais, jog egzistuoja tiesinis ryšys tarp akcijų rinkos indekso X ir tam tikros akcijos kainos Y , analizavo pasirinktos įmonės akcijos kainų jautrumą, pritaikydami elastingumo charakteristikas, kai veikia rinkos svyravimai.

X įtaką Y galima užrašyti paprasta regresijos lygtimi:

$$Y = a + bX + \varepsilon ; \quad (1)$$

čia a ir b yra konstantos.

Teorinėse minimo tyrimo įžvalgose nurodoma, jog elastingumas kinta, kai kinta X . Tai reiškia, kad akcijos kainų jautrumas nėra vienodas esant skirtingoms akcijų rinkos indekso (X) reikšmėms. Be to, elastingumo reikšmė priklauso nuo parametrų a ir b . Kai X ir Y susiję pagal (1) lygtį, tada elastingumas išreiškiamas tokia formule:

$$\eta = \frac{bX}{a+bX} ; \quad (2)$$

$\eta = 1$, kai akcijos kainos grąža yra tokia pati kaip ir rinkos grąža. Tai reiškia, kad akcijos kaina pakyla (nukrenta) 1 %, kai indeksas pakyla (nukrenta) 1 %. Tokiu atveju $a = 0$. Jei $\eta > 1$, tai reiškia, kad akcijos grąža yra proporcingai didesnė nei rinkos grąža. Kadangi akcijų kaina ir rinkos indeksas dažniausiai koreliuoja teigiamai, tai regresijos kreivės nuolydis „ b “ taip pat yra teigiamas, todėl $\eta > 1$ tada, kai „ a “ yra neigiamas. Jei $\eta < 1$, tai reiškia, kad akcijos grąža yra proporcingai mažesnė nei rinkos grąža. Tokiu atveju akcijos kaina padidėja (sumažėja) mažiau nei 1 %, kai indeksas padidėja (sumažėja) 1%. Be to, $\eta < 1$ būna tada, kai „ a “ yra teigiamas.

Antony'io ir Jeevanand'o [7] tyrimo rezultatus galima perteikti taip: kadangi beta (šiuo atveju taikytas „ b “) yra konstanta, t. y. tiesios kreivės nuolydžio reikšmė, tai vien tik ji negali būti naudojama, norint patikrinti, kaip jautriai akcijų kainos reaguoja į rinkos indekso pokyčius. Tam turi būti taikomi elastingumo skaičiavimai. Taip pat nustatyta, jog jautrumas būna didesnis, kai indekso reikšmė labai didelė, ir mažesnis, kai indekso reikšmė vidutiniška. Norint įvertinti akcijų grąžų pokyčius, veikiamus rinkos grąžos, rekomenduojama remtis ne pastaraisiais duomenimis, o taikyti akcijų kainos elastingumo reikšmes. Apibendrinant, turto kaina grįšta elastingumu bus patrauklesnė nei pastovi beta reikšmė.

Apibendrinant galima teigti, jog *globali* jautrumo analizė įvairiuose tyrimuose taikoma dažniau nei *lokali*. O pagrindinis tokių analizių tikslas yra nustatyti veiksnius, kurių reikšmių pasikeitimas labiausiai sąlygoja modelio rezultato pasikeitimą. Šiame poskyryje apžvelgtas tyrimas, susijęs su jautrumo analize, kuriame taikyta tiesinė regresija ir elastingumo skaičiavimai, manoma, yra pakankamai informatyvus, siekiant parodyti matematiškai išreikštą šios analizės pusę ir pateikti visapusišką pavyzdį. Toliau labiau gilinamasi į tai, kokie tyrimai yra atlikti kitų pasaulio tyrėjų, norint nustatyti akcijų kainų ar indeksų pokyčiams įtakos turinčius veiksnius, kokie kintamieji dažniausiai pasirenkami analizei, kokie tyrimo metodai taikomi ir kita.

1.3. Veiksnių įtaka akcijų pokyčiams bei jų kategoriniai aspektai

Graham'as ir Dodd'as [8] dar 1934 metais išskyrė keturias esmines dedamąsias, kurių tinkamas įvertinimas vertybinių popierių analizės metu, anot jų, galėtų padėti atsakyti į klausimą: „Ar pirkti, ar parduoti, ar laikyti tam tikrą finansinį instrumentą?“. Tos dedamosios yra:

- *Asmuo*. Asmeninės savybės daugiau ar mažiau turi įtakos kiekvieno finansinio instrumento pasirinkimui. Šios savybės gali būti patirtis ir kompetencija finansų srityje, asmens temperamentas, kažkam teikiamas pirmumas ir pan.
- *Laikas*. Laikas, kuriuo analizuojamos tam tikros problemos, gali turėti įtakos, jog bus pateikiamos skirtingos išvados apie tą patį objektą. Vienu metu viskas gali atrodyti daug geriau nei kitu. Pavyzdžiui, padidėjus ar sumažėjus obligacijų pajamingumui, o kartu –

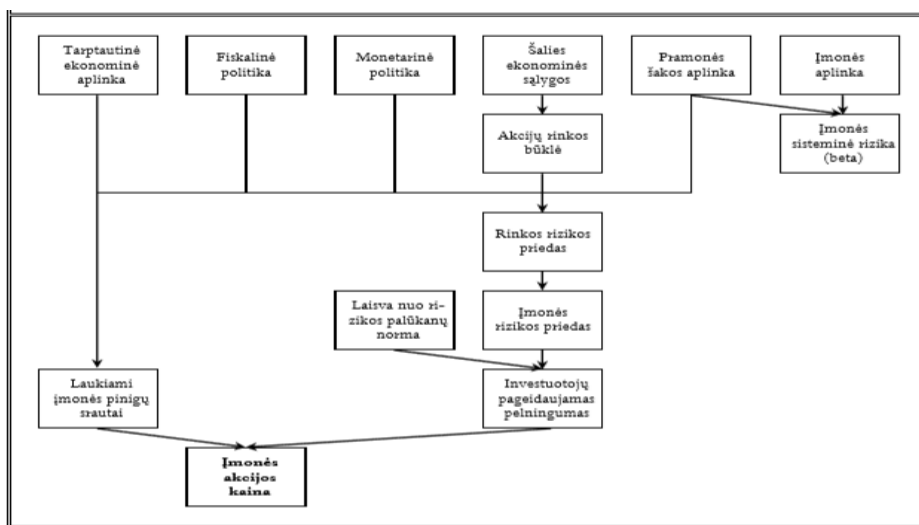
sumažėjus ar padidėjus jų kainai, investuotojų nuotaikos gali būti skirtingos, todėl tinkamas laiko pirkti ir parduoti įvertinimas yra labai svarbu.

- *Kaina*. Kaina yra neatskiriama kiekvieno sprendimo dėl vertybinių popierių dalis. Tai ypač svarbu akcijų atveju, nes dažnai išskyla pavojus, kad bus perkama ar parduodama pervertinta arba neįvertinta kaina.
- *Vertybinio popieriaus rūšis*. Tam tikro finansinio instrumento pasirinkimo už tam tikrą kainą problema gali būti išspręsta pakeitus formą: vietoj klausimų „Koks vertybinis popierius?“ ir „Už kokią kainą?“, galime paklausti „Kokioje įmonėje?“ ir „Kokios pasiūlyto įsipareigojimo sąlygos?“. Tai suteikia daugiau išsamumo ir subalansavimo.

Apibrėžus visus minėtus elementus, autorių teigimu, galėtų būti iškeltas apibendrinantis klausimas: „Ar turėtų tam tikras vertybinis popierius *VP* būti perkamas už kainą *K*, laiku *L*, asmens *A*?“.

Toje pačioje knygoje Graham‘as ir Dodd‘as [8] aprašė kiekybinius ir kokybinius faktorius, turinčius didelės reikšmės atliekant vertybinių popierių analizę. Pirmajai kategorijai priskiriami įmonės statistiniai rodikliai, o kokybiniais gali būti pavadinami tokie veiksniai kaip verslo pobūdis; tam tikri ūkio šakos, kurioje veikia įmonė, aspektai; fizinės, geografinės charakteristikos; valdymo ypatumai; sektoriaus, pramonės ir, apskritai, verslo perspektyvos.

Bagdonas ir Klimašauskas [9] parodo schematiškai, kokie veiksniai gali daryti poveikį įmonės akcijų kainos kitimui (žr. 2 pav.).



2 pav. Akcijų kainos kitimui įtaką darantys veiksniai

Graham‘o ir Dodd‘o [8] pastebėjimai apie vertybinių popierių kainai įtakos turinčius veiksnius buvo bene, pirmieji. Kiek vėliau Chen‘as, Roll‘as, Ross‘as [10] bei Bodurtha, Cho, Senbet‘as [11], vėlgi vieni pirmųjų, analizavo makroekonominių veiksnių poveikį akcijų rinkoms. Tai buvo du, vienas kitą papildantys tyrimai, kurių metu nustatyta, jog egzistuoja reikšmingas ryšys tarp akcijų rinkų ir įvairių makroekonominių jėgų.

Anot lietuvių tyrėjų Boreikos ir Pilinkaus [12], akcijų kainų ryšys su makroekonomika yra tiesioginis. Šį ryšį autoriai pagrindžia teigdami, jog makroekonominė aplinka turi įtakos vartotojų elgsenai, kuri daro poveikį vertybinių popierių rinkai. Pavyzdžiui, padidėjus kainų lygiui, sumažėja vartojimas, tai, savo ruožtu, mažina įmonių pajamas ir kartu gali sąlygoti padidėjusį nedarbą bei

kitus makroekonominis rodiklius, o visa tai ne itin teigiamus investuotojų lūkesčius ir akcijų rinkų kitimo tendencijas.

Dauguma tyrėjų visame pasaulyje, atlikusių tyrimus šioje srityje, analizavo būtent makroekonominių rodiklių ir įvairių šalių įmonių akcijų kainų ar akcijų rinkų indeksų tarpusavio ryšius. Keleto naujausius tyrimus atlikusių autorių darbų susistemimas pateikiamas 1 lentelėje, nurodant analizuotą šalį ar kelias šalis, pasirinktus makroekonominis rodiklius ir taikytus tyrimo metodus. Apžvelgus minėtus duomenis pastebėta, jog tarp tyrimuose taikomų makroekonominių rodiklių dominuoja bendrasis vidaus produktas (toliau BVP), infliacija, palūkanų norma, valiutų kursas, piniginės lėšos, pramonės gamybos indeksas, tiesioginės užsienio investicijos. Mariappan'as, Hari ir Jyotishi [14] įtraukė kelis kiek kitokius rodiklius – interneto vartotojų skaičių, realaus ir efektyvaus valiutų kurso indeksą, akcijų apyvartą. Jareno ir Negrut'as [17], analizavę įtaką Jungtinių Amerikos Valstijų akcijų rinkai, savo tyrime pritaikė nedarbo lygio duomenis, o kai kurių Azijos šalių tyrėjai be keleto jau minėtų veiksnių taip pat panaudojo naftos ir aukso kainos rodiklius.

Beveik visų atliktų analizių metu daugiau ar mažiau taikyti vienodi ekonominiai tyrimo metodai: aprašomoji statistika, stacionarumo tikrinimo testai (ADF ir PP), Granger'io priežastingumo ryšio nustatymo testas, Johansen'o ko-integracinis metodas, koreliacinė analizė, regresijos modelių sudarymas, Vektorių klaidų koregavimo modelis (VECM). Kiek daugiau „naujovių“ įnešė Ndlovu, Faisal'as, Resatoglu, et al. [23], pasiūlydami dispersijos išskaidymo metodo taikymą (VDC) ir impulso atsako funkcijų (IRF) sudarymą.

Apžvelgus minėtų tyrimų rezultatus pastebėta, kad visi Inumula'o, Tadamarla'o, Deeppa'o [22], Chauque'o, Rayappan'o [21], Hunjra'o, Chani, Shahzad'o, et al. [15], Ndlovu, Faisal'o, Resatoglu, et al [23] pasirinkti makroekonominiai kintamieji turi įtakos analizuotų atskirų įmonių akcijų kainų ar akcijų rinkų indeksų kitimui. Duy'o [18] analizės atveju akcijų rinkos indekso pokyčiui jokios įtakos neturėjo pramonės gamybos indeksas ir mažmeninė naftos kaina, John'o ir Duke'o II [13] atveju - investicijos, Ouma'o ir Muriu [16] atveju - palūkanų norma, o Jareno ir Negrut'o [17] tyrimo metu nenustatyta ryšio tarp akcijų rinkos indekso ir infliacijos. Ligocka, Pražak'as, Stavarek'as [19] savo darbe nurodo, jog tarp penkių (iš šešių analizuotų) kompanijų akcijų kainų ir kai kurių makroekonominių rodiklių nustatyti statistiškai reikšmingi ryšiai, tačiau iš jų nebuvo nė vienos, kurių ryšiai sietų su visais faktoriais. Minėtos vienos kompanijos akcijų kainai įtakos neturėjo nė vienas kintamasis. Kabeer'as [20] nurodo, jog savo tyrimo metu pavyko nustatyti visų analizuotų kintamųjų statistiškai reikšmingą nevienodą poveikį pasirinktų šalių akcijų rinkoms, t. y. vienu šalių atveju ryšiai buvo su vienais kintamaisiais, kitų – su kitais. Apibendrinant galima teigti, jog labiausiai atskirų įmonių akcijų kainų ar akcijų rinkų indeksų pokyčius lemia tokie makroekonominiai rodikliai kaip BVP, piniginės lėšos, valiutų kursas, palūkanų norma ir infliacija.

Visų minėtų darbų autoriai atsižvelgia tik į makroekonominis veiksnis, kurie, žinoma, yra svarbūs, tačiau nereikėtų pamiršti, kad akcijų kainų ar akcijų rinkų indeksų pokyčiams įtaką taip pat daro atskirų ūkio šakų duomenys, įmonių finansiniai rodikliai, politiniai, emociniai bei kiti kintamieji. Būtent į tai savo darbuose atkreipė dėmesį Chen'as, Kim'as, Kim'as [24], Zopiatis, Savva, Lambertides'as, et al. [25], Cheng'as, Tzeng'as, Kang'as [26], kurie, be kitų rodiklių, tyrė politinių įvykių, stichinių nelaimių, didelių sporto renginių, finansų krizių, teroristinių atakų, karinių konfliktų įtaką akcijų rinkoms.

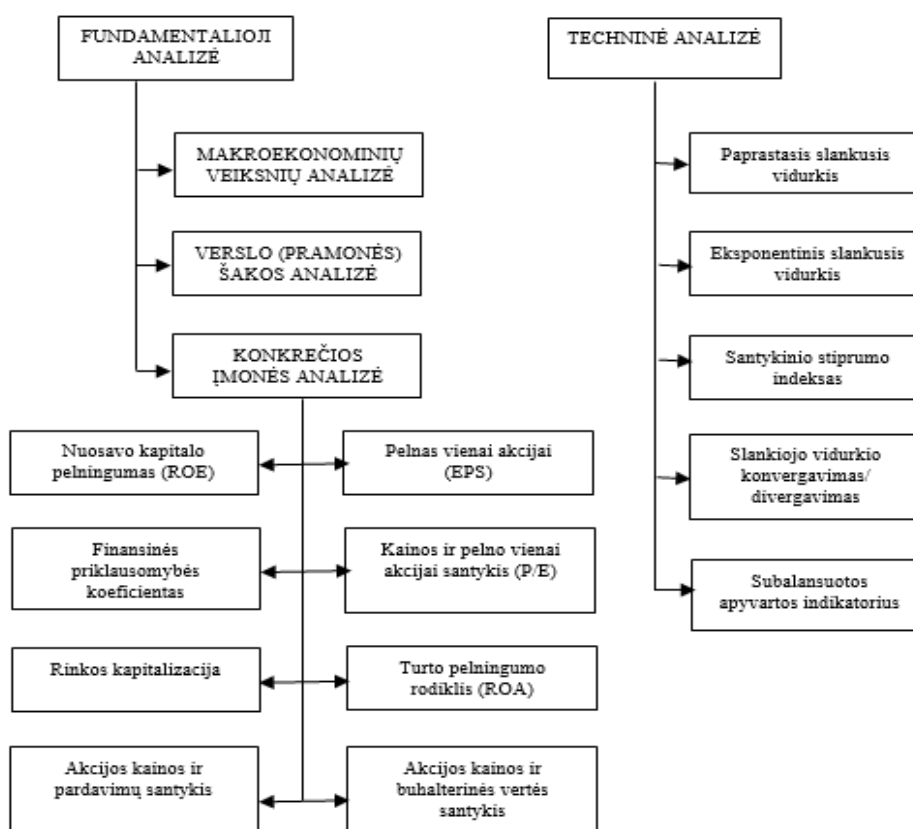
1 lentelė. Akcijų ir makroekonominių rodiklių ryši analizavusių tyrėjų darbų apibendrinimas

Eil. Nr.	Šaltinis	Šalis	Makroekonominiai rodikliai	Taikyti metodai
1.	John, Duke II (2013)	Nigerija	BVP, finansų sektoriaus plėtra, akcijų rinkos likvidumas, infliacija, vidaus taupymo rodiklis, investicijos	Stacionarumo nustatymo testas (ADF), liekamųjų paklaidų ko-integracinis regresijos modelis
2.	Mariappan, Hari, Jyotishi (2013)	24 šalys iš viso pasaulio	Palūkanų norma, interneto vartotojų skaičius, BVP, tiesioginės užsienio investicijos, piniginės lėšos, akcijų apyvarta, realaus ir efektyvaus valiutų kurso indeksas	Tiesinės regresijos modelis
3.	Hunjra, Chani, Shahzad, et al. (2014)	Pakistanas	Palūkanų norma, valiutų kursas, infliacija, BVP	Aprašomoji statistika, stacionarumo tikrinimo (ADF) ir Granger'io prieš. ryšio nustatymo testai, Johansen'o ko-integracinis metodas
4.	Ouma, Muriu (2014)	Kenija	Piniginės lėšos, palūkanų norma, infliacija, valiutų kursas	Regresijos modelis, stacionarumo nustatymo testas (ADF)
5.	Jareno, Negrut (2016)	JAV	BVP, infliacija, pramonės gamybos indeksas, palūkanų norma, nedarbo lygis	Pearson'o koreliacijos koeficiento taikymas, taškų sklaidos diagramos
6.	Duy (2016)	Vietnamas	Palūkanų norma, piniginių lėšos, valiutų kursas, pramonės gamybos indeksas, infliacija, mažmeninė naftos kaina, aukso kainos indeksas	Stacionarumo nustatymo testai (ADF ir PP), Johansen'o ko-integracinis metodas, Vektorių klaidų koregavimo modelis (VECM)
7.	Ligocka, Pražak, Stavarek (2016)	Šveicarija	BVP, palūkanų norma, infliacija	Aprašomoji statistika, stacionarumo nustatymo testai (ADF ir PP), koreliacinė analizė, Johansen'o ko-integracinis metodas, Vektorių klaidų koregavimo modelis (VECM), Granger'io priežastingumo ryšio nustatymo testas
8.	Kabeer (2017)	Bangladešas, Indija, Pakistanas, Šri Lanka, Kinija	Užsienio valiutų kursas, tiesioginės užsienio investicijos, infliacija	Aprašomoji statistika, Pearson'o koreliacijos koeficiento taikymas, regresijos modelis
9.	Chauque, Rayappan (2018)	Malaizija	Valiutų kursas, infliacija	Stacionarumo nustatymo testai (ADF ir PP), heteroskedastiškumo ir Granger'io priežastingumo ryšio nustatymo testai, daugialypės regresijos modelis
10.	Inumula, Tadamarla, Deeppa (2018)	Indija	Neapdirbtos naftos kaina, Indijos <i>Nifty</i> indeksas, pramonės gamybos indeksas, didmeninės kainos indeksas, valiutų kursas	Aprašomoji statistika, Karlpearson'o koreliacijos koeficiento taikymas, daugialypės tiesinės regresijos modelis
11.	Ndlovu, Faisal, Resatoglu, et al. (2018)	Pietų Afrikos Respublika	Piniginės lėšos, infliacija, palūkanų norma, valiutų kursas	Regresijos modelis, stacionarumo (ADF) ir Granger'io priežastingumo ryšio nustatymo testai, Johansen'o ko-integracinis metodas, VECM, Dispersijos išskaidymo metodas (VDC), Impulso atsako funkcijos (IRF)

Chen'as, Kim'as, Kim'as [24] ištyrė, kad dideli sporto renginiai ir politiniai įvykiai turėjo įtakos viešbučių akcijų gražų kitimui, o Zopiatis, Savva, Lambertides'as, et al. [25] atlikę kitą tyrimą nustatė, jog netikėti ne makroekonominiai įvykiai iššaukė trumpalaikį poveikį analizuojamiems akcijų rinkų indeksams, o grįžimas į pirminę padėtį įvykdavo per 2-3 dienas. Nurodoma, kad pastebėti skirtumai pagal incidento tipą: teroristiniai išpuoliai ir stichinės nelaimės sukėlė

reikšmingiausių pokyčių. Pirmųjų atveju tai įvykdavo kitą dieną po įvykio, o stichinių nelaimių atveju – tą pačią dieną. Nustatyta, kad kariniai konfliktai turėjo mažai įtakos. Taip pat pastebėta, kad tik Jungtinėse Amerikos valstijose įvykę incidentai turėdavo globalų poveikį beveik visiems indeksams, o kiti labiausiai paveikdavo regionines akcijų rinkas. Cheng'as, Tzeng'as, Kang'as [26] taip pat nustatė, kad nemažai ne makroekonominių įvykių turėjo reikšmingos įtakos *Taivano elektronikos* akcijų grąžai.

Mokslinėje literatūroje taip pat nurodoma, kad, apskritai, išskiriamos dvi vertybinių popierių, o kartu ir akcijų rinkos, analizės mokyklos: *fundamentali* bei *techninė*. Kiti autoriai tiesiog aprašydavo, kokie veiksniai, kokiai mokyklai priskiriami, o Isidore ir Christie [27] savo darbe, kuriame lygina šias dvi analizės rūšis, pateikia schematiškai, kokie rodikliai sudaro kiekvieną jų (žr. 3 pav.).



3 pav. Fundamentaliosios ir techninės analizės sudėtinės dalys

Fundamentali

Teigiama, kad šią analizę taiko nemaža dalis investuotojų, investuojančių ilgajam laikotarpiui. Metodo logika yra tokia: jei įmonės ilgalaikiai finansiniai rodikliai geri, jei ją supa palanki verslo aplinka, tada labai tikėtina, jog *ilgalaikės* akcijos perspektyvos taip pat bus geros ir tokia investicija yra teigiamas pasirinkimas. Fundamentaliai „stiprios“ akcijos savyje turi potencialo generuoti teigiamą grąžą: šios kompanijos uždirba pinigus, didina pelną ir augina savo verslą. Vis dėlto fundamentaliosios analizės problema yra tokia, jog ji neparodo *dabartinės* akcijų būklės. „Gera kompanija“ ir „gera akcija“ nėra tas pats. Įmonė gali uždirbti labai daug pinigų, atlikti nuostabų darbą, didinant pajamas ir pelną, tačiau jos akcijos kaina trumpuoju laikotarpiu smuks žemyn [28].

Autoriai Petrusheva ir Jordanoski's [29] taip pat išdėsto panašias mintis, teigdami, jog atliekant fundamentaliąją analizę galima bandyti nustatyti akcijos vertę ateityje, atsižvelgiant į įvairius ekonominius kintamuosius, vadinamus fundamentaliaisiais. Analizė susideda iš visos ekonomikos, šakos, kurioje įmonė veikia, ir pačios įmonės rodiklių nagrinėjimo. Anksčiau minėtoje schemoje (žr. 3 pav.) nurodyti įmonių rodikliai, kuriuos būtų galima panaudoti atliekant fundamentaliąją analizę, dažnai jau būna pateikti įmonių finansinėse ataskaitose ir jų skaičiuoti patiems nereikia, todėl čia platesnis šių rodiklių apibūdinimas nepateikiamas.

Techninė analizė

Investuotojai, kurie remiasi technine analize, visai nekreipia dėmesio į fundamentaliuosius veiksnius. Apskritai, kai kurie jų susikoncentruoja tik į diagramas, grafikus, parodančius pirkimus bei pardavimus, ir net negalėtų pasakyti, kokio verslo akciją analizuoja. Išimtinai remiantis tik technine analize ir siekiant identifikuoti investavimo galimybes, mažai atsižvelgiama į riziką, susijusią su konkrečia įmone [28].

Anot Petrusheva's ir Jordanoski'o [29], ši analizė grindžiama prielaida, kad visi fundamentalieji veiksniai jau atspindi akcijų kainą. Techninės analizės metu orientuojamasi tik į akcijų kainą ir prekybos apimtį, kurios duomenys pateikiami lentelių ir grafikų pavidalu. Analizuodami šias diagramas, investuotojai gali gauti informaciją apie tendą ir dėsningumą kainų pokyčiuose, prekybos apimtis ir tarpusavio priklausomybę. Kitaip tariant, čia analizuojami istoriniai duomenys tikint, kad istorija kartojasi.

Young'as [30] apibūdino keletą svarbių techninių/matematinių rodiklių (nurodytų ir pateiktoje schemoje), naudojamų atliekant techninę analizę:

- *Paprastasis slankusis vidurkis* (angl. *Simple moving average (SMA)*). Šis rodiklis taikomas prognozuoti akcijų kainų tendą ir pasikeitimo laiką. Čia pagrindinis akcentas yra „judanti“ vidutinė linija, kuri ir taikoma prognozėms: jei, žiūrint grafiškai, paprastojo slankiojo vidurkio linija nukrenta (pakyla) virš akcijų kainos linijos, tai parodo, kad tikėtina, jog akcijų kaina kils (kris). Linijos kryptis nurodo dabartinės tendencijos kryptį. Ji gali būti naudojama trumpo ir ilgo laikotarpio prognozėms.
- *Eksponentinis slankusis vidurkis* (angl. *Exponential moving average (EMA)*). Rodiklis panašus į prieš tai aptartąjį, tačiau šio skirtumas toks, jog čia daugiau „svorio“ suteikiama kelioms paskutinėms buvusioms dienoms, o ankstesnis laikotarpis mažiau svarbus. Šis įrankis yra „galingesnis“ nei paprastasis slankusis vidurkis.
- *Santykinio stiprumo indeksas* (angl. *Relative strength index (RSI)*). Indeksas apskaičiuojamas pagal tokią formulę:

$$100 - (100 / (1 + (\text{aukščiausių dienos kainų vidurkis} / \text{žemiausių dienos kainų vidurkis}))) \quad (3)$$

Šis rodiklis „seka“ kainą ir gali įgyti reikšmes nuo 0 iki 100. Didelė jo reikšmė parodo, kad akcija buvo pervertinta, o maža – jog akcija buvo neįvertinta.

- *Slankiojo vidurkio konvergavimas/ divergavimas* (angl. *Moving Average Convergence/Divergence (MACD)*). Tai yra impulso indikatorius, parodantis dviejų slankiųjų vidurkių sąryšį. Čia išvedamos dvi linijos: *MACD*, kuri gaunama iš 12 dienų eksponentinio slankiojo vidurkio atėmus 26 dienų eksponentinį vidurkį, ir *signalo linija*,

kuri yra 9 dienų eksponentinio vidurkio linija. Kai *MACD* linija pasislenka aukščiau (žemiau) *signalo linijos*, išryškėja trendo aukštėjimo (žemėjimo) pradžia.

- *Subalansuotos apyvartos indikatorius* (angl. *On Balance Volume (OBV)*). Tai dar vienas impulso indikatorius, kuris skaičiuojamas pridėdant kasdienės prekybos apyvartos, kai akcijų kaina kilo, ir atimant kasdienės apyvartos, kai kaina krito. Smunkanti *OBV* linija parodo, kad akcijų kainos taip pat sumažės, o kylanti – jog akcijų kainos taip pat didės.

Čia aprašyti, bene, svarbiausi techniniai rodikliai, kurie yra labai naudingi, analizuojant istorinius akcijų kainų duomenis. Vis dėlto yra ir kitų, kurie verti paminėti: *Bollinger'io juostų indikatorius* ir *stochastinis osciliatorius*.

- *Bollinger'io juostų indikatorius* (angl. *Bollinger Bands Indicator*). Šis indikatorius yra vienas iš kintamumo rodiklių. Jis parodo, kada rinka yra pasyvi, o kada joje aktyviai vyksta prekyba. Kai rinka yra pasyvi, *Bollinger'io juostos* tampa siauros (mažas kintamumas), kai pagyvėja - juostos prasiplečia (didelis kintamumas). Kuo kintamumas didesnis, tuo investavimas į tokias akcijas tampa rizikingesnis. Kainos šalia viršutinės juostos yra aukštos, šalia apatinės – žemos. *Bollinger'io juostų grafike* kainas atspindi stačiakampiais, kurie simbolizuoja kainų kritimą arba kilimą, „aplipdyta“ diagrama. Vidurinė juosta rodo slankiojo vidurkio (rekomenduojama 20 periodų) judėjimą, viršutinė juosta yra vidurinė juosta, „pereinanti“ į viršų tam tikrą standartinių nuokrypių skaičių (dažniausiai 2), apatinė – vidurinė juosta, „pereinanti“ žemyn tam tikrą standartinių nuokrypių skaičių (dažniausiai 2). Standartinis nuokrypis yra kintamumo matas, kuris apibrėžia juostų plotį. Aukštas kintamumas iššaukia žemą kintamumą, o žemas – aukštą [31].
- *Stochastinis osciliatorius* (angl. *Stochastic Oscillator / Stochastic Momentum Index*). Šį rodiklį galima apibūdinti taip: kainoms didėjant, akcijų uždarymo kainos būna arčiau kainų intervalo viršutinės dalies, ir, atvirkščiai, nuosmukio metu uždarymo kainos būna arčiau kainų intervalo apatinės dalies. Stochastiniame procese naudojamos dvi linijos: %K linija ir %D linija. %D linija yra svarbesnė ir ji iššaukia pagrindinius signalus. Esminis stochastinio osciliatoriaus tikslas yra nustatyti, kur paskutinė uždarymo kaina siejasi su pasirinkto laikotarpio kainų intervalu. Keturiolika čia yra bene dažniausiai taikomas akcijų kainų periodas [32].

Fundamentaliąją analizę atlikti yra sudėtingiau nei techninę, taip pat tai užima daugiau laiko, tačiau ji labai naudinga tada, kai investuojama ilgesniam laikotarpiui ir norima bent apytiksliai prognozuoti akcijų kainas ateityje. Techninė analizė labai naudinga siekiant stebėti pokyčius, tendencijas trumpuoju laikotarpiu ir priimti kiek galima efektyvesnius sprendimus „čia ir dabar“.

Apibendrinant poskyrio informaciją galima teigti, jog akcijų kainų ar akcijų rinkų indeksų pokyčiams įtakos turi ne vienas ir net ne keli, o keliolika skirtingų faktorių iš įvairių sričių. Jie gali būti makroekonominiai, apibūdinantys tam tikrą sektorių, konkrečią įmonę, paremti investuotojų emocijomis ir panašiai. Apžvelgus kitų autorių atliktus tyrimus šioje srityje pastebėta, jog labiausiai skirtingų įmonių akcijų kainų ar akcijų rinkų indeksų pokyčius lemia tokie makroekonominiai rodikliai kaip BVP, piniginių lėšos, valiutų kursas, palūkanų norma ir infliacija. Buvo ir kita dalis tyrėjų, kurie pastebėjo, jog politiniai įvykiai, stichinės nelaimės, dideli sporto renginiai, finansų krizės, teroristinės atakos, kariniai konfliktai taip pat daro nemažą įtaką šių vertybinių popierių pokyčiams. Apskritai, atliekant akcijų kainų ar akcijų rinkų indeksų kitimo analizę galima remtis

dviem vertybinių popierių rinkos analizės mokyklomis: fundamentaliają bei techninę. Pirmoji naudinga, kai investuojama ilgesniu laikotarpiu, o antroji – trumpuoju.

Kaip jau daugiau ar mažiau aptarta, daugelį, o gal net ir visus investuotojus dominantis aspektas – akcijų kainų pokyčiai ateityje, jų prognozavimas, nuspėjamumas. Vieni tyrėjai bando kainų reikšmes prognozuoti remdamiesi istoriniais laiko eilutės duomenimis (techninės analizės sritis), tačiau kiti tam prieštarauja ir teigia, jog akcijų kainų tokiu būdu tiksliai prognozuoti neįmanoma, kadangi, kaip minėta anksčiau, veikia daug išorės ir vidinių faktorių, dėl kurių nuolatinio kintamumo gali būti labai sudėtinga tai padaryti. Net jei prognozuojama ne vien tik pagal istorinius laiko eilutės duomenis, o bandoma ateities kainas nuspėti pagal faktorių poveikį ir pokyčius, susijusius su jais (fundamentaliosios analizės sritis), vis tiek galima gauti tik apytiksles, o ne tikslias reikšmes.

Todėl kiti tyrėjai visame pasaulyje vis dažniau stengiasi atrasti būdus, kurie galėtų padėti kiek galima tiksliau prognozuoti kainų kintamumą diena į priekį ar truputį ilgesniu laikotarpiu, kai norima rasti ne tikslias reikšmes, o apskritai nustatyti ar kaina kils, ar nekils. Dėl to kitame poskyryje apžvelgiami kai kurių autorių darbai, susiję su įmonių akcijų kainų ar akcijų rinkų indeksų kitimo prognozavimu, taikant mašininio mokymosi (angl. *Machine Learning*) metodus. Čia labiau gilinantis ne į klausimą „Kokia bus kaina?“, bet „Kaip kaina keisis?“

1.4. Mašininio mokymosi metodų taikymas prognozuojant akcijų kitimą

Klasifikavimas, o, kartu ir prognozavimas, panaudojant mašininio mokymosi algoritmus, plačiai taikomas daugelyje gyvenimo sričių: medicinoje, kai norima prognozuoti (ne)susirgimo vienokia ar kitokia liga tikimybę, kredito institucijų veikloje, kai norima, nustatyti, ar bus gražinama paskola, ar ne, ir panašiai. Finansų rinkos taip pat nėra išimtis. Čia daug dėmesio skiriama akcijų kainų kitimo prognozavimui, kadangi investuotojams yra labai svarbu bent kažkiek galėti numatyti ateities įvykius, t. y. galėti priimti kiek galima palankesnius sprendimus tam, kad nebūtų patiriami dideli nuostoliai, arba, atvirkščiai, gaunama kiek galima daugiau pelno.

Klasių prognozavimo modelių kūrimas susideda iš trijų pagrindinių dalių:

1. numatymo, ką norima prognozuoti, ir duomenų matricos sudarymo su reikiamais kintamaisiais;
2. klasifikavimo algoritmų pasirinkimo tam, kad būtų galima įgyvendinti prognozavimo užduotį;
3. vertinimo tam, kad būtų įvertintas sudarytų modelių tikslumas, tinkamumas ir prognozių patikimumas.

Labai informatyvų darbą šia tema parašė ir nemažą tyrimą atliko Jungtinių Amerikos Valstijų tyrėjas Weng'as [33], kuris sudarė įmonės *Apple* akcijų kainos kitimo diena į priekį prognozavimo modelius. Pasirinktas istorinis akcijų kainos analizės laikotarpis – 3 metai. Duomenų matrica sudaryta iš tokių duomenų blokų: 1) viešai prieinamos *Apple* akcijų kainos informacijos, kuri apima dieninę atidarymo, uždarymo, aukščiausią, žemiausią kainą ir prekybos apimtį (angl. *trade volume*), kainos ir pelno vienai akcijai santykio (P/E), kaip kompanijos „sveikatą“ apibūdinančio rodiklio, bei *NASDAQ* ir *DJIA* indeksų; 2) dažniausiai naudojamų techninių indikatorių, kurie atspindi kainų pokyčius laikui bėgant. Tai yra: stochastinis osciliatorius (%K), Williams'o %R

rodiklis, arba kitaip Williams'o procentinė sritis, ir santykinio stiprumo indeksas (RSI); 3) kasdienių naujienų apie dominančią akciją iš *Google News* skaičiavimų; 4) unikalių lankytojų, apsilankiusių susijusiuose *Wikipedia* puslapiuose, skaičius per dieną.

Po to, panaudojant SVM rekursinį požymių pašalinimo metodą (angl. *Recursive Feature Elimination (RFE)*), atrinkti tie kintamieji, kurie gali turėti didžiausios įtakos prognozavimo tikslumui. Modeliai sudaryti pritaikant tris mašininio mokymosi algoritmus: neuroninius tinklus (angl. *Artificial Neural Networks (ANN)*), atraminių vektorių metodą (angl. *Support Vector Machine (SVM)*) ir sprendimų medį (angl. *Decision Tree (DT)*). Šie modeliai, įgyvendinus 10 kartų kryžminį patikrinimą (angl. *10-fold cross validation*), lyginami ir vertinami pagal ploto po ROC kreive reikšmę AUC (pagrindinis tikslumo vertinimo metodas) bei septynis kitus matus.

Prieš atliekant klasifikavimą/prognozavimą, autorius iškėlė penkis vienos dienos į priekį tikslus, kurie gali būti aktualūs investuotojams: 1) dabartinės prekybos dienos akcijų *atidarymo kainos* ir praėjusios prekybos dienos *uždarymo kainos* palyginimas; 2) dabartinės dienos akcijų *atidarymo kainos* ir praėjusios dienos *atidarymo kainos* palyginimas; 3) ir 4) atitinka panašią logiką kaip pirmi du tik čia dabartinę dieną apibūdina *uždarymo kaina*, o praėjusią prekybos dieną – *atidarymo kaina*; 5) dabartinės dienos akcijų *sandorių skaičiaus* (prekybos apimčių) ir praėjusios dienos *sandorių skaičiaus* palyginimas. Visi šie tikslai transformuoti į dvejetainę išraišką: 0 reiškia, jog nekils, o 1 reiškia, jog reikšmė kils, palyginus su praėjusia diena. Visa tai, kas parašyta, galima susisteminti lentelės pavidalu (žr. 2 lentelę).

2 lentelė. Kitimo prognozavimo viena diena į priekį tikslai pagal Weng'ą [33]

Tikslai	Formulė
1 tikslas	$Atidarymo(i+1) - Uždarymo(i)$
2 tikslas	$Atidarymo(i+1) - Atidarymo(i)$
3 tikslas	$Uždarymo(i+1) - Uždarymo(i)$
4 tikslas	$Uždarymo(i+1) - Atidarymo(i)$
5 tikslas	$Prekybos\ apimtys(i+1) - Prekybos\ apimtys(i)$

Autoriaus gauti rezultatai parodė, jog aukščiausios sudarytų modelių AUC reikšmės buvo apie 0,89, kai taikyti *SVM* ir *DT* klasifikavimo metodai 2-4 tikslų atveju. O tikslumas (angl. *accuracy*) siekė apie 85 %.

Dar vieną panašų tyrimą atliko Kim'as [34], kuris teigia, jog atraminių vektorių metodas (*SVM*) yra perspektyvus algoritmas prognozuojant finansines laiko eilutes. Jis savo darbe taikė tik techninius indikatorius, kuriuos pasitelkiant prognozavo Korėjos indekso *KOSPI* kitimą viena diena į priekį. Techninių indikatorių iš viso pasirinkta 12.

Čia taip pat buvo sudarytas tikslo stulpelis, kai 0 reiškia, jog kitos dienos indekso reikšmė bus mažesnė nei dabartinės dienos, o 1 reiškia, jog indekso reikšmė kitą dieną bus didesnė. Iš viso pasirinktos 2928 prekybos dienos (10 metų dieniniai duomenys).

Sudarius modelius su skirtingais parametrais ir gavus rezultatus paaiškėjo, jog aukščiausia prognozavimo tikslumo reikšmė, taikant *SVM* metodą, buvo tik apie 65 %. Autoriaus manymu, modelių tikslumą galima pagerinti radus optimalius *SVM* parametrus arba analizei pasirinkus kitokias laiko eilutes.

Palyginus su anksčiau aprašytu tyrimu, šio metu sudaryto geriausio modelio tikslumo rezultatas buvo apie 20 % žemesnis. Vadinasi, viena iš priežasčių, kodėl sudarytų prognozavimo modelių tikslumas pagerėja, galėtų būti tai, jog kintamieji pasirenkami iš įvairių duomenų šaltinių ir optimalių parametrų taikymas čia nebeturi itin didelės reikšmės.

3 lentelėje pateikiama susisteminta jau aprašytų ir kelių kitų tyrimų informacija.

3 lentelė. Akcijų kainų kitimo prognozavimą atlikusių tyrėjų darbų apibendrinimas

Eil. Nr.	Šaltinis	Duomenų šaltiniai	Taikyti metodai	Tikslumas
1.	Kim (2003)	Techniniai indikatoriai	SVM	Tikslumas – 65 %
2.	Al-Radaideh, Assaf, Alnagi (2013)	Prekybos duomenys	DT	Tikslumas – 55 %
3.	Leung, MacKinnon, Wang (2014)	Finansiniai įmonės duomenys	SSVM	Tikslumas – 59 %
4.	Lee, Ryu, Lee (2017)	Naujienų skaičius iš svetainės <i>NAVER</i>	LR (<i>logistinė regresija</i>), SVM, ANN, DT	Tikslumas – 72 %
5.	Weng (2017)	Prekybos duomenys, techniniai indikatoriai, <i>Wikipedia</i> lankytojų skaičius, <i>Google</i> naujienų skaičius, kiti sugeneruoti kintamieji	ANN, SVM, DT	AUC – 0,89 Tikslumas – 85 %

Iš lentelėje pateiktų duomenų akivaizdžiai matoma, jog aukščiausias sudarytų prognozavimo modelių tikslumas pasiektas tada, kai kintamieji pasirinkti iš įvairių duomenų šaltinių. Taip pat galima teigti, kad didelę reikšmę, sudarant tokius modelius, turi duomenys, susiję su naujienomis, kadangi abiem atvejais, kai tai taikyta, tikslumas buvo aukštas (85 % ir 72 %). Labiausiai nepasiteisino atvejis, kai prognozuojant pasirinkti tik akcijų prekybos duomenys (uždarymo, atidarymo, mažiausia kaina ir kt.) ir sprendimų medžio klasifikatorius (*DT*), kadangi tada gautas itin žemas tikslumas – 55 %.

Weng'as [33] atliko išsamesnę tokio pobūdžio tyrimų apžvalgą ir teigė, jog kitų tyrėjų sudarytų modelių tikslumas nesiekdavo 83 %.

Apžvelgus kitų šalių autorių įgyvendintus tyrimus, verta labiau pasigilinti, kokie tyrimai atlikti, susiję su Baltijos šalių įmonių akcijų kainų ar akcijų rinkų indeksų kitimo prognozavimu, bei kokie rezultatai gauti čia.

1.5. Akcijų Baltijos šalyse kitimo tyrimų apžvalga

Pasaulyje dažnai analizuojamos garsiausių kompanijų akcijos bei svarbiausi akcijų rinkų indeksai, kadangi investuotojams labai naudinga ir reikalinga bent kažkiek numatyti jų pokyčius ateityje. Ne mažiau svarbios, jei investuojama čia, yra ir Baltijos šalių akcijų rinkų tendencijos bei įvairių įmonių akcijos. Dėl šios priežasties toliau apžvelgiama, kokie Lietuvos, Latvijos ir Estijos vertybinių popierių prognozavimo tyrimai atlikti kitų tyrėjų, kokios problemos iškyla, ir ką būtų galima padaryti kitaip.

Akcijų kainų prognozavimo tyrimą įgyvendino Grigoryan'as [35], kuris, pritaikydamas neuroninių tinklų metodą, atliko vienos iš Estijos įmonių akcijų uždarymo kainos prognozavimą. Istorinių

duomenų laikotarpis – 34 mėnesiai. Pasirinkti dviejų rūšių kintamieji: akcijų prekybos rodikliai (uždarymo, atidarymo, aukščiausia, žemiausia kaina, prekybos apimtys, apyvartos duomenys) bei 30 techninių indikatorių (visi duomenys dieniniai).

Tam, kad prognozavimo metu būtų naudojami tik reikšmingi požymiai, kintamieji atrenkami pritaikant pagrindinių komponenčių analizę (angl. *Principal Component Analysis (PCA)*). Taip iš 36 kintamųjų liko 10, suteikiančių daugiausiai informacijos apie pagrindinį duomenų rinkinį.

Čia nebuvo prognozuojama diena į priekį, kaip ankstesnių tyrimų metu, ar atliekamas įprastas klasifikavimas. Buvo stebima, kaip neuroninių tinklų metodą būtų galima pritaikyti prognozuojant finansinių laiko eilučių konkrečias reikšmes pasirinktu laiko horizontu, kai sudaryto modelio rezultatai palyginami su tikrosiomis reikšmėmis. Gauti rezultatai parodė, kad pasirinktas metodas ir taikyti duomenys yra tinkami atlikti tokias akcijų kainų prognozes, nes paklaida (*MSE* reikšmė) gauta maža.

Dar vieną tokio tipo tyrimą atliko Maksvytienė ir Safonovas [36]. Jie tyrė ekonometrinių modelių pritaikymo galimybes, prognozuojant Lietuvos akcijų rinkos indekso (*OMXV*) pokyčius. Čia naudojami mėnesiniai duomenys, pasirenkant paskutinės mėnesio dienos indekso reikšmę, kai bendras laikotarpis – 15 metų. Sudaromi keturi vienfaktoriniai modeliai, naudojant tik indekso *OMXV* duomenis. Taikomi metodai yra: klasikinis laiko eilutės išskaidymas, *Holt* eksponentinis glodinimas, *Holt-Winters* eksponentinis glodinimas ir *ARIMA (1,1,1)*. Kito etapo metu sudaromi daugialypės tiesinės regresijos modeliai, kai vėl pasirinktos *OMXV* indekso reikšmės ir 7 makroekonominių rodiklių duomenys.

Gauti rezultatai parodė, kad, sudarant vienfaktorius modelius, indekso kitimo prognozės ilguoju laikotarpiu yra labai prastos, paklaidos gaunamos didelės, o prognozuojant vienu periodu (vienu mėnesiu) į priekį – paklaidos mažesnės, dėl to ir prognozavimas kiek tikslesnis. Atliekant tiesinę regresinę analizę buvo nustatyta, jog 5 iš 7 makroekonominių rodiklių neturėjo įtakos indekso kitimui. Su likusiais dviem kintamaisiais sudarytas modelis buvo nelabai tinkamas prognozuojant indeksą.

Šie du su prognozavimu susiję tyrimai buvo vieni iš nedaugelio, atliktų analizuojant Baltijos šalių įmonių akcijas ar akcijų rinkų indeksus. Dauguma tyrėjų vertino makroekonominių ir kitų rodiklių poveikį akcijų kainų, indeksų kitimui, įgyvendindami tarpusavio ryšio nustatymo analizę. Tokius tyrimus darė Boreika, Pilinkus [12], Marcišauskienė, Cibulskienė [37], Hsing'as [38], Vaitkė, Martinkutė-Kaulienė [39], Pilinkus [40], Bagdonas, Klimašauskas [9]. Jų darbuose taikyti metodai ir nustatyti akcijų kainų ar akcijų indeksų kitimui įtakos turintys reikšmingi kintamieji pateikiami 4 lentelėje.

Įvertinus lentelėje pateiktus duomenis galima teigti, jog sąveikai nustatyti dažniausiai buvo taikyta koreliacinė regresinė analizė. Autoriai analizavo tiek atskirų, tiek visų trijų Baltijos šalių akcijų rinkas, tačiau rezultatai čia pateikiami apibendrinti ir atskirai pagal šalis neišskiriami. Matoma, jog dažniausiai nustatyti įtaką darantys makroekonominiai rodikliai buvo BVP, pinigų kiekis, valdžios sektoriaus skola, nedarbo lygis, valiutų kursas, tiesioginės užsienio investicijos.

4 lentelė. Įvairių veiksnių poveikį Baltijos šalių akcijoms vertinusių tyrėjų darbų apibendrinimas

Eil. Nr.	Šaltinis	Taikyti metodai	Reikšmingi kintamieji
1.	Bagdonas, Klimašauskas (2005)	Koreliacinė regresinė analizė	Pinigų kiekis, valiutų kursas, nedarbo lygis,
2.	Boreika, Pilinkus (2009)	Koreliacinė regresinė analizė	BVP, pinigų kiekis, nedarbo lygis, individualaus vartojimo išlaidos, statybos sąnaudų kainų indeksas
3.	Pilinkus (2010)	Granger'io priežastingumo ryšio nustatymo testas, vektorių autoregresija, Johansen'o ko-integracijos testas	BVP, eksportas, importas, pinigų kiekis, tiesioginės užsienio investicijos, valdžios sektoriaus skola
4.	Marcišauskienė, Cibulskienė (2013)	Koreliacinė analizė ir daugialypė regresinė analizė	Pinigų kiekis, tiesioginės užsienio investicijos, BVP, nedarbo lygis, valdžios sektoriaus skola
5.	Hsing (2014)	GARCH modelis	Valdžios sektoriaus skola, BVP, Vokietijos akcijų rinkos indeksas, valiutų kursas, palūkanų norma, numatoma infliacijos lygis, euro zonos vyriausybių obligacijų pajamingumas
6.	Vaitkė, Martinkutė-Kaulienė(2018)	Koreliacinė regresinė analizė	Ekonominių sentimentų rodiklis, Lietuvos investuotojų pasitikėjimo indeksas

1.6. Darbo temos ir uždavinių pagrindimas

Atlikus su darbo tema susijusios literatūros apžvalgą pastebėta, jog tyrėjus visame pasaulyje ypač dominantis objektas yra įvairūs veiksniai, darantys poveikį akcijų kainų ar akcijų rinkų indeksų kitimui. Iš to išplaukiantys prognozavimo tyrimai yra paremti kuo kokybiškesnių ir tikslesnių modelių sudarymo būdų paieškomis. Nors vieni autoriai bando prognozuoti tikslias reikšmes, kiti mano, jog to padaryti neįmanoma, todėl apsiriboja trumpo laikotarpio akcijų kainų ar akcijų rinkų indeksų kitimo spėjimais. Šiame darbe taip pat pasirenkamas pastarasis būdas. Apžvelgus nemažai tokių darbų pastebėta, jog vieni tyrėjai labiau kreipia dėmesį į kuo geresnių modelių parametrų parinkimą, kiti labiau akcentuoja taikomus prognozavimo metodus, dar kiti labai daug reikšmės skiria duomenims, tačiau beveik nekreipia dėmesio į kintamųjų atrinkimą. Su klasifikavimu susijusioje literatūroje dažnai teigiama, jog sudarant klasifikavimo/prognozavimo modelius yra svarbu atrinkti reikšmingus veiksniai tam, kad būtų galima pagerinti tikslumą ir tokių modelių kokybę. Žinoma, tam įgyvendinti galima taikyti įprastus kintamųjų atrankos metodus, tačiau atsižvelgiant į tai, kad jautrumo analizės metu yra siekiama nustatyti, kurie kintamieji daro didžiausią įtaką stebimo dydžio reikšmių pasikeitimui, nuspręsta tai susieti su klasifikavimu ir stebėti, ar taip atrinkus reikšmingus kintamuosius pagerėja modelių tikslumas. Kitaip tariant, šiame darbe pagrindinis dėmesys skiriamas nustatyti, ar pritaikant skirtingus jautrumo analizės metodus ir tokiu būdu atrinkus reikšmingus kintamuosius, galima pagerinti klasifikavimo tikslumą, kai prognozuojamas atskirų įmonių akcijų kainų ir akcijų rinkų indeksų kitimas atitinkamai viena diena (techninės analizės sritis) ir vienu mėnesiu (fundamentaliosios analizės sritis) į priekį.

Atsižvelgiant į kitų tyrėjų atliktus akcijų kainų kitimo prognozavimo viena diena į priekį tyrimus, šiame darbe, sudarant tokį duomenų rinkinį, nuspręsta naudoti akcijų prekybos rodiklius ir keletą sugeneruotų techninių indikatorių, atsisakant duomenų iš įvairių naujienų šaltinių. Kai prognozuojamas akcijų rinkų indeksų kitimas vienu mėnesiu į priekį, į duomenų rinkinį įtraukiami kitų autorių nustatyti kaip reikšmingi tokie kintamieji: BVP, valiutų kursas, palūkanų norma, vartotojų kainų indeksas, valdžios sektoriaus skola, nedarbo lygis. Taip pat pridėdama ir keliolika kitų rodiklių.

Pasirenkant tyrime taikomus mašininio mokymosi metodus, nuspręsta naudoti tris pagrindinius, kuriuos taikė nemažai kitų tyrėjų (atraminių vektorių metodas, neuroniniai tinklai ir sprendimų medis), bei du papildomus savo nuožiūra.

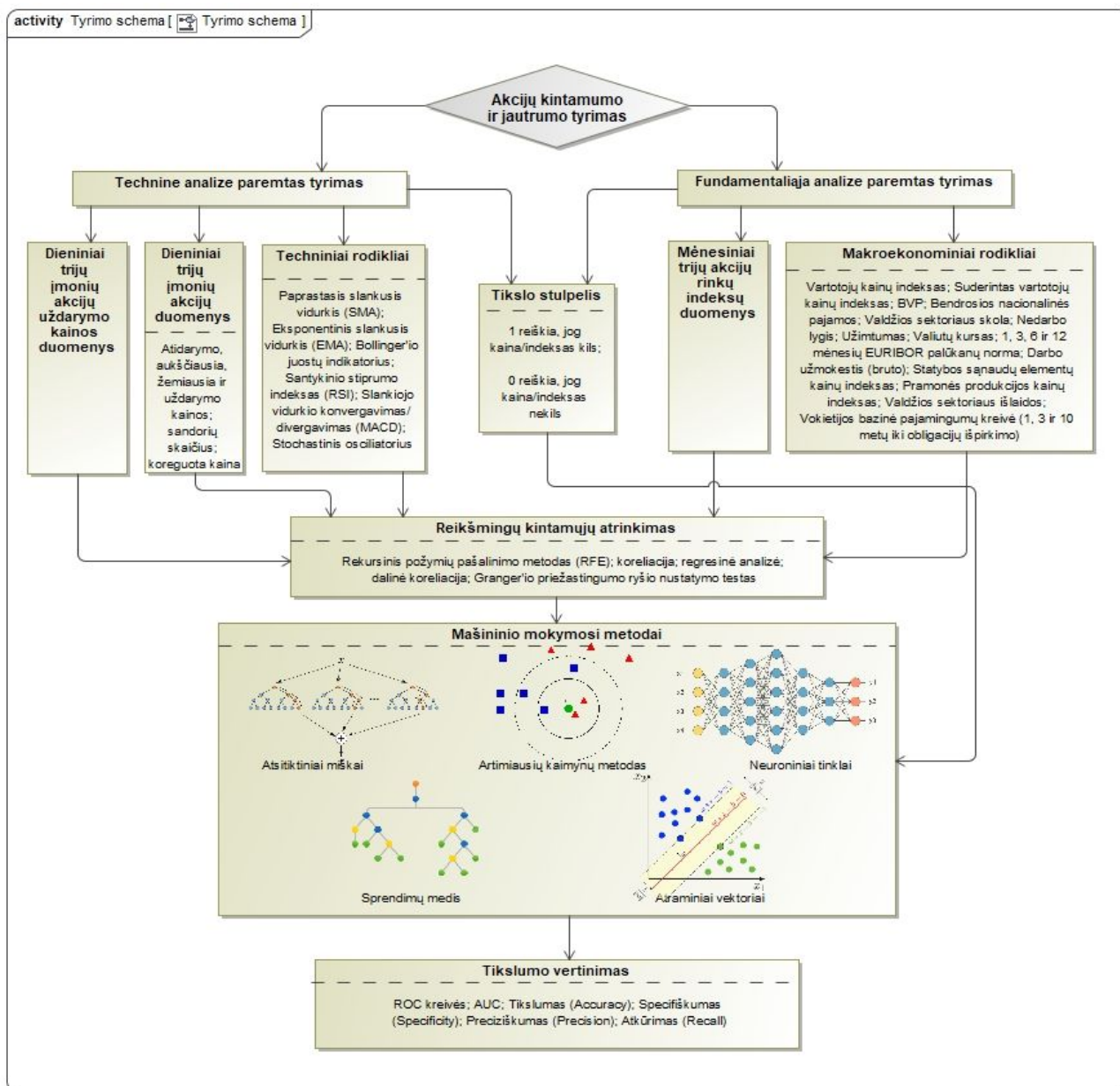
Atliekant tyrimą, pasirinkti Baltijos šalių akcijų rinkų duomenys, kadangi, kaip pastebėta, čia tokio pobūdžio tyrimai neatlikti.

2. Tyrimų metodai ir priemonės

Praeitame skyriuje nagrinėjama tema apibrėžta teoriniu aspektu, aprašant kitų autorių atliktus tyrimus ir gautus rezultatus. Išryškinta ir problema, kurios sprendimo galimybės pateikiamos šiame skyriuje. Čia, pirmiausia, aprašoma tyrimo, kuris susideda iš dviejų dalių, struktūra, po to apžvelgiami taikomi tyrimo metodai, galiausiai, apibrėžiama programinė realizacija.

2.1. Tyrimo struktūra

Šiame darbe du pagrindiniai žodžiai yra „kintamumas“ (angl. *volatility*) ir „jautrumas“ (angl. *sensitivity*). Dėl to tyrimas atliekamas toks, kad būtų kiek galima labiau išryškintas Baltijos šalių vertybinių popierių, šiuo atveju akcijų ir akcijų rinkų indeksų, kintamumas, paremtas prognozavimu bei pagal skirtingus jautrumo analizės metodus atrinktais reikšmingais kintamaisiais. Kaip minėta anksčiau, tyrimas susideda iš dviejų dalių, o detalus to vaizdavimas schematiškai matomas 4 pav.



4 pav. Atliekamo tyrimo schema

1.3. poskyryje jau buvo minėta, jog mokslinėje literatūroje nurodoma, kad išskiriamos dvi vertybinių popierių rinkos analizės mokyklos: *fundamentalioji* bei *techninė*. Kai kurie autoriai prie šių dviejų dar linkę pridėti ir *sentimentų* (teksto) analizę. Mąstant apie šio tyrimo gaires, buvo nuspręsta apsiriboti fundamentaliosios ir techninės analizės aspektais, atsisakant sentimentų analizės, kuri galėtų būti kito magistro darbo pagrindas.

Taip natūraliai atsirado dvi akcijų kintamumo ir jautrumo tyrimo dalys: „Technine analize paremtas tyrimas“ ir „Fundamentaliąja analize paremtas tyrimas“, kadangi norėta ištirti trumpo laikotarpio (*dienos*) ir ilgesnio laikotarpio (*mėnesio*) akcijų pokyčius, kuriems turi įtakos istoriniai akcijų duomenys bei įvairūs makroekonominiai rodikliai. Nemažai tyrėjų, atlikdami fundamentaliąją analizę, priklausomu kintamuoju renkasi ne konkrečių įmonių akcijas, o akcijų indeksus. Kadangi makroekonominiai, įvairių ūkio šakų duomenys dažniausiai būna mėnesiniai, toks analizės būdas indeksų atveju yra prasmingesnis nei atskirų akcijų. Be to, taip galima stebėti visos akcijų rinkos, o ne tik pavienių akcijų kaitą. Techninės analizės atveju pastebimos priešingos tendencijos, kadangi čia analizuojami trumpo laikotarpio pokyčiai, todėl tai labai aktualu atskirų akcijų prekybos metu. Dėl to šiame darbe, atliekant technine analize paremtą tyrimą, pasirinktos trijų įmonių akcijos (po vieną iš kiekvienos Baltijos šalies), o atliekant fundamentaliąją analize paremtą tyrimą – šalių akcijų rinkų indeksai: *OMX Tallinn*, *OMX Riga*, *OMX Vilnius*.

Daugiau ar mažiau apžvelgus dvilypę šio darbo tyrimo struktūrą, pereinama prie platesnio tiriamojame dalyje taikomų metodų aprašymo.

2.2. Tyrimo metodai ir modelių vertinimo priemonės

Siekiant įgyvendinti darbe iškeltą tikslą, pagrindinis atliekamo tyrimo akcentas bus duomenų klasifikavimas į dvi klases, pritaikant penkis mašininio mokymosi metodus: atsitiktiniai miškai (angl. *Random Forests (RF)*), artimiausių kaimynų metodas (angl. *k-Nearest Neighbors (kNN)*), neuroniniai tinklai (angl. *Neural Networks (nnet)*), sprendimų medis (angl. *Decision Tree (DT)*) ir atraminių vektorių metodas (angl. *Support Vector Machine (SVM)*). Šie metodai pasirinkti įvertinus, ką taikė kiti tyrėjai, atlikdami tokio pobūdžio tyrimus, ir remiantis savo subjektyvia nuomone apie kitokių metodų galimą pritaikomumą. Sudarytų modelių kokybės ir tikslumo vertinimui literatūroje dažnai išskiriami tokie matai: plotas po kreive (angl. *Area Under The Curve (AUC)*), kuris šiame darbe taikomas kaip pagrindinis tokio pobūdžio matas, tikslumas (angl. *Accuracy*), specifiškumas (angl. *Specificity*), preciziškumas (angl. *Precision*), atkūrimas (angl. *Recall*); ROC kreivė (angl. *Receiver Operating Characteristics*). Kadangi kiekvienas klasifikavimo etapas atliekamas su skirtingais duomenų rinkiniais, reikšmingų kintamųjų atrankai taikomi tokie metodai: rekursinis požymių pašalinimo metodas (angl. *Recursive Feature Elimination (RFE)*), koreliacija, regresinė analizė, dalinė koreliacija, Granger'io priežastingumo ryšio nustatymo testas. Toliau visa tai, kas parašyta, apžvelgiama detalčiau.

2.2.1. Mašininio mokymosi metodai

Atsitiktiniai miškai

Tai yra mokymosi „su mokytoju“ metodas, kurį, kaip galima suprasti iš pavadinimo, sudaro į miškus panašios struktūros, susidedančios iš sprendimų medžių. Šie medžiai gali būti „klasifikavimo medžiai“ ir „regresijos medžiai“. Dėl to atsitiktinis miškas taip pat gali būti

pritaikomas spręsti klasifikavimo ir regresijos problemas. Jei vertinti sprendimų medžio metodą atskirai, testuojant duomenis klasifikavimo metu, jis „perduoda“ tik vieną apmokytą medį. Čia išryškėja atsitiktinių miškų pranašumas, kadangi testavimo etapui šis metodas „perduoda“ keletą apmokytų medžių. Ši savybė atsitiktinių miškų metodą padaro pranašesniu prieš įprastą sprendimų medžio techniką [44].

Kai klasifikavimo metu sugeneruojama daug sprendimų medžių, jie „balsuoja“ už populiariausią klasę. Apibendrinant galima pateikti tokį apibrėžimą: atsitiktinis miškas yra klasifikatorius, susidedantis iš rinkinio kitų klasifikatorių – sprendimų medžių – $\{h(x, \Theta_k), k = 1, \dots\}$, kur $\{\Theta_k\}$ yra nepriklausomi vienodai pasiskirstę atsitiktiniai vektoriai, o kiekvienas medis atiduoda vieną balsą už populiariausią įvesties x klasę [45].

Artimiausių kaimynų metodas

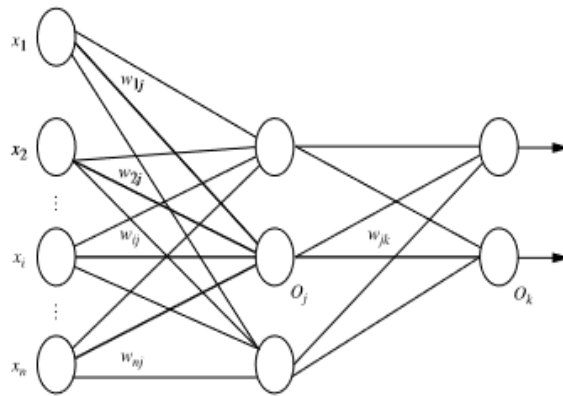
Trumpiausias šio metodo apibūdinimas būtų toks: panašios reikšmės atstovauja panašias klases. Artimiausių kaimynų metodas yra pagrįstas mokymusi pagal analogiją, kai testavimo imties įrašai yra lyginami su apmokymų imties įrašais, vertinant panašumą. Apmokymo įrašai apibūdinami n atributų. Kiekvienas įrašas atspindi vieną tašką n dimensijos erdvėje. Tokiu būdu visi apmokymų imties įrašai patenka į šią modelio erdvę. Kai atsiranda nežinomas įrašas, artimiausių kaimynų klasifikatorius modelio erdvėje ieško tų k apmokymų įrašų, kurie yra artimiausi nežinomajam. Šie k apmokymų imties įrašai yra k „artimiausi kaimynai“ nežinomam įrašui. Čia „artumas“ apibrėžiamas kaip atstumo metrika, pvz., Euklido atstumas. [46] Šis atstumas tarp dviejų taškų ar įrašų, tarkime, $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ ir $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$, matuojamas taip:

(4)

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2};$$

Neuroniniai tinklai

Neuroninis tinklas yra sujungtų įvesties/išvesties vienetų rinkinys, kai kiekvienas ryšys susietas su tam tikru svoriu. Modelio sudarymo metu tinklas „mokosi“ koreguodamas svorius taip, kad būtų galima korektiškai prognozuoti įėjimo įrašų klases. 5 paveikslėlyje matomas neuroninio tinklo pavyzdys. Šis tinklas sudarytas iš *įvesties sluoksnio* (pirmas lygmuo), *paslėpto sluoksnio* (antras lygmuo) ir *išvesties sluoksnio* (trečias lygmuo). Kiekvienas sluoksnis susideda iš vienetų. Paslėptų sluoksnių gali būti ne vienas. Kai įvestys praeina *įvesties sluoksni*, būna pridodamas svoris ir nukreipiama *paslėptajam sluoksniui*. Išvestys iš to gali patekti į kitą *paslėptąjį sluoksni* ir taip toliau. Paskutinio *paslėptojo sluoksnio* svertiniai išėjimai yra įvedami į tuos vienetus, kurie sudaro *išvesties sluoksni*, perteikiantį tinklo prognozes [46].



5 pav. Neuroninio tinklo schema [46]

Sprendimų medis

Apie sprendimų medį jau truputį buvo rašyta aptariant atsitiktinius miškus. Tai yra hierarchinis mokymosi „su mokytoju“ metodas, kurio struktūra primena medį. Egzistuoja dvi šių medžių kategorijos: klasifikavimo ir regresijos medžiai. Klasifikavimui, kaip ir šiuo atveju, metodas taikomas, bene, dažniausiai. Čia padalijimo kriterijumi yra medžio mazgas, kur „priimamas“ sprendimas. Šis kriterijus suskirsto apmokymo duomenis į dvi ar daugiau dalių. Pavyzdžiui, jei suskirstymo kriterijus yra amžius ≤ 30 , tai kairėje medžio pusėje galėtų būti visi stebiniai, kai amžius ne didesnis už 30, o dešinėje – daugiau nei 30. Padalijimo kriterijaus tikslas yra taip suskirstyti duomenis, kad „susimaišymas“ tarp kintamųjų, priklausančių skirtingoms klasėms, kiekvienoje šakoje būtų kuo mažesnis [47].

Tam, kad duomenų padalijimo kokybė skirtingų sprendimų medžio algoritmų pritaikymo atveju būtų kuo aukštesnė, apibrėžiami keli dydžiai, pasižymintys skirtingomis savybėmis:

1. *Klaidos lygis* (angl. *Error rate*). Tarkime, kad p yra dominuojančiai klasei priklausančių duomenų taškų rinkinio S atveju dalis. Tada klaidos lygis bus $1-p$. Jei rinkinys S suskirstomas į atskirus rinkinius $S_1 \dots S_r$, tada bendras padalijimo klaidų lygis bus apibrėžiamas kaip svertinis atskirų rinkinių S_i klaidų lygių vidurkis, kai S_i svoris yra $|S_i|$. Padalijimas, kai klaida mažiausia, bus pasirenkamas iš kitų alternatyvų.
2. *Gini indeksas* (angl. *Gini index*). Šis indeksas $G(S)$ duomenų rinkiniui S gali būti apskaičiuojamas pagal apmokymo duomenų, esančių rinkinyje S , klasių pasiskirstymą $p_1 \dots p_k$.

(5)

$$G(S) = 1 - \sum_{j=1}^k p_j^2;$$

Bendras Gini indeksas, kai rinkinys S yra suskaidomas į mažesnius rinkinius $S_1 \dots S_r$ gali būti apibrėžiamas kaip kiekvieno S_i svertinis indekso reikšmių $G(S_i)$ vidurkis, kai S_i svoris yra $|S_i|$.

(6)

$$\text{Gini - Split}(S \Rightarrow S_1 \dots S_r) = \sum_{i=1}^r \frac{|S_i|}{|S|} G(S_i);$$

Suskaidymas su mažiausiu Gini indeksu yra pasirenkamas iš kitų alternatyvų.

3. *Entropija* (angl. *Entropy*). Šis matas $E(S)$ duomenų rinkiniui S gali būti apskaičiuojamas pagal apmokymo duomenų, esančių mazge, klasių pasiskirstymą $p_1 \dots p_k$.

(7)

$$E(S) = - \sum_{j=1}^k p_j \log_2(p_j);$$

Bendra entropijos reikšmė, kai rinkinys S yra suskaidomas į mažesnius rinkinius $S_1 \dots S_r$ gali būti apibūdinamas kaip kiekvieno S_i svertinis mato verčių $E(S_i)$ vidurkis, kai S_i svoris yra $|S_i|$.

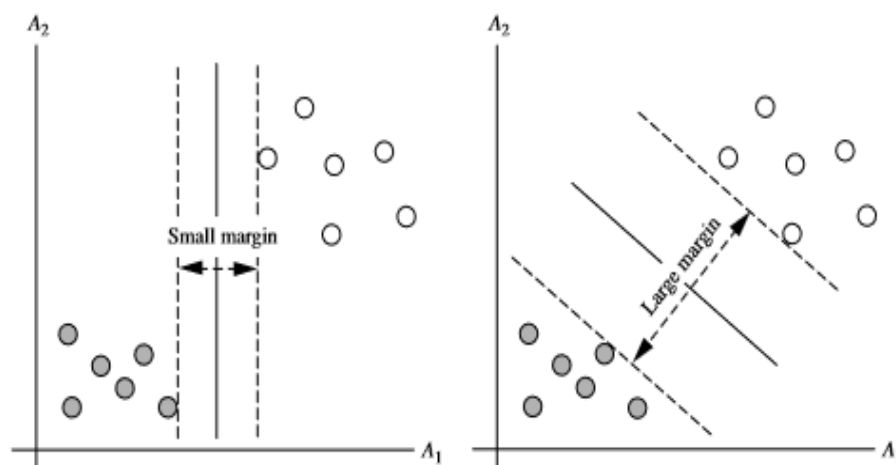
(8)

$$\text{Entropy - Split}(S \Rightarrow S_1 \dots S_r) = \sum_{i=1}^r \frac{|S_i|}{|S|} E(S_i);$$

Labiau pageidaujamos yra žemesnės entropijos reikšmės. [47]

Atraminių vektorių metodas

Šis metodas skirtas klasifikuoti tiek tiesiškai, tiek netiesiškai pasiskirsčiusius duomenis. Paprasčiausias atvejis – tai klasifikavimo į dvi klases problema, kai klasės yra atskiriamos tiesiškai. Tokį atskyrimą apibūdina tiesios linijos, kurių gali būti begalė. Tikslas yra rasti vieną geriausią, kuri, kaip tikimasi, iššauktų mažiausią klasifikavimo klaidą. Kitaip tai būtų galima apibūdinti, kaip siekį rasti geriausią hyper plokštumą (angl. *hyperplane*). Atraminių vektorių metodas šią problemą sprendžia ieškodamas maksimalių paraščių hyper plokštumos [46]. 6 paveiksle matomi du klasių atskyrimo būdai, kai paraštės yra skirtingos.



6 pav. Atraminių vektorių metodo tiesinio klasių atskyrimo schema su skirtingomis paraštėmis [46]

Abiem šiais atvejais hyper plokštumos gali korektiškai klasifikuoti duomenis, tačiau tikimasi, kad hyper plokštuma su didesnėmis paraštėmis tai darys tiksliau. Dėl to (apmokymo metu) atraminių vektorių metodas ieško hyper plokštumos su didesnėmis paraštėmis. Atskiriančią hyper plokštumą galima būtų užrašyti taip:

$$W \cdot X + b = 0; \quad (9)$$

čia W yra svorio vektorius $W = \{w_1, w_2, \dots, w_n\}$; n yra atributų skaičius; X apibūdina apmokymo duomenis; b yra skaliaras, dažnai įvardijamas tiesiog nuolydžiu [46].

Atraminiais vektoriais vadinami abiejų klasių taškai, esantys ant paraščių krašto. Jie vieninteliai daro įtaką sprendimo ribai.

Egzistuojant netiesiškam atskyrimui, įvedamas „branduolio triukas“ (angl. *kernel trick*), kuris perkelia įvestis į aukštesnę erdvę tam, kad būtų galima atskirti klases [48].

Toliau bus aprašomi klasifikavimo modelių vertinimo ir atrankos metu taikomi matai.

2.2.2. Klasifikavimo modelių vertinimas ir atranka

Šiame etape svarbų vaidmenį atlieka sumaišymų matrica (angl. *confusion matrix*). Tai yra naudingas įrankis analizuojant, kaip gerai klasifikatorius gali atpažinti skirtingų klasių įrašus. Vizualinė sumaišymų matricos išraiška pateikiama 5 lentelėje.

5 lentelė. Sumaišymų matrica, nurodant teigiamas ir neigiamas reikšmes iš viso [46]

		Prognozuojama klasė		Iš viso
		Taip	Ne	
Tikroji klasė	Taip	Teisingai teigiama (TP)	Klaidingai neigiama (FN)	Teigiamų (P)
	Ne	Klaidingai teigiama (FP)	Teisingai neigiama (TN)	Neigiamų (N)
Iš viso		Teigiamų prognozuotų (P')	Neigiamų prognozuotų (N')	Teigiamų ir neigiamų (P+N)

Norint suprasti sumaišymų matricos esmę, būtina žinoti, ką reiškia keturi „kertiniai akmenys“, naudojami skaičiuojant daugelį modelių vertinimo matų. Perpratus juos, bus lengviau suvokti ir šių matų reikšmę.

- *Teisingai teigiama* (angl. *True positive (TP)*). Tai yra teigiamos klasės reikšmės, kurios suklasifikuotos teisingai.
- *Teisingai neigiama* (angl. *True negative (TN)*). Tai yra neigiamos klasės reikšmės, kurios suklasifikuotos teisingai.
- *Klaidingai teigiama* (angl. *False positive (FP)*). Tai yra tie neigiamos klasės įrašai, kurie neteisingai buvo priskirti/prognozuoti teigiamai klasei.
- *Klaidingai neigiama* (angl. *False negative (FN)*). Tai yra tie teigiamos klasės įrašai, kurie neteisingai buvo priskirti/prognozuoti neigiamai klasei [46].

Keli klasifikavimo modelių vertinimo matai su nurodymu, koks rodiklis tai yra, ir jų apskaičiavimo formulės pateikiamos 6 lentelėje.

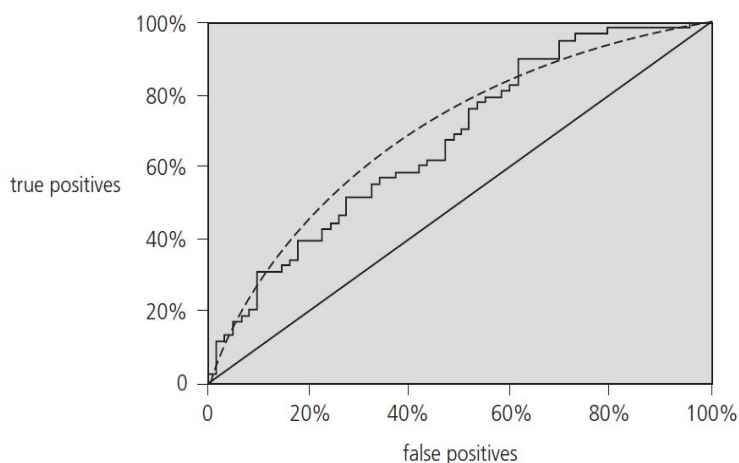
6 lentelė. Modelių vertinimo matai ir jų skaičiavimo formulės [46]

Matas	Formulė
Tikslumas (patvirtinimo rodiklis)	$(TP + TN) / (P + N)$
Jautrumas/atkūrimas (teisingai teigiamos reikšmės rodiklis)	TP / P
Specifiškumas (teisingai neigiamos reikšmės rodiklis)	TN / N
Preciziškumas	$TP / (TP + FP)$

Plačiau šiuos matus galima apibūdinti taip:

- *tikslumas* (angl. *accuracy*). Tai yra klasifikatoriaus klasifikavimo tikslumo procentinė išraiška, parodanti, kaip tiksliai reikšmės suskirstomos pagal klases;
- *jautrumas* (angl. *sensitivity*). Apibūdina teisingai prognozuotų teigiamų reikšmių santykį su visa teigiama klase;
- *specifiškumas* (angl. *specificity*). Apibūdina teisingai prognozuotų neigiamų reikšmių santykį su visa neigiama klase;
- *atkūrimas* (angl. *recall*). Galutinė šio mato išraiška yra kaip ir jautrumo, tačiau korektiškas apibūdinimas skamba taip: apibrėžia teisingai prognozuotų teigiamų reikšmių santykį su teisingai prognozuotų teigiamų ir klaidingai prognozuotų neigiamų reikšmių suma ($TP / (TP + FN)$);
- *preciziškumas* (angl. *precision*). Apibrėžiamas, kaip teisingai prognozuotų teigiamų reikšmių santykis su teisingai prognozuotų teigiamų ir klaidingai prognozuotų teigiamų reikšmių suma.

Dar vienas labai svarbus matas yra AUC (angl. *Area Under The Curve*) – ploto po ROC (angl. *Receiver Operating Characteristics*) kreive reikšmė ir, žinoma, pati ROC kreivė. Ši kreivė parodo klasifikatoriaus veikimo kokybę, neatsižvelgiant į klasės pasiskirstymą ar klaidos kainą. Čia vertikali ašis (žr. 7 pav.) apibūdina teigiamų reikšmių skaičių, išreikštą procentais, palyginus su visomis teigiamomis reikšmėmis ($tp = (TP / (TP + FN)) \times 100\%$), o horizontali ašis parodo neigiamų reikšmių skaičių, išreikštą procentais, palyginus su visomis neigiamomis reikšmėmis ($fp = (FP / (FP + TN)) \times 100\%$) [49].



Paveikslėlyje matoma nelygi kreivė labai priklauso nuo konkrečios testavimo imties duomenų, o lygesnė ji gaunama tada, kai taikomas kryžminis patikrinimas (angl. *cross-validation*) [49].

Anot Witten'o ir Frank'o [49], kuo didesnis plotas po ROC kreive, kuris vadinamas AUC, tuo geresnis sudarytas modelis. Ši reikšmė interpretuojama kaip tikimybė, kad klasifikatorius „teiks pirmenybę“ klasifikuoti atsitiktinai parinktas reikšmes teigiamomis, nei neigiamomis.

2.2.3. Jautrumo analizės metodai

Kaip teigia Aggarwal'as [47], kintamųjų atranka turėtų būti pirmas klasifikavimo proceso įgyvendinimo etapas, kadangi kai kurie jų gali turėti mažai įtakos prognozuojant skirtingų klasių reikšmes. Net priešingai, tokie kintamieji darys neigiamą poveikį klasifikavimo modelių tikslumui ir bus skaičiavimų neefektyvumo šaltinis. Dėl to pagrindinis kintamųjų atrankos metu taikomų algoritmų tikslas yra atrinkti labiausiai informatyvius požymius, susijusius su prognozuojamomis klasėmis. Egzistuoja trys atrankos metodų tipai:

1. *filtravimo metodai*. Tai yra aiškus matematinis kriterijus, leidžiantis įvertinti kintamojo ar kelių kintamųjų kokybę, kai pašalinami nereikšmingieji;
2. „*apgaubimo*“ *metodai*. Čia daroma prielaida, kad klasifikavimo algoritmas yra tinkamas įvertinti, kaip gerai jis veikia su tam tikrais kintamaisiais. Tada požymių atrankos metodas „apgaukia“ šį klasifikavimo algoritmą, siekiant sudaryti reikšmingų kintamųjų rinkinį;
3. *įterpimo metodai*. Jei sudarytas klasifikavimo modelis savyje talpina informaciją apie reikšmingiausius kintamuosius, tada tokie požymiai yra izoliuojami, o klasifikatorius apmokomas su „apkarpytais“ kintamaisiais [47].

Šiame darbe svarbūs kintamieji taip pat atrenkami, tačiau tai daroma pritaikant keturis jautrumo analizėje naudojamus metodus ir vieną tradicinį kintamųjų atrankos algoritmą tam, kad būtų galimybė palyginti. Šie metodai yra:

1. *Rekursinis požymių pašalinimo metodas* (angl. *Recursive Feature Elimination (RFE)*). Jis čia pateikiamas kaip tradicinis kintamųjų atrankos algoritmas. Ši procedūra, pirmiausia, optimizuoja svorius, tada apskaičiuoja visų kintamųjų vertinimo kriterijų ir pašalina tuos, kurių gauto kriterijaus reikšmė yra mažiausia. Tai yra atgalinio (angl. *backward*) požymių pašalinimo pavyzdys. Tie kintamieji, kurie atrinkti kaip svarbūs, nebūtinai yra patys svarbiausi individualiai. Tik taikomi drauge jie yra optimalūs [50].
2. *Koreliacija*. Dar kitaip tai vadinama koreliacine analize, kai norima nustatyti asociaciją ar ryšį tarp dviejų (ar daugiau) kintamųjų. Čia remiamasi linijinės priklausomybės prielaida tarp kiekybinių dydžių. Koreliacinės analizės rezultatas yra koreliacijos koeficientas, kurio reikšmės kinta nuo -1 iki +1. Kuo arčiau vieneto, tuo priklausomybė yra didesnė. Skaičiuojami Pearson'o ir Spearman'o koreliacijos koeficientai. Pearson'o koeficientas taikomas tada, kai du kintamieji tenkina tokias prielaidas: ryšys yra linijinis, kintamieji vienas nuo kito nepriklausomi ir pasiskirstę pagal normalųjį skirstinį. Jei taip nėra, skaičiuojamas Spearman'o koreliacijos koeficientas, kuris gali būti naudojamas kaip neparаметrinis Pearson'o koeficiento atitikmuo [51]. Šiame darbe dėl prielaidų netenkinimo skaičiuojamas Spearman'o koreliacijos koeficientas ir pasirenkami tie reikšmingi kintamieji, su kuriais gauta koeficiento reikšmė yra didesnė nei 0,5.

3. *Regresinė analizė.* Šiame darbe taikoma daugialypė tiesinė regresija, kurios įgyvendinimo metu modeliuojamas ryšys tarp nepriklausomų kintamųjų ir priklausomo (tikslu) kintamojo. Daug įvairių reiškinių negali būti apibūdinami linijine priklausomybe tarp kintamųjų, todėl sudaromi kitokie regresijos modeliai. Tam, kad tiesinės regresijos modelis būtų tinkamas prognozuoti, reikia tenkinti tokias prielaidas: ryšys tarp priklausomo ir nepriklausomų kintamųjų yra linijinis, nėra multikolinearumo problemos tarp nepriklausomų kintamųjų, tenkinama homoskedastiškumo prielaida, liekamosios paklaidos pasiskirsčiusios pagal normalųjį skirstinį [52]. Šio darbo duomenys netenkino multikolinearumo nebuvimo, tačiau čia nebuvo siekiama prognozuoti, o tik atrinkti reikšmingus kintamuosius, todėl prielaidos netenkinimas ignoruotas ir pasirinkti tie veiksniai, kurie pirmo modelio sudarymo metu buvo nurodyti svarbiais.
4. *Dalinė koreliacija.* Čia taip pat skaičiuojami koreliacijos koeficientai, bet, skirtingai nei įprastos koreliacijos atveju, dalinė koreliacija yra metodas skirtas identifikuoti koreliaciją tarp dviejų kintamųjų, kai kiti kintamieji laikomi konstantomis [53].
5. *Granger'io priežastingumo ryšio nustatymo testas* (angl. *Granger causality test*). 1969 metais Granger'is [54] pasiūlė atlikti priežastinio ryšio nustatymo testą, taikant laiko eilutės duomenis. Anot autoriaus, X yra Y priežastis, kai tai naudinga prognozuojant Y . Čia „naudinga“ reiškia tai, kad X gali pagerinti Y prognozavimo tikslumą, jei atsižvelgiama tik į istorines Y reikšmes. Egzistuoja kelios situacijos, kai šis testas gali būti pritaikomas: 1) dviejų kintamųjų ir jų *lagų* atveju, 2) kai pasirinkti daugiau nei du kintamieji, 3) galiausiai, priežastingumas gali būti testuojamas VAR sistemoje.

Šiame poskyryje aprašyti visi tyrime taikomi tyrimo metodai ir kita reikalinga informacija. Toliau trumpai bus apibūdinama viso to įgyvendinimo programinė realizacija.

2.3. Programinė realizacija

Visa praktinė darbo dalis įgyvendinama naudojant programą R. Ji pasirinkta dėl savo paprastumo, įvairiapusio pritaikomumo ir plataus naudojimo visame pasaulyje.

Kadangi beveik visų darbe taikomų duomenų matavimų skalės buvo skirtingos, jie standartizuoti tam, kad atliekamas tyrimas būtų korektiškesnis, o gauti rezultatai tikslesni ir patikimesni. Taip pat vyravo nemažas klasių disbalansas, kai reikšmių skirtumas tarp klasių kartais siekdavo keliasdešimt procentų, todėl buvo atliekamas duomenų balansavimas. Tai, kad didelis klasių disbalansas yra problema, kai norima vertinti klasifikavimo modelių tikslumą, savo straipsnyje pažymi ir Tharwat'as [58]. Pavyzdžiui, tikslumo matas (angl. *accuracy*) yra jautrus nesubalansuotiems duomenims. Tarkime, kad kažkokiam duomenų rinkinyje neigiamos klasės reikšmių yra 95 %, o teigiamos – 5 %. Pasirinktas klasifikatorius tokius duomenis gali labai tiksliai priskirti neigiamai klasei ir tikslumas bus aukštas, tačiau realaus vertinimo prasme tai nebus korektiška. Dėl šios priežasties, kaip minėta, atliktas duomenų balansavimas, pritaikant *ROSE* (angl. *Random Over-Sampling Examples*) paketą. Čia apmokymo duomenų rinkinys yra „papildomas“ mažumos klasės reikšmėmis ir taip abiejų klasių santykis tampa apylygis. Dėl kai kurių savybių *ROSE* funkcija yra pranašesnė nei kiti tokio pobūdžio metodai [59].

¹ Standartizavimo metu duomenys keičiami taip, kad iš atsitiktinio (standartizuojamo) dydžio atimamas pasiskirstymo vidurkis, kuris lygus 0, ir padalinama iš pasiskirstymo standartinio nuokrypio, kurio reikšmė yra 1 ($z = (X - \mu) / \sigma$).

Duomenyse praleistų reikšmių nenustatyta, išsiskiriantys stebiniai nė vienu atveju neviršijo daugiau nei 5 % rekomenduojamos ribos.

Klasifikavimo modelių sudarymo metu taikomas paketas *CARET*. Modelius bandyta sudaryti ir su įprastomis mašininio mokymosi funkcijomis, tačiau tikslumas nebuvo toks geras kaip naudojant minėtąjį paketą. Čia kiekvieno metodo pritaikymo atveju sudaromas parametru „tinklelis“ ir taip stengiamasi rasti geriausią modelį. Tam, kad nekiltų persimokymo problema, atliekamas duomenų kryžminis patikrinimas (angl. *cross-validation*), tačiau šiuo atveju taikomas dviejų „raundų“ kryžminis patikrinimas (angl. *nested* arba *two-round cross validation*), kai pirmu etapu derinami parametrai, o antru vertinamos klaidos. Visi programinės realizacijos kodai pateikiami 1, 2, 3, 4, 5 prieduose.

3. Tyrimo rezultatai ir jų aptarimas

Pagrindinė atliekamo tyrimo paskirtis yra padėti investuotojams priimti kiek galima tikslesnius sprendimus, susijusius su akcijų pirkimu, pardavimu ar laikymu. Pasaulyje yra sukurta įrankių, padedančių vertinti ateities akcijų kainų ar akcijų indeksų kitimą, todėl šio tyrimo rezultatai papildo tai, kas jau egzistuoja, ir gali būti pritaikoma kasdienėje veikloje. Tyrimas susideda iš dviejų dalių. Pirmiausia, analizuojamas akcijų kainų kitimas viena diena į priekį, o kartu ir to prognozavimas. Čia taikomi kiekvienos iš pasirinktų įmonių akcijų kainų istoriniai duomenys, juos susiejant su praeitos dienos ir esamos dienos uždarymo kainos skirtumu: „kaina kils“ arba „kaina nekils“ („liks tokia pati“ taip pat patenka į „kaina nekils“ grafą). Iškeliama prielaida, jog kainos kintamumo prognozavimas gali būti tikslesnis, jei atrenkami reikšmingi kintamieji ir tai daroma panaudojant įvairius jautrumo analizės metodus. Kitoje tyrimo dalyje apžvelgiamos visų trijų Baltijos šalių akcijų rinkų indeksų svyravimų galimos prognozės mėnesiu į priekį, kai tikslo stulpelio grafų reikšmės tokios pat kaip pirmoje dalyje, o nepriklausomi kintamieji yra įvairūs makroekonominiai rodikliai. Čia taip pat atliekamas kintamųjų atrinkimas skirtingais metodais ir prielaida pritaikoma ta pati. Atsižvelgiant į tai, kas paminėta, toliau bus aprašomi šių dviejų tyrimo dalių įgyvendinimo metu gauti rezultatai prieš tai pateikiant tyrime taikomų duomenų apžvalgą.

3.1. Duomenų apžvalga

3.1.1. Įmonių akcijų kainų kitimo prognozavimo metu naudojami duomenys

Analizei pasirinktos trijų (po vieną iš Lietuvos, Latvijos ir Estijos), iš pirmo žvilgsnio sėkmingų, įmonių akcijų duomenys, kurie pateikiami *Nasdaq Baltic* interneto svetainėje. Pasirinkta po vieną iš kiekvienos šalies dar ir dėl to, kad galbūt kažkokie vidiniai valstybės ekonominiai niuansai, kurie analizuojamu laikotarpiu lėmė vienokį ar kitokį kainų pasikeitimą, galėtų daryti įtaką dabar atliekamo prognozavimo metu. Tiriamas laikotarpis – nuo 2009-01-02 iki 2018-12-28. Tai duomenų matricoje sudaro po 2495 eilutes kiekvienu atveju. Gali atrodyti, jog toks skaičius nėra didelis, kai šio darbo pagrindas turėtų būti didieji duomenys, tačiau, kaip savo darbe teigia Oli. Sigo, Selvam'as, Maniam'as, et al. [41], jie naudoja didžiuosius duomenis, kadangi visą duomenų matricą sudaro 29718 stebinių: pasirinktos trijų įmonių akcijos, kurias kiekvieną apibrėžia keturi kintamieji ir 2476 prekybos dienos. Šiame magistro darbe ir šiame tyrime kiekvienos iš trijų įmonių duomenų matricą sudaro 42415 stebinių (17 kintamųjų ir 2495 prekybos dienos), todėl tai taip pat galima įvardinti esant didžiais duomenimis.

Indeksas/Akcijos	2009.01.02	2018.12.28	+/-%
– GRG1L	0,0869 EUR	1,325 EUR	1 424,99 ↑
– GRD1R	4,5959 EUR	6,5 EUR	41,43 ↑
– TKM1T	2,08 EUR	8,42 EUR	304,81 ↑

8 pav. Įmonių akcijų kainų pokyčiai ilguoju laikotarpiu

Pasirinktų įmonių tam tikrą sėkmingumą, subjektyvia nuomone, apibūdina akcijų kainų kilimas ilguoju laikotarpiu, išreikštas procentais (žr. 8 pav.) ir vizualus kainų dinamikos grafikų įvertinimas (žr. 9 pav.). Visa ši informacija pateikiama iš *Nasdaq Baltic* interneto svetainės. 8 paveikslėlyje matoma, jog visų trijų įmonių akcijų kainos per 10 metų pakilo nuo keliasdešimt iki daugiau nei tūkstančio procentų, o 9 pav. grafikas parodo šią kainų dinamiką vizualiai visu analizuojamu

laikotarpiu. Lietuvos įmonės rezultatai (oranžinė kreivė) vizualiai atrodo tikrai įspūdingai, tačiau jų pradinė akcijų kaina analizuojamo laikotarpio pradžioje buvo tik 9 euro centai, todėl ir pokytis toks didžiulis, kai pakilo iki 1,3 euro. Tuo pačiu metu Estijos įmonės (mėlyna kreivė) akcijų kaina pakilo daugiau nei 6 eurai, tačiau vizualiai tai neatrodo taip įspūdingai.



9 pav. Įmonių akcijų kainų grafikai

Kaip minėta, tyrimui pasirinkta po vieną Lietuvoje, Latvijoje ir Estijoje veikiančią įmonę (iš *Nasdaq Baltic* skelbiamo Baltijos Oficialiojo prekybos sąrašo). Jos yra:

- *AB „Grigeo“* (sutrumpintai GRG1L). Tai įmonė, užsiimanti sanitarinio buitinio popieriaus ir jo gaminių bei gofruotojo kartono ir pakuočių iš jo gamyba. Rinka: *Nasdaq Vilnius*;
- *AS „Grindeks“* (sutrumpintai GRD1R). Įmonė užsiima farmacijos produktų gamyba. Rinka: *Nasdaq Riga*;
- *AS „Tallinna Kaubamaja Grupp“* (sutrumpintai TKMIT). Įmonės veikla yra prekių didmeninė prekyba ir pardavimas. Rinka: *Nasdaq Tallinn*.

Neapsiribojant tik subjektyviu analizuojamų įmonių akcijų vertinimu, pasirinkta pora objektyvių informacijos šaltinių tam, kad būtų galima susidaryti išsamesnį „vaizdą“ apie šias akcijas. Dėl to nubrėžiamos Bollinger'io juostos (angl. *Bollinger Bands*) laikotarpiu nuo 2018-01-02 iki 2018-12-28 ir skaičiuojami efektyvumo indeksai (EI) nuo 2009 iki 2018 metų imtinai bei tik 2018 metais kiekvienos įmonės atveju atskirai.

Apie Bollinger'io juostas buvo rašyta 1.3. poskyryje. Pasikartojant galima paminėti, jog šis indikatorius yra vienas iš kintamumo rodiklių. Jis parodo, kada rinka yra pasyvi, o kada joje aktyviai vyksta prekyba. Kai rinka yra pasyvi, Bollinger'io juostos tampa siauros (mažas kintamumas), kai pagyvėja - juostos prasiplečia (didelis kintamumas). Kuo kintamumas didesnis, tuo investavimas į tokias akcijas tampa rizikingesnis. Kainos šalia viršutinės juostos yra aukštos, šalia apatinės – žemos. Grafike kainas atspindi stačiakampiais, kurie simbolizuoja kainų kritimą arba kilimą, „aplipdyta“ diagrama. Vidurinė juosta rodo slankiojo vidurkio (rekomenduojama 20 periodų) judėjimą, viršutinė juosta yra vidurinė juosta, „pereinanti“ į viršų tam tikrą standartinių nuokrypių skaičių (dažniausiai 2), apatinė – vidurinė juosta, „pereinanti“ žemyn tam tikrą

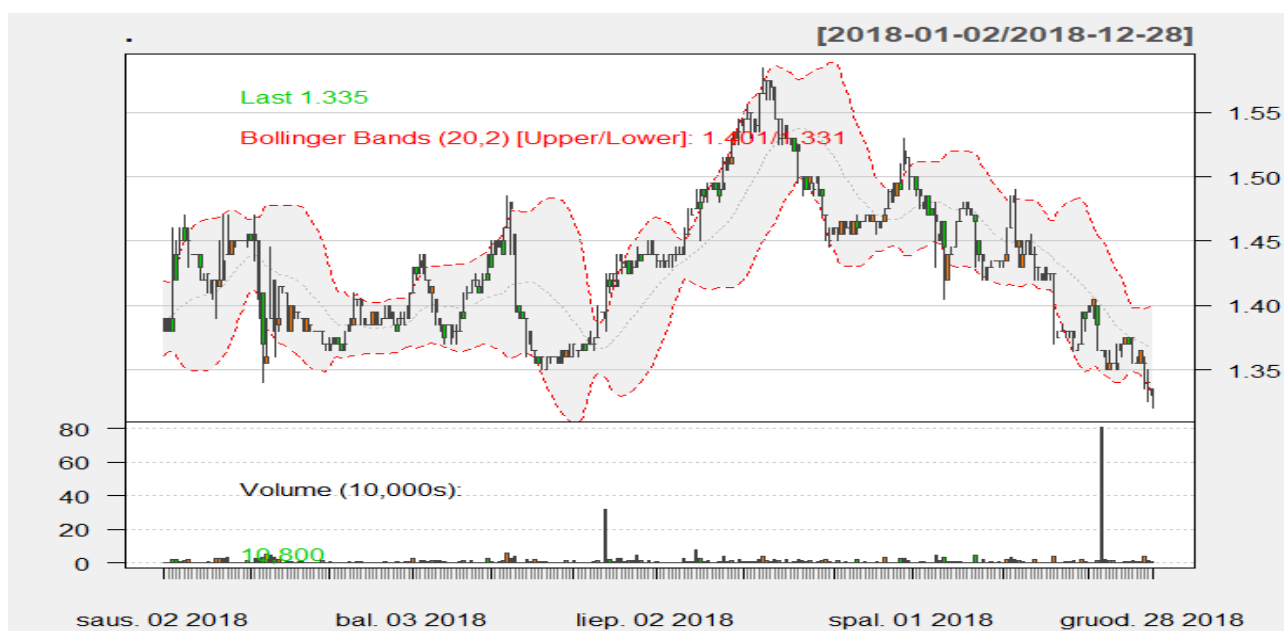
standartinių nuokrypių skaičių (dažniausiai 2). Standartinis nuokrypis yra kintamumo matas, kuris apibrėžia juostų plotį. Aukštas kintamumas iššaukia žemą kintamumą, o žemas – aukštą.

Kristoufek'as, Vosvrda's [42] apibrėžė efektyvumo indeksą (angl. *Efficiency Index*), kuris parodo, kiek efektyvi yra rinka ar pavienės akcijos. Kuo rinka/akcijos efektyvesnės, tuo labiau kainos tampa nebe prognozuojamos, o atsitiktinės. Autoriai, skaičiuodami šį indeksą, į vieną apjungė kelis matavimus: Hurst eksponentę (angl. *Hurst exponent*), Fraktalinį dydį (angl. *Fractal dimension*) ir Apytikslę entropiją (angl. *Approximate entropy*). Kuo gauta indekso reikšmė yra mažesnė, tuo rinka/akcijos efektyvesnės. Tyrėjų gauta mažiausia tam tikros rinkos efektyvumo indekso reikšmė buvo 0,0619, tačiau tam, kad akcijų rinką ar akcijas būtų galima įvardinti esant efektyviomis, šis dydis neturėtų viršyti vieneto. Taip pat autoriai padarė išvadą, kad rinkos/akcijų efektyvumas yra tiesiogiai susijęs su rinkos išsivystymo lygiu.

Viską aprašius teoriniu aspektu, galima apžvelgti praktiškai, kaip atrodo analizuojamų įmonių Bollinger'io juostos, ir kokie gaunami apskaičiuoti efektyvumo indeksai pagal kiekvieną atvejį atskirai.

3.1.1.1. Lietuvos įmonė AB „Grigeo“

Grafike (žr. 10 pav.) galima matyti, kad beveik visu analizuojamu laikotarpiu – 2018 metais – vyko aktyvi prekyba šios įmonės akcijomis. Juostos buvo plačios, vadinasi, kintamumas didelis. Taip pat matyti, kad aukštas kintamumas iššaukė žemą kintamumą, o žemas – aukštą. Be viso to dar pastebima, jog akcijų kaina metų pabaigoje smuko.



10 pav. Bollinger'io juostos pagal įmonės AB „Grigeo“ akcijų uždarymo kainos duomenis

Apskaičiuotas efektyvumo indeksas laikotarpiu nuo 2009 iki 2018 metų yra 0,6593. Tai parodo, kad akcijos, vertinant pagal to meto duomenis, buvo efektyvios. O gauti tik 2018 metų efektyvumo indekso rezultatai dar kiek mažesni – 0,5436. Tai parodo, kad tais metais šios įmonės akcijos buvo tikrai efektyvios.

3.1.1.2. Latvijos įmonė AS „Grindeks“

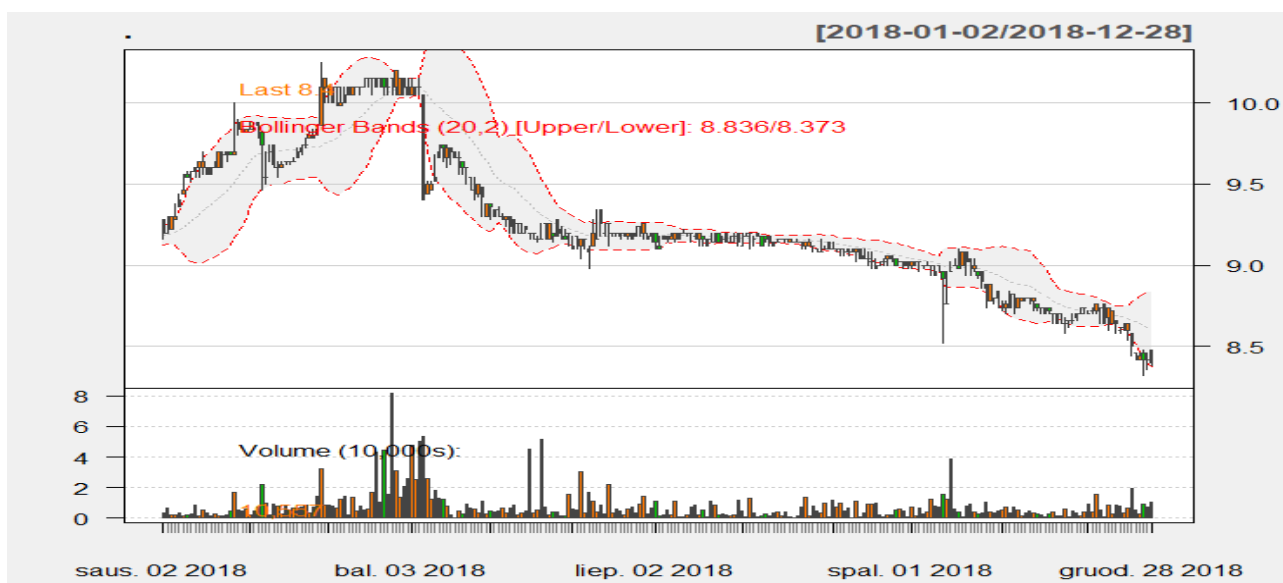
Pateiktame grafike (žr. 11 pav.) matoma, jog tik pirmą ir paskutinį 2018 metų ketvirtį vyko kiek aktyvesnė prekyba įmonės akcijomis, o likusiu laiku Bollinger'io juostos buvo siauros, vadinasi, kintamumas ne itin didelis. Taip pat matyti, kad aukštas kintamumas iššaukė žemą kintamumą, o žemas – aukštesnį. Pati akcijų kaina visu laikotarpiu žemėjo ir tik metų pabaigoje pradėjo kilti.



11 pav. Bollinger'io juostos pagal įmonės AS „Grindeks“ akcijų uždarymo kainos duomenis

Apskaičiuotas efektyvumo indeksas laikotarpiu nuo 2009 iki 2018 metų yra 0,5868. Tai parodo, jog akcijos buvo efektyvios. Gautas tik 2018 metų efektyvumo indekso rezultatas yra kiek aukštesnis – 0,6042. Lyginant su Lietuvos įmone, AS „Grindeks“ akcijos visu analizuojamu laikotarpiu buvo efektyvesnės, tačiau vien tik 2018 metais – jau efektyvesnės AB „Grigeo“.

3.1.1.3. Estijos įmonė AS „Tallinna Kaubamaja Grupp“



12 pav. Bollinger'io juostos pagal įmonės AS „Tallinna Kaubamaja Grupp“ akcijų uždarymo kainos duomenis

12 paveikslėlyje matoma, jog aktyvi prekyba AS „Tallinna Kaubamaja Grupp“ akcijomis vyko pirmaisiais 2018 metų mėnesiais, vėliau akcijų kaina pradėjo mažėti, Bollinger'io juostos siaurėti, kintamumas taip pat buvo nedidelis. Tik metų pabaigoje galima išvelgti nedidelius teigiamų pokyčių ženklus.

Apskaičiuotas šių akcijų efektyvumo indeksas laikotarpiu nuo 2009 iki 2018 metų yra 0,6298. Tai parodo, jog akcijos efektyvios. Gautas vien 2018 metų efektyvumo indeksas kiek didesnis – 0,6506, kas rodo mažesnę efektyvumą.

Apibendrinant galima teigti, jog vertinant ilguoju laikotarpiu, efektyviausios buvo Latvijos įmonės AS „Grindeks“ akcijos, tačiau paskutinių, 2018 metų, duomenys rodo, jog Lietuvos įmonė AB „Grigeo“ demonstravo gerus rezultatus ir tais metais jos akcijos buvo efektyviausios.

Kaip jau aptarta, technine analize paremtoje tyrimo dalyje naudojami trijų įmonių akcijų duomenys. Jų prekybos rodikliai duomenų matricose sudaro po šešis kintamuosius kiekvienu atveju: atidarymo, aukščiausia, žemiausia, uždarymo kainos, prekybos apimtys ir koreguota kaina. Kita dalis kintamųjų yra iš istorinių uždarymo kainos duomenų apskaičiuoti techniniai rodikliai (pasirenkami tie, kurie buvo aprašyti teorinės dalies 1.3. poskyryje), taikant 20 periodų (dienų). Šie rodikliai yra:

- paprastasis slankusis vidurkis (SMA);
- eksponentinis slankusis vidurkis (EMA);
- Bollinger'io juostų indikatorius;
- santykinio stiprumo indeksas (RSI);
- slankiojo vidurkio konvergavimas/divergavimas (MACD);
- stochastinis osciliatorius.

SMA, EMA ir RSI atvejais išvedama po vieną apskaičiuotų duomenų stulpelį. Generuojant Bollinger'io juostų indikatorius duomenis, pateikiamos apatinės (*dn*) ir viršutinės (*up*) juostų reikšmės bei procentiniai skaičiavimai (*pctB*). MACD atveju apskaičiuojamas kainos osciliatorius (*macd*) ir osciliatoriaus signalo linija (slankusis osciliatoriaus vidurkis) (*signal*), o ieškant stochastinio osciliatoriaus reikšmių, išvedami stochastinio greito %K (*fastK*), stochastinio greito %D (*fastD*) ir stochastinio lėto %D (*slowD*) stulpeliai.

Taigi, iš viso kiekvieną duomenų matricą sudaro po vieną priklausomą kintamąjį (kiekvienos akcijos uždarymo kaina) ir 17 nepriklausomų kintamųjų. Kai visi duomenys surinkti ir apskaičiuoti, generuojamas tikslo (angl. *target*) stulpelis, kuris bus būtinas taikant mašininio mokymosi metodus. Šio stulpelio reikšmės yra 1 ir 0. Kai apskaičiuojamas akcijų uždarymo kainos esamos dienos ir praėjusios dienos skirtumas, tada 1 reiškia, jog kaina kils, o 0 – jog kaina nekils (bus tokia pati arba kris). Tam, kad uždarymo kainos kitimo prognozavimas diena į priekį būtų korektiškesnis, pirmieji šeši kintamieji (atidarymo, aukščiausia, žemiausia, uždarymo kainos, prekybos apimtys ir koreguota kaina) paslenkami viena diena į priekį, nekeičiant kitų kintamųjų ir tikslo stulpelio pozicijų. Taip daroma dėl to, kad esamos dienos uždarymo kaina būtų prognozuojama naudojant praeitos dienos prekybos duomenis.

3.1.2. Baltijos šalių akcijų rinkų indeksų kitimo prognozavimo metu naudojami duomenys

Atliekant šį tyrimą, pasirinkti Lietuvos, Latvijos ir Estijos akcijų rinkų indeksų *OMX Tallinn*, *OMX Riga*, *OMX Vilnius* mėnesiniai duomenys, kurie pateikiami *Nasdaq Baltic* interneto svetainėje. Toje pačioje interneto svetainėje teigiama, jog į minėtus indeksus įtraukiamos visos Baltijos šalių akcijų biržų Oficialiajame ir Papildomajame prekybos sąraše esančių bendrovių akcijos. Šie indeksai atspindi dabartinę kiekvienos Baltijos šalies ar bendros visų šalių vertybinių popierių rinkos situaciją ir jos pokyčius [43]. Analizuojamas laikotarpis – nuo 2009-01 iki 2018-12. Tai duomenų matricioje sudaro po 120 eilučių kiekvienos šalies atveju.

13 paveikslėlyje pateikiami analizuojamų indeksų grafikai. Šaltinis: *Nasdaq Baltic* interneto svetainė. Estijos ir Lietuvos indeksų atvejais (oranžinė ir mėlyna linijos) pastebimas trendas, o Latvijos indeksui ilgą laiką „sekėsi“ prasčiausiai, kol nuo 2017 metų beveik susilygino su Lietuvos akcijų rinkų indeksu.



13 pav. Baltijos šalių akcijų rinkų indeksų grafikai

Kaip ir įmonių akcijų analizės metu, taip ir čia apskaičiuojami akcijų rinkų indeksų efektyvumo indeksai (EI) nuo 2009 iki 2018 metų imtinai. Gauti rezultatai parodė, kad efektyviausias tuo laikotarpiu buvo Lietuvos indeksas - 0,5599, nuo jo nedaug atsiliko Estijos - 0,5679, o mažiausiai efektyvus, palyginus su kitais dviem, analizuojamu laikotarpiu buvo Latvijos indeksas, kurio reikšmė yra 0,6088.

Kadangi tai yra fundamentaliąją akcijų analize paremta tyrimo dalis, siekiant įgyvendinti akcijų rinkų indeksų kitimo prognozavimą mėnesiu į priekį, pasirenkama po devyniolika vienodų makroekonominių rodiklių kiekvienos šalies atveju. Visa informacija (pavadinimas, sutrumpintas pavadinimas, kintamojo aprašymas, šaltinis, iš kur paimti duomenys) apie šiuos rodiklius pateikiama 7 lentelėje.

Taigi, iš viso kiekvieną duomenų matricą sudaro po vieną priklausomą kintamąjį (akcijų rinkos indeksą) ir 19 nepriklausomų kintamųjų. Kai visi duomenys surinkti, generuojamas tikslo (angl. *target*) stulpelis, kuris bus būtinas taikant mašininio mokymosi metodus, atliekant klasifikavimą/indeksų reikšmių kitimo prognozavimą. Šio stulpelio reikšmės, kaip ir ankstesnėje

tyrimo dalyje, yra 1 ir 0. Kai apskaičiuojamas akcijų rinkų indeksų esamo mėnesio ir praėjusio mėnesio skirtumas, tada 1 reiškia, jog indekso reikšmė kils, o 0 – jog nekils (bus tokia pati arba kris).

7 lentelė. Tyrime taikomi makroekonominiai rodikliai ir jų aprašymas

Eil. Nr.	Rodiklis	Sutrumpintas pavadinimas	Rodiklio aprašymas	Šaltinis
1.	Vartotojų kainų indeksas	<i>VKI</i>	Pagrindinis vartotojų kainų infliacijos rodiklis	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
2.	Suderintas vartotojų kainų indeksas	<i>SVKI</i>	Vartotojų kainų indeksas, apskaičiuojamas pagal ES mastu suderintą metodologiją	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
3.	Bendrasis vidaus produktas	<i>BVP</i>	Per tam tikrą laikotarpį sukurtų galutinio vartojimo priekių ir paslaugų pinigine išraiška	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
4.	Bendrosios nacionalinės pajamos	<i>BNP</i>	Bendra šalies viduje pagamintų produktų vertė (BVP) ir pajamos, gautos iš kitų šalių (ypač iš palūkanų ir dividendų)	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
5.	Valdžios sektoriaus skola	<i>SKOLA</i>	Pinigų suma, kurią valstybės institucijos yra skolingos skolintojams	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
6.	Nedarbo lygis	<i>NEDARBAS</i>	Nedarbingų asmenų, kurie galėtų ir norėtų dirbti (bedarbių), tačiau neturi tokios galimybės, santykis su visa darbo jėga	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
7.	Užimtumas	<i>UZIMTUMAS</i>	Dirbančių darbingo amžiaus žmonių skaičiaus santykis, palyginti su visa darbo jėga	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
8.	Valiutų kursas	<i>VALIUTOS</i>	Euro (EUR) ir Jungtinių Amerikos Valstijų dolerio (USD) santykis	Europos centrinis bankas
9.	EURIBOR 1	<i>EURIBOR_1</i>	1 mėnesio tarpbankinė palūkanų norma, už kurią euro zonoje bankai yra pasirengę paskolinti lėšų eurai kitiems bankams	Europos pinigų rinkų institutas
10.	EURIBOR 3	<i>EURIBOR_3</i>	3 mėnesių tarpbankinė palūkanų norma, už kurią euro zonoje bankai yra pasirengę paskolinti lėšų eurai kitiems bankams	Europos pinigų rinkų institutas
11.	EURIBOR 6	<i>EURIBOR_6</i>	6 mėnesių tarpbankinė palūkanų norma, už kurią euro zonoje bankai yra pasirengę paskolinti lėšų eurai kitiems bankams	Europos pinigų rinkų institutas
12.	EURIBOR 12	<i>EURIBOR_12</i>	12 mėnesių tarpbankinė palūkanų norma, už kurią euro zonoje bankai yra pasirengę paskolinti lėšų eurai kitiems bankams	Europos pinigų rinkų institutas
13.	Darbo užmokestis (bruto)	<i>UZMOKESTIS</i>	Darbo užmokestis iki mokesčių, tenkantis vienam samdomam darbuotojui	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
14.	Statybos sąnaudų elementų kainų indeksas	<i>STATYBOS_KI</i>	Santykinis rodiklis, kuriuo išreiškiamas naujos statybos statinių sąnaudų elementų bendrasis kainų pokytis per tam tikrą laikotarpį.	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
15.	Pramonės produkcijos kainų indeksas	<i>PRODUKCIJOS_KI</i>	Rodiklis, parodantis produkcijos apimties pokyčius per tam tikrą laikotarpį	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
16.	Valdžios sektoriaus išlaidos	<i>ISLAIDOS</i>	Pinigų suma, kurią išleidžia valstybinės institucijos	Lietuvos/Latvijos/Estijos statistikos duomenų bazė
17.	Vokietijos YTM kreivė 1Y	<i>YTM_1Y</i>	Vienų metų iki obligacijų išpirkimo Vokietijos bazinė pajamingumų kreivė	Europos centrinis bankas
18.	Vokietijos YTM kreivė 3Y	<i>YTM_3Y</i>	Trijų metų iki obligacijų išpirkimo Vokietijos bazinė pajamingumų kreivė	Europos centrinis bankas
19.	Vokietijos YTM kreivė 10Y	<i>YTM_10Y</i>	Dešimties metų iki obligacijų išpirkimo Vokietijos bazinė pajamingumų kreivė	Europos centrinis bankas

Iškeliant prielaidą, jog makroekonominiai veiksniai ne iš karto daro įtaką indeksų kitimui, sudaroma keletas prognozavimo modelių su skirtingais duomenų rinkiniais: kai makroekonominiai duomenys nėra paslenkami, paslenkami vienu ir dviem periodais į priekį arba, kitaip tariant, vėluojama vienu ar dviem mėnesiais. Taip daroma nekeičiant tikslo stulpelio pozicijų.

Apžvelgus abiejose tyrimo dalyse naudojamus duomenis, toliau aprašomi gauti prognozavimo rezultatai.

3.2. Įmonių akcijų kainų kitimo prognozavimo diena į priekį rezultatai

Jau minėta, kad šiame darbe analizuojami Lietuvos įmonės AB „Grigeo“, Latvijos įmonės AS „Grindeks“ ir Estijos įmonės AS „Tallinna Kaubamaja Grupp“ akcijų kainų istoriniai duomenys ir šių kainų kintamumas. Todėl gauti rezultatai toliau pateikiami pagal kiekvieną įmonę atskirai.

3.2.1. Lietuvos įmonė AB „Grigeo“

Atliekant tyrimą, pirmiausia, galimai reikšmingi kintamieji, taikomi prognozuojant akcijų kainos kitimą, atrenkami pagal vieną įprastą kintamųjų atrankos metodą (*RFE*) ir keturis jautrumo analizės metodus. Gauti rezultatai pateikiami 8 lentelėje.

8 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės AB „Grigeo“ atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji
1.	Rekursinis požymių pašalinimo metodas (<i>RFE</i>)	<i>Koreguota kaina, žemiausia kaina</i>
2.	Koreliacija	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, uždarymo kaina, koreguota kaina, SMA, EMA, dn, up</i>
3.	Regresinė analizė	<i>Žemiausia kaina, koreguota kaina, rsi</i>
4.	Dalinė koreliacija	<i>SMA, dn, up</i>
5.	Granger'io priežastingumo ryšio nustatymo testas	<i>Atidarymo kaina, aukščiausia kaina, dn, up, rsi</i>

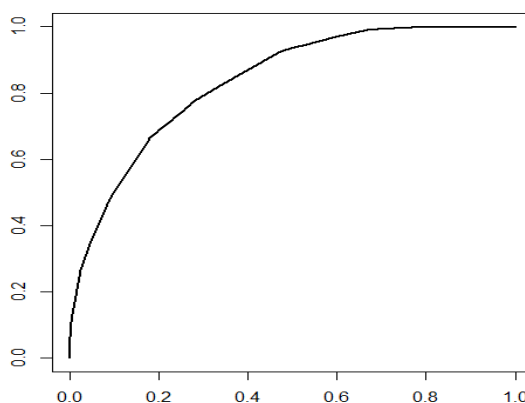
Pastebima, kad iš septyniolikos tyrime taikomų kintamųjų beveik kiekvieno metodo panaudojimo atveju atrenkama tik mažoji jų dalis. Dažniausiai pasitaikančiais galima išskirti *koreguotą kainą, žemiausią kainą, dn, up*. Tai yra vieni tų rodiklių, kurie galimai turėtų neblogai apibūdinti įmonės AB „Grigeo“ akcijų uždarymo kainos dinamiką.

Pagal šiuos gautus rezultatus atliekamas klasifikavimas, panaudojant taip pat penkis metodus. To rezultatai matomi 9 lentelėje. Šioje lentelėje nurodyti pavadinimai atitinka tokias reikšmes: *Metodas* – taikytas klasifikavimo metodas; *Visi kintamieji* – modeliai sudaryti su visais kintamaisiais; *RFE* – modeliai sudaryti, kai kintamieji atrinkti taikant rekursinį požymių pašalinimo metodą; *Koreliacija* – modeliai sudaryti, kai kintamieji atrinkti taikant koreliacinę analizę; *Regresija* - modeliai sudaryti, kai kintamieji atrinkti taikant regresinę analizę; *Dalinė koreliacija* - modeliai sudaryti, kai kintamieji atrinkti taikant dalinę koreliacinę analizę; *Granger'io testas* - modeliai sudaryti, kai kintamieji atrinkti taikant Granger'io priežastingumo ryšio nustatymo testą. Išvesti skaičiai yra ploto po ROC kreive įvertis, kuris vadinamas AUC, ir šiame darbe taikomas kaip pirminis, ir bene svarbiausias, klasifikavimo kokybės įvertinimo matas.

9 lentelė. AB „Grigeo“ duomenų klasifikavimo rezultatai pagal tikslumo matą AUC

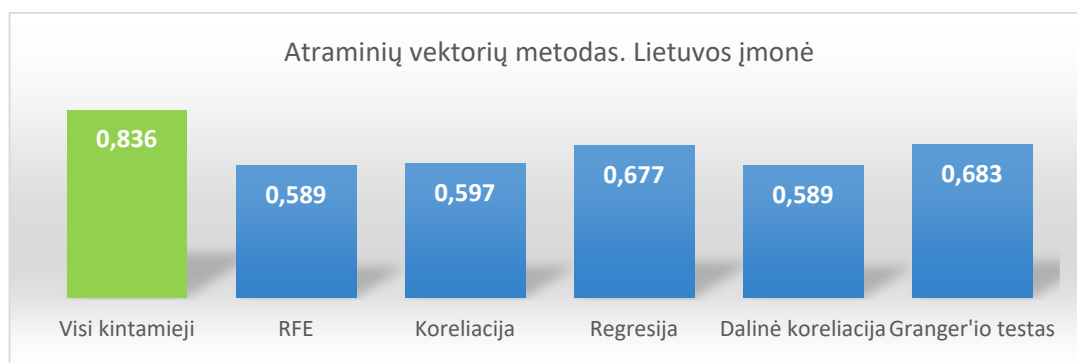
Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger'io testas
RF	0,662	0,513	0,506	0,532	0,502	0,521
kNN	0,500	0,505	0,510	0,539	0,501	0,512
nnet	0,500	0,505	0,510	0,539	0,501	0,512
DT	0,727	0,564	0,563	0,651	0,569	0,620
SVM	0,836	0,589	0,597	0,677	0,589	0,683

Apžvelgus gautus duomenis galima teigti, jog geriausiai klasifikavo atraminių vektorių klasifikatorius (*SVM*). Tiek tada, kai pasirinkti visi kintamieji, tiek tuo metu, kai jie atrinkti pagal skirtingus atrankos metodus. Lentelėje geltona spalva pažymėtas plotas rodo, kuris klasifikatorius klasifikavo geriausiai, kai naudojami visi kintamieji. Raudona spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis nebuvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tos geriausios AUC reikšmės, kurios gautos klasifikuojant pagal skirtingą kintamųjų atrankos metodą. Taigi, apskritai, aukščiausias AUC įvertis 0,836 gautas tada, kai naudoti visi kintamieji ir taikytas Atraminių vektorių metodas.



14 pav. Geriausio modelio ROC kreivė pagal AB „Grigeo“ akcijų kainos duomenis

14 paveikslėlyje matoma išvesta šio geriausio modelio ROC kreivė. Idealiu atveju ji turėtų siekti kairįjį ir viršutinį grafiko kraštus bei sudaryti pusės kvadrato formą. Šiuo atveju taip nėra, tačiau rezultatas nėra prastas. Iš žemiau esančios diagramos (žr. 15 pav.) galima įvertinti, kaip vizualiai atrodo visi AUC įverčiai, kurie gauti taikant *SVM* metodą. Kai kintamieji atrinkti, pritaikant Granger'io priežastingumo ryšio nustatymo testą, rezultatas buvo antras aukščiausias, bet gerokai atsiliekantis nuo pirmojo.



15 pav. Geriausi AUC įverčiai pagal AB „Grigeo“ akcijų kainos duomenis

10 lentelėje pateikiami kiti prognozavimo įvertinimo matai pagal gautus aukščiausius AUC įverčius kiekvieno kintamųjų atrankos metodo pritaikymo atveju atskirai, kai pasirinktas slenkstis yra 0,5 (metodas *SVM*). Tuo atveju, kai klasifikavimo metu naudoti visi kintamieji, gautas tikslumas (angl. *accuracy*) yra didžiausias – 70,34 proc. Preciziškumas (angl. *precision*) ir atkūrimas (angl. *recall*) taip pat turi būti kiek galima didesni, kadangi jie yra galimybės padaryti I ir II rūšies klaidą indikatoriai (kuo reikšmės mažesnės, tuo tikimybės padaryti šias klaidas yra didesnės). Šiuo atveju I rūšies klaida reikštų tai, jog klasifikatorius tam tikrą reikšmę priskirtų kainos kilimo kategorijai, tačiau iš tiesų akcijų kaina nekiltų. Šios rūšies klaidą apibūdina preciziškumas, kuris, kaip matoma pagal klasifikavimo rezultatus lentelėje, nėra labai aukštas. Vis dėlto duomenys pateikti, kai pasirinktas tikslumo slenkstis yra 0,5 (rinktis tokį slenkstį rekomenduojama literatūroje). Jei būtų pasirinktas kitas slenkstis, skaičiai būtų truputį kitokie. Dėl to šiame darbe svarbiausias sudarytų modelių kokybės vertinimo matas bus AUC, o kiti pateiktieji – tik paprasto informacinio pobūdžio. Žvelgiant iš teorinės pusės, kaip minėta 2.2.2 skyrelyje, AUC yra tikimybė, kad modelis „teiks pirmenybę“ klasifikuoti atsitiktinai parinktą teigiamą reikšmę (pvz. kainos didėjimą) nei atsitiktinai parinktą neigiamą (pvz. akcijų kainos nedidėjimą) [49]. Dėl to galima teigti, kad tai yra informatyvus rodiklis.

10 lentelė. AB „Grigeo“ akcijų kainos kitimo prognozavimo tikslumo matai. Slenkstis 0,5

Metodas	Tikslumas	Atkūrimas	Specifiškumas	Preciziškumas	AUC
Visi kintamieji	0,7034	0,8483	0,6276	0,5438	0,836
RFE	0,5355	0,6674	0,4664	0,3956	0,589
Koreliacija	0,5399	0,6884	0,4621	0,4011	0,597
Regresija	0,6088	0,6208	0,6026	0,4497	0,677
Dalinė koreliacija	0,5559	0,6091	0,5281	0,4031	0,589
Granger'io testas	0,6132	0,6301	0,6044	0,4545	0,683

3.2.2. Latvijos įmonė AS „Grindeks“

Kaip pirmos įmonės atveju, taip ir čia, atliekant tyrimą, pirmiausia, galimai reikšmingi kintamieji, taikomi prognozuojant akcijų kainos kitimą, atrinkami pagal penkis šiame darbe naudojamus kintamųjų atrankos metodus. Gauti rezultatai matomi 11 lentelėje.

11 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės AS „Grindeks“ atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji
1.	Rekursinis požymių pašalinimo metodas (RFE)	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, uždarymo kaina, koreguota kaina, SMA, EMA, dn, up, pctB, rsi</i>
2.	Koreliacija	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, uždarymo kaina, koreguota kaina, SMA, EMA, dn, up</i>
3.	Regresinė analizė	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, uždarymo kaina, koreguota kaina, SMA, EMA, dn, pctB, rsi, macd, signal, fastK, fastD, slowD</i>
4.	Dalinė koreliacija	<i>macd, signal, fastK, fastD, slowD</i>
5.	Granger'io priežastingumo ryšio nustatymo testas	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, prekybos apimtys, koreguota kaina, pctB, fastK</i>

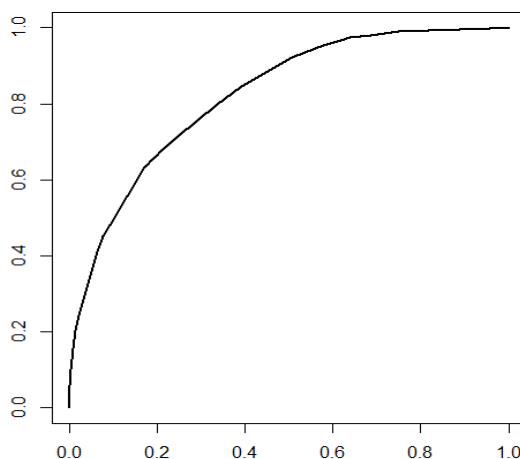
Verta pažymėti, kad čia kintamųjų kiekvieno atrankos metodo pritaikymo atveju atrinkta žymiai daugiau nei tada, kai analizuoti Lietuvos įmonės akcijų kainos duomenys. Dažniausiai pasikartojančiais galima išskirti tokius pagrindinius prekybos rodiklius: *atidarymo kainą, aukščiausią kainą, žemiausią kainą, uždarymo kainą, koreguotą kainą* bei keletą techninių indikatorių: *SMA, EMA, dn, pctB, fastK*. Tai yra vieni tų rodiklių, kurie turėtų turėti nemažą įtaką, prognozuojant įmonės *AS „Grindeks“* akcijų uždarymo kainos dinamiką.

Pagal šiuos gautus rezultatus atliekamas kainos kitimo prognozavimo viena diena į priekį modelių sudarymas, panaudojant penkis klasifikavimo metodus. To rezultatai matomi 12 lentelėje AUC įverčio išraiška.

12 lentelė. AS „Grindeks“ duomenų klasifikavimo rezultatai pagal tikslumo matą AUC

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger'io testas
RF	0,599	0,502	0,503	0,513	0,672	0,628
kNN	0,578	0,589	0,513	0,567	0,663	0,574
nnet	0,578	0,589	0,513	0,567	0,663	0,574
DT	0,722	0,624	0,529	0,717	0,742	0,700
SVM	0,799	0,696	0,577	0,802	0,824	0,721

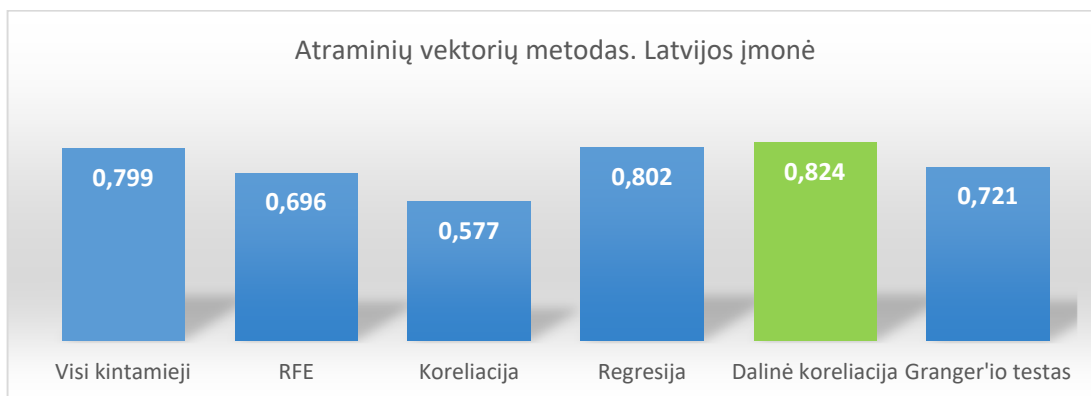
Įvertinus gautus rezultatus galima teigti, jog geriausiai vėl klasifikavo atraminių vektorių klasifikatorius (*SVM*): tiek tada, kai pasirinkti visi kintamieji, tiek tais atvejais, kai jie atrinkti pagal skirtingus kintamųjų atrankos metodus. Lentelėje geltona spalva pažymėtas plotas rodo, kuris klasifikatorius klasifikavo geriausiai, kai naudojami visi kintamieji. Žalia spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis buvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tos geriausios AUC reikšmės, kurios gautos klasifikuojant pagal taikomą kintamųjų atrankos metodą. Taigi, čia skirtumas nuo ankstesnės įmonės yra toks, kad net dviem atvejais buvo prognozuojama geriau, kai kintamieji atrinkti pritaikius regresiją ir dalinę koreliaciją. Pastarasis metodas buvo tas, dėl kurio gauta didžiausia AUC reikšmė – 0,824. Įdomu yra tai, jog čia didžiausią įtaką akcijų kainos kitimui turėjo vien techniniai rodikliai: *macd, signal, fastK, fastD, slowD*. Pagal svarbą juos galima išdėstyti taip: *signal, fastD, macd, slowD, fastK*.



16 pav. Geriausio modelio ROC kreivė pagal AS „Grindeks“ akcijų kainos duomenis

16 paveikslėlyje matoma išvesta šio geriausio modelio ROC kreivė. Kaip minėta anksčiau, idealiu atveju ji turėtų siekti kairiąją ir viršutinę grafiko sienas bei sudaryti pusės kvadrato formą. Šiuo atveju taip nėra, tačiau rezultatas nėra prastas, o net priešingai - galima įvardinti geru.

Žemiau esanti diagrama (žr. 17 pav.) parodo vizualų visų AUC reikšmių, gautų taikant *SVM* metodą, vaizdą. Jei Lietuvos įmonės atveju atotrūkis nuo geriausio modelio prognozavimo kokybės vertinimo rezultato buvo žymiai didesnis, čia net keli gauti įverčiai mažai skiriasi tarpusavyje.



17 pav. Geriausi AUC įverčiai pagal AS „Grindeks“ akcijų kainos duomenis

13 lentelėje matomi kiti šios įmonės akcijų uždarymo kainos kitimo prognozavimo įvertinimo matai tų modelių, kurių AUC reikšmės buvo aukščiausios kiekvieno kintamųjų atrankos metodo pritaikymo atveju atskirai, kai slenkstis yra 0,5 (metodas *SVM*). Modelių sudarymo metu pasirenkant visus kintamuosius, gautas tikslumas yra didžiausias – 73,47 proc., atkūrimas – dalinės koreliacijos pritaikymo metu (76,97 proc.), o preciziškumo didžiausia reikšmė gauta tada, kai vėl pasirenkami visi kintamieji (66,31 proc.). Vis dėlto nereikia pamiršti, kad taikytas slenkstis yra 0,5. Jei būtų kitas slenkstis, rezultatai būtų kitokie. Dėl to šiame darbe, kaip minėta anksčiau, labiau kreipiamas dėmesys į AUC reikšmę.

13 lentelė. AS „Grindeks“ akcijų kainos kitimo prognozavimo tikslumo matai. Slenkstis 0,5

Metodas	Tikslumas	Atkūrimas	Specifiškumas	Preciziškumas	AUC
Visi kintamieji	0,7347	0,6369	0,7962	0,6631	0,799
RFE	0,6281	0,6017	0,6447	0,5160	0,696
Koreliacija	0,5808	0,2780	0,7714	0,4337	0,577
Regresija	0,7230	0,7023	0,7361	0,6263	0,802
Dalinė koreliacija	0,7194	0,7697	0,6878	0,6082	0,824
Granger'io testas	0,6569	0,5550	0,7211	0,5561	0,721

3.2.3. Estijos įmonė AS „Tallinna Kaubamaja Grupp“

Tradiciskai, prieš atliekant tyrimą, pirmiausia, galimai reikšmingi kintamieji, taikomi prognozuojant įmonės akcijų uždarymo kainos kitimą, atrenkami pagal tuos pačius penkis kintamųjų atrankos metodus. Gauti rezultatai matomi 14 lentelėje.

Galima nurodyti, jog tiek su akcijų prekyba tiesiogiai susijusių rodiklių, tiek techninių indikatorių čia išskiriama apylygiai. Dažniausiai pasikartojantys yra šie: *aukščiausia kaina*, *žemiausia kaina*, *koreguota kaina*, *EMA*, o kintamasis *prekybos apimtys* visų trijų įmonių atvejais atrenkamas itin retai. Visi minėtieji kintamieji yra vieni tų rodiklių, kurie galimai turėtų turėti nemažą įtaką, prognozuojant įmonės *AS „Tallinna Kaubamaja Grupp“* akcijų uždarymo kainos kitimą.

14 lentelė. Atrinkti reikšmingi kintamieji pagal įmonės *AS „Tallinna Kaubamaja Grupp“* atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji
1.	Rekursinis požymių pašalinimo metodas (RFE)	<i>Žemiausia kaina, koreguota kaina</i>
2.	Koreliacija	<i>Atidarymo kaina, aukščiausia kaina, žemiausia kaina, uždarymo kaina, koreguota kaina, SMA, EMA, dn, up</i>
3.	Regresinė analizė	<i>Aukščiausia kaina, koreguota kaina, SMA, EMA, macd, signal</i>
4.	Dalinė koreliacija	<i>EMA, rsi, macd, signal</i>
5.	Granger'io priežastingumo ryšio nustatymo testas	<i>Aukščiausia kaina, žemiausia kaina, prekybos apimtys, dn, up, pctB, rsi, fastK, fastD</i>

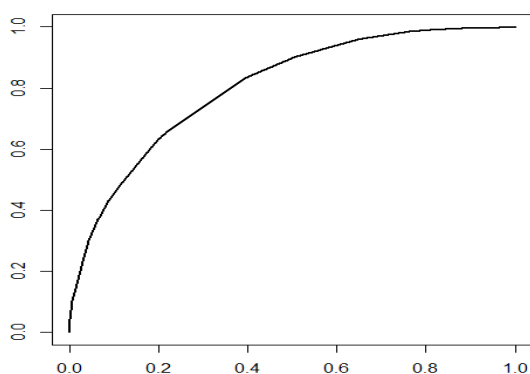
Pagal šiuos gautus rezultatus atliekamas kainos kitimo prognozavimo viena diena į priekį modelių sudarymas, panaudojant penkis klasifikavimo metodus. Gauti rezultatai matomi 15 lentelėje AUC įverčio išraiška.

15 lentelė. *AS „Tallinna Kaubamaja Grupp“* duomenų klasifikavimo rezultatai pagal tikslumo matą AUC

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger'io testas
RF	0,540	0,503	0,504	0,516	0,521	0,627
kNN	0,510	0,504	0,516	0,521	0,555	0,502
nnet	0,510	0,504	0,516	0,521	0,555	0,502
DT	0,726	0,516	0,514	0,514	0,635	0,721
SVM	0,805	0,525	0,538	0,545	0,693	0,791

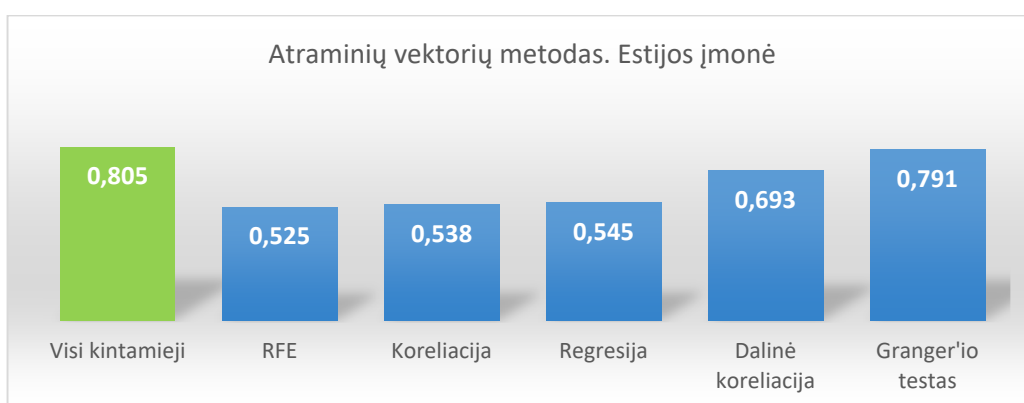
Apžvelgus gautus rezultatus galima teigti, jog vėl, kaip ir ankstesniais kartais, geriausiai klasifikavo atraminių vektorių klasifikatorius (*SVM*): tiek tada, kai pasirinkti visi kintamieji, tiek tais atvejais, kai jie atrinkti pagal atskirus kintamųjų atrankos metodus. Lentelėje geltona spalva pažymėtas plotas rodo, kuris klasifikatorius klasifikavo geriausiai, kai naudojami visi kintamieji. Raudona spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis nebuvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tų geriausių modelių AUC reikšmės, kurios gautos klasifikuojant pagal skirtingą jautrumo analizės metodą. Kintamųjų atrankos metodų prasme šie rezultatai yra tokie patys kaip ir Lietuvos įmonės atveju: geriausiai klasifikuojama, kai naudojami visi kintamieji, antras geriausias rezultatas gaunamas, kai pritaikomas Granger'io priežastingumo ryšio nustatymo testas, kokybiškiausiai visa tai darantis klasifikatorius kaip, beje, ir Latvijos įmonės atveju – *SVM*. Vienintelis skirtumas čia yra toks, jog AUC reikšmė (0,791), gauta kintamųjų atrankai panaudojus Granger'io priežastingumo ryšio nustatymo testą, yra mažai besiskirianti nuo tos, kuri gauta prognozavimui pasirinkus visus kintamuosius – 0,805.

18 paveikslėlyje matoma nubrėžta geriausio gauto modelio ROC kreivė. Galima pasikartoti, kad idealiu atveju ji turėtų siekti kairiąją ir viršutinę grafiko sienas bei sudaryti pusės kvadrato formą. Šiuo atveju taip nėra, tačiau rezultatą vis tiek galima įvardinti geru.



18 pav. Geriausio modelio ROC kreivė pagal AS „Tallinna Kaubamaja Grupp“ akcijų kainos duomenis

Žemiau esanti diagrama (žr. 19 pav.) parodo vizualų visų AUC reikšmių, gautų taikant *SVM* metodą, vaizdą. Matoma, kad šiuo atveju „konkurencija“ yra didesnė tik tarp dviejų sudarytų geriausių modelių (reikšmės 0,805 ir 0,791). Kiti gauti AUC įverčiai, kai taikomi skirtingi kintamųjų atrankos metodai, yra žymiai mažesni.



19 pav. Geriausi AUC įverčiai pagal AS „Tallinna Kaubamaja Grupp“ akcijų kainos duomenis

Kiti AS „Tallinna Kaubamaja Grupp“ akcijų uždarymo kainos geriausių prognozavimo modelių tikslumo vertinimo matai, kai pasirinktas slenkstis 0,5, pateikiami 16 lentelėje. Beveik pagal visus vertinamus rodiklius geriausiai „atrodo“ tas modelis, kuris sudaromas pasirinkus visus kintamuosius: tikslumas – 70,90 proc., atkūrimas – 75,55 proc., preciziškumas – 63,98 proc. Kadangi tokie matai šiame darbe yra tik paprasto informacinio pobūdžio, toliau detalai nenagrinėjami.

16 lentelė. AS „Tallinna Kaubamaja Grupp“ akcijų kainos kitimo prognozavimo tikslumo matai. Slenkstis 0,5

Metodas	Tikslumas	Atkūrimas	Specifiškumas	Preciziškumas	AUC
Visi kintamieji	0,7090	0,7555	0,6733	0,6398	0,805
RFE	0,5070	0,4336	0,5634	0,4328	0,525
Koreliacija	0,5299	0,3939	0,6343	0,4528	0,538
Regresija	0,5343	0,4474	0,6010	0,4628	0,545
Dalinė koreliacija	0,6441	0,5996	0,6782	0,5888	0,693
Granger'io testas	0,6978	0,7454	0,6612	0,6283	0,791

Aprašius gautus rezultatus galima apibendrinti pirmą tyrimo dalį, kai analizuota galimybė kuo tiksliau prognozuoti įmonių akcijų uždarymo kainų kitimą viena diena į priekį, pritaikant vieną įprastą ir keturis jautrumo analizės metodus, siekiant atrinkti reikšmingus kintamuosius. Pastebėta, jog pritaikius jautrumo analizės metodus įmanoma pagerinti sudaromų modelių prognozavimo tikslumą. Kai kintamieji atrinkti pagal rekursinį požymių pašalinimo metodą (*RFE*), tikslumas nė vienu atveju nebuvo itin aukštas. Nors Lietuvos ir Estijos įmonių atvejais geriausi modeliai buvo sudaryti tada, kai prognozavimui naudoti visi kintamieji, vis dėlto tikrai vertėtų atkreipti dėmesį į Granger'io priežastingumo ryšio nustatymo testą, kai parodoma, ar egzistuoja priežastinis ryšys tarp priklausomo ir nepriklausomų kintamųjų. Naudojant šiuo testu atrinktus veiksnius, geriausių sudarytų modelių AUC vidurkis buvo 0,732. Taip pat geri modeliai buvo gauti, kai kintamieji atrinkti pagal dalinės koreliacijos (AUC reikšmių vidurkis 0,702) ir regresijos (AUC reikšmių vidurkis 0,675) įgyvendinimo rezultatus. Kalbant apie klasifikavimo metodus, tokiems duomenims vienareikšmiškai reikėtų rinktis atraminių vektorių klasifikatorių.

Pačių geriausių sudarytų modelių AUC įverčiai (0,836, 0,824 ir 0,805) parodė, jog panaudojant vien tik istorinius akcijų kainos duomenis galima gauti gerus prognozavimo rezultatus viena diena į priekį. Weng'as [33], kuris taip pat atliko akcijų kainų kitimo prognozavimo viena diena į priekį tyrimą, savo duomenų matricą papildė su naujienomis susijusiais požymiais dėl ko, kaip manoma, prognozavimo tikslumas būdavo aukštas. Jo geriausių sudarytų modelių AUC buvo apie 0,89. Tai yra tik nedaug aukštesnis rezultatas nei gauta šio tyrimo metu.

Analizuotos įmonės buvo parinktos po vieną iš kiekvienos Baltijos šalies, tačiau esminių skirtumų tarp valstybių neįžvelgiama – visais atvejais akcijų kainų kitimas viena diena į priekį prognozuotas panašiai gerai.

3.3. Akcijų rinkų indeksų kitimo prognozavimo mėnesiu į priekį rezultatai

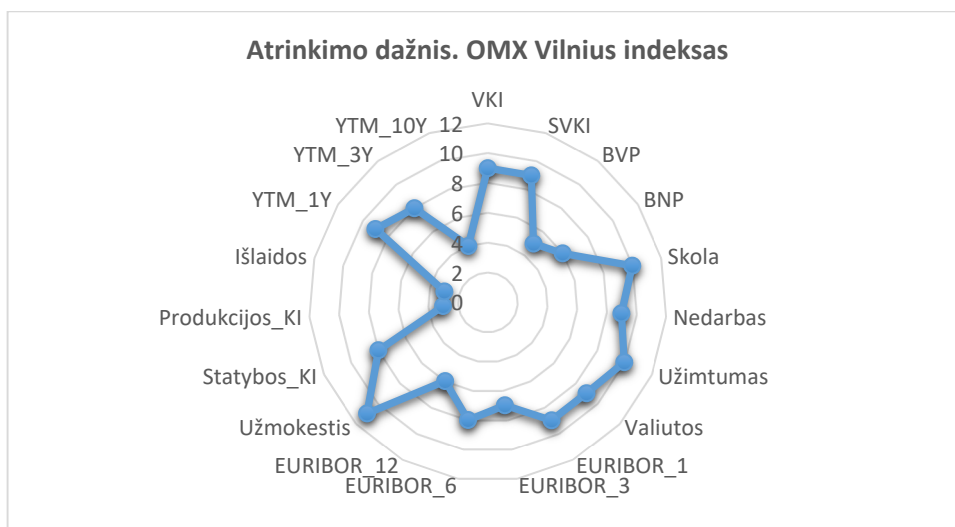
Norint atlikti kitimo prognozavimo mėnesiu į priekį analizę, pasirinkti visų trijų Baltijos šalių akcijų rinkų indeksai: *OMX Vilnius*, *OMX Riga*, *OMX Tallinn* bei devyniolika makroekonominių rodiklių. Spėjant, jog ekonominiai pokyčiai gali ne iš karto turėti įtakos indeksų svyravimui, prognozavimo modeliai sudaromi keliais etapais: kai makroekonominiai rodikliai nepaslenkami, paslenkami vienu periodu (mėnesiu) į priekį ir paslenkami dviem periodais į priekį arba, kitaip tariant, vėluojama vienu ar dviem mėnesiais. Gauti rezultatai toliau pateikiami pagal kiekvieną indeksą atskirai.

3.3.1. Lietuvos akcijų rinkos indeksas OMX Vilnius

Pirmiausia, kaip ir pirmoje tyrimo dalyje, pritaikius penkis kintamųjų atrankos metodus, atrenkami reikšmingi kintamieji, kurie galimai galėtų turėti nemažos įtakos, prognozuojant Lietuvos akcijų rinkos indekso kitimą mėnesiu į priekį. 17 lentelėje jie matomi susisteminti pagal atskirą metodą ir paslinkimo lygį, o 20 paveikslėlis atspindi, kaip dažnai tam tikras rodiklis buvo išskiriamas kaip reikšmingas.

17 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Vilnius indekso atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji		
		Nepaslinkta	Paslinkta 1 periodu	Paslinkta 2 periodais
1.	Rekursinis požymių pašalinimo metodas (RFE)	VKI, SVKI, skola, nedarbas, užimtumas, EURIBOR_6/12, užmokestis, statybos_KI, produkcijos_KI	VKI, SVKI, skola, nedarbas, užimtumas, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y	VKI, SVKI, skola, nedarbas, užimtumas, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y_3Y
2.	Koreliacija	VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, YTM_1Y/3Y/10Y	VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, YTM_1Y/3Y/10Y	VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, YTM_1Y/3Y/10Y
3.	Regresinė analizė	VKI, SVKI, skola, nedarbas, valiutos, EURIBOR_1, užmokestis, statybos_KI, išlaidos	VKI, SVKI, BNP, skola, nedarbas, valiutos, EURIBOR_1, užmokestis, išlaidos	VKI, SVKI, skola, nedarbas, užimtumas, valiutos, EURIBOR_1, užmokestis, išlaidos
4.	Dalinė koreliacija	Skola, YTM_1Y/3Y/10Y	Statybos_KI, YTM_1Y/3Y	BVP, BNP, valiutos, EURIBOR_3, YTM_1Y/3Y
5.	Granger'io priežastingumo ryšio nustatymo testas	BNP, užimtumas, EURIBOR_6	Užimtumas, valiutos, užmokestis	BVP, užimtumas, valiutos, EURIBOR_1/3/6, užmokestis, YTM_1Y/3Y



20 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Vilnius indeksas

Dažniausiai pasikartojusiais rodikliais galima nurodyti darbo užmokestį, gyventojų užimtumą, valdžios sektoriaus skolą, o rečiausiai - valdžios sektoriaus išlaidas, pramonės produkcijos kainų indeksą, Vokietijos bazinės pajamingumą kreivės 10 metų iki obligacijų išpirkimo duomenis, BVP. Pastebima, kad duomenų paslinkimas vienu ar dviem periodais į priekį turėjo įtakos tam, kad kai kurie rodikliai būdavo nurodomi kaip reikšmingi, jei taip nebuvo anksčiau. Pavyzdžiui, dalinės koreliacijos ir Granger'io priežastingumo ryšio nustatymo atvejais, BVP bei analizuojamo indekso dinamikos ryšys išryškėdavo tik po dviejų mėnesių.

Kaip ankstesnėje tyrimo dalyje, taip ir čia pagal gautus reikšmingų faktorių atrankos rezultatus atliekamas indekso kitimo prognozavimas vienu mėnesiu į priekį modelių sudarymas. Pritaikomi tie

patys anksčiau aptarti penki klasifikavimo metodai: atsitiktinių miškų (*RF*), artimiausių kaimynų (*kNN*), neuroninių tinklų (*nnet*), sprendimų medžio (*DT*) bei atraminių vektorių (*SVM*). Gauti modelių kokybės vertinimo įverčiai AUC pateikiami 18, 19, 20 lentelėse pagal duomenų paslinkimo lygį atskirai.

18 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,532	0,531	0,562	0,547	0,574	0,607
kNN	0,525	0,504	0,514	0,528	0,523	0,520
nnet	0,525	0,504	0,514	0,528	0,523	0,520
DT	0,542	0,504	0,504	0,509	0,527	0,525
SVM	0,512	0,515	0,504	0,513	0,518	0,543

19 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,534	0,589	0,591	0,564	0,599	0,557
kNN	0,504	0,504	0,540	0,592	0,504	0,544
nnet	0,504	0,504	0,540	0,592	0,504	0,544
DT	0,521	0,558	0,535	0,505	0,590	0,540
SVM	0,523	0,508	0,523	0,540	0,560	0,614

20 lentelė. OMX Vilnius indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,588	0,579	0,611	0,556	0,522	0,579
kNN	0,571	0,539	0,557	0,541	0,504	0,532
nnet	0,571	0,539	0,557	0,541	0,504	0,532
DT	0,532	0,534	0,523	0,511	0,504	0,517
SVM	0,507	0,514	0,504	0,522	0,504	0,519

Apžvelgus gautus rezultatus galima teigti, jog čia, priešingai nei pirmoje dalyje, nebėra vieno metodo, kuris klasifikuotų geriausiai. Pagal klasifikavimo kokybę vyrauja atsitiktinių miškų klasifikatoriaus modeliai, tačiau aukščiausia AUC reikšmė (0,614) gauta tada, kai buvo taikytas atraminių vektorių metodas, paslinkus duomenis vienu periodu į priekį ir panaudojant Granger'io priešastingumo ryšio tikrinimo metu atrinktus kintamuosius. 18, 19, 20 lentelėse geltona spalva pažymėti plotai rodo, kuris klasifikatorius klasifikavo geriausiai, kai taikomi visi kintamieji. Žalia spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis buvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tų geriausių modelių AUC reikšmės, kurios gautos klasifikuojant pagal skirtingą kintamųjų atrankos metodą. Patys kokybės vertinimo įverčiai nėra labai aukšti nė vienu atveju, tačiau lyginant, kai klasifikuojama su visais duomenimis, ir kai kintamieji yra atrenkami, pastarojo būdo pritaikymo metu rezultatai akivaizdžiai pagerėja. Taip pat matoma, kad duomenų paslinkimas pasiteisino,

kadangi neretai AUC reikšmės gautos aukštesnės, nei to nedarius. Taigi, įvertinus visus sudarytus modelius galima teigti, jog geriausiai buvo klasifikuojama, kai kintamieji atrinkti pritaikius Granger'io priežastingumo ryšio nustatymo testą ir koreliacinę analizę. O pačiam geriausiam modeliui (pagal AUC reikšmę 0,614) gauti užteko iš visų devyniolikos kintamųjų, jog būtų panaudoti šie rodikliai: *gyventojų užimtumas, valiutų kursas, darbo užmokestis*. Jie, žvelgiant iš jautrumo analizės pusės, pagal svarbą pasiskirstė taip: *valiutų kursas, gyventojų užimtumas, darbo užmokestis*.

3.3.2. Latvijos akcijų rinkos indeksas OMX Riga

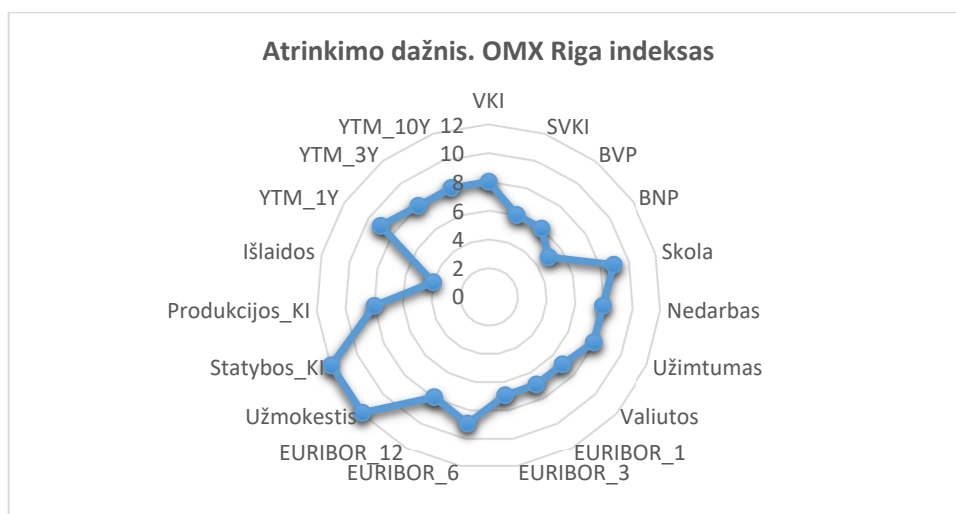
Iš pradžių, kaip ir visais kitais atvejais, pritaikius penkis kintamųjų atrankos metodus, atrenkami reikšmingi kintamieji, kurie galimai galėtų turėti įtakos, prognozuojant Latvijos akcijų rinkos indekso kitimą mėnesiu į priekį. 21 lentelėje jie matomi susisteminti pagal atskirą metodą ir duomenų paslinkimo lygį, o įvertinus 21 paveikslėlyje pateiktą grafiką galima sužinoti, kaip dažnai kiekvienas nepriklausomas kintamasis buvo išskiriamas, kaip turintis įtakos priklausomam kintamajam - akcijų rinkos indeksui.

21 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Riga indekso atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji		
		Nepaslinkta	Paslinkta 1 periodu	Paslinkta 2 periodais
1.	Rekursinis požymių pašalinimo metodas (RFE)	<i>VKI, SVKI, BVP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, išlaidos, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, skola, nedarbas, užimtumas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y/3Y/10Y</i>
2.	Koreliacija	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, valiutos, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, YTM_1Y/3Y/10Y</i>
3.	Regresinė analizė	<i>Skola, nedarbas, užimtumas, užmokestis, statybos_KI, išlaidos, YTM_1Y/3Y</i>	<i>Skola, nedarbas, užimtumas, užmokestis, išlaidos, YTM_1Y</i>	<i>Skola, užimtumas, EURIBOR_1, užmokestis, išlaidos, YTM_10Y</i>
4.	Dalinė koreliacija	<i>BNP, EURIBOR_6, užmokestis, statybos_KI, produkcijos_KI</i>	<i>Užimtumas, EURIBOR_6/12, statybos_KI, produkcijos_KI, YTM_1Y/3Y</i>	<i>Valiutos, EURIBOR_12, užmokestis, statybos_KI</i>
5.	Granger'io priežastingumo ryšio nustatymo testas	<i>VKI, užimtumas, užmokestis, YTM_10Y</i>	<i>VKI, statybos_KI</i>	<i>EURIBOR_3/6, statybos_KI</i>

Kaip matoma grafike (žr. 21 pav.), dažniausiai pasikartojusiais rodikliais galima nurodyti *darbo užmokestį ir statybos sąnaudų elementų kainų indeksą*, o rečiausiai - *valdžios sektoriaus išlaidas, bendrąsias nacionalines pajamas, BVP, suderintą vartotojų kainų indeksą*. Atkreipiamas dėmesys, kad tokių rodiklių kaip *darbo užmokestis, valdžios sektoriaus skola, valdžios sektoriaus išlaidos, BVP* panašus atrinkimo dažnis buvo būdingas ir Lietuvos indekso atveju. Apžvelgus 21 lentelę

pastebima, kad duomenų paslinkimas vienu ar dviem periodais į priekį turėjo įtakos tam, kad kai kurie rodikliai būdavo nurodomi kaip reikšmingi, jei taip nebuvo anksčiau, ir atvirkščiai.



21 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Riga indeksas

Kaip jau tapo įprasta, nustačius įtakos priklausomam kintamajam turinčius nepriklausomus kintamuosius, atliekamas akcijų rinkos indekso kitimo prognozavimo vienu mėnesiu į priekį modelių sudarymas. Tam vėl pritaikomi tie patys anksčiau minėti penki klasifikavimo metodai. Gauti modelių kokybės vertinimo įverčiai AUC pateikiami 22, 23, 24 lentelėse pagal duomenų paslinkimo lygį atskirai.

22 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,622	0,637	0,659	0,565	0,619	0,579
kNN	0,502	0,502	0,506	0,502	0,505	0,506
nnet	0,502	0,502	0,506	0,502	0,505	0,506
DT	0,516	0,526	0,515	0,518	0,502	0,566
SVM	0,502	0,511	0,517	0,516	0,517	0,526

23 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,634	0,634	0,638	0,523	0,624	0,502
kNN	0,505	0,505	0,502	0,512	0,507	0,531
nnet	0,505	0,505	0,502	0,512	0,507	0,531
DT	0,502	0,502	0,517	0,503	0,502	0,502
SVM	0,504	0,504	0,509	0,539	0,516	0,557

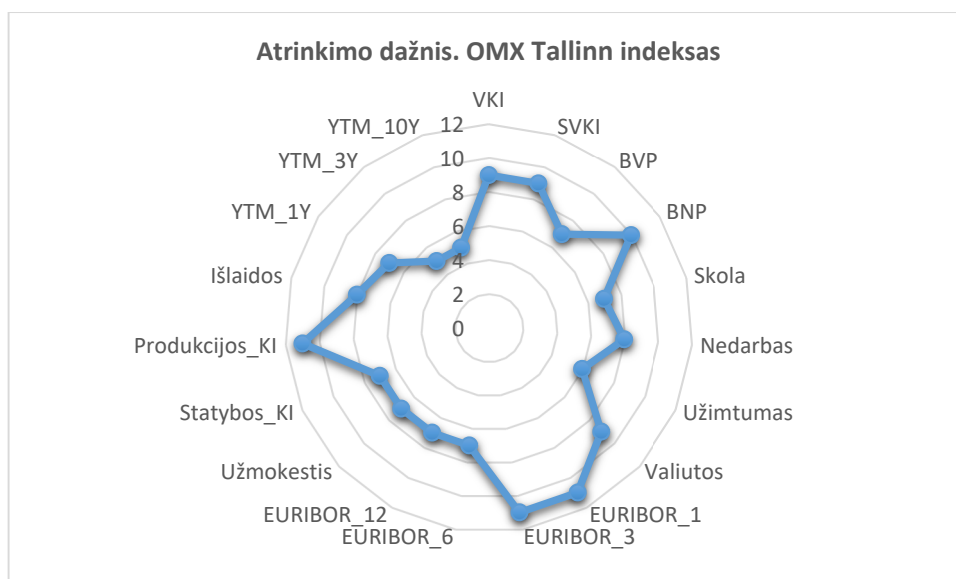
24 lentelė. OMX Riga indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,615	0,622	0,639	0,543	0,615	0,539
kNN	0,505	0,503	0,502	0,549	0,507	0,515
nnet	0,505	0,503	0,502	0,549	0,507	0,515
DT	0,502	0,502	0,502	0,506	0,502	0,514
SVM	0,502	0,509	0,532	0,510	0,502	0,556

Įvertinus gautus rezultatus, pirmiausia, pastebima tai, jog čia AUC reikšmės, apskritai, yra aukštesnės nei Lietuvos indekso kitimo prognozės atveju. Vertinant pagal klasifikatorių, vėl vyrauja atsitiktinių miškų metodu sudaryti modeliai. Aukščiausias AUC įvertis (0,659) gautas tada, kai buvo taikytas atsitiktinių miškų metodas, duomenys nepaslinkti nė vienu periodu į priekį, o reikšmingi kintamieji atrinkti pagal koreliaciją. Aukščiau matomose lentelėse, kaip ir anksčiau, geltona spalva pažymėti plotai rodo, kuris klasifikatorius klasifikavo geriausiai, kai taikomi visi kintamieji. Žalia spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis buvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tų geriausių modelių AUC reikšmės, kurios gautos klasifikuojant pagal skirtingą kintamųjų atrankos metodą. Lyginant, kai klasifikuojama su visais duomenimis, ir kai kintamieji yra atrenkami, pastarojo būdo pritaikymo metu rezultatai pagerėja keliais punktais. Šį kartą duomenų paslinkimas neturėjo didesnės įtakos, norint sudaryti geriausią modelį, – toks gautas tada, kai duomenys nebuvo paslenkami. Taigi, įvertinus visus gautus rezultatus galima teigti, jog geriausiai buvo klasifikuojama, kai kintamieji atrinkti pritaikius koreliaciją, o taikomas klasifikatorius - atsitiktinių miškų metodas. Šiuo atveju, sudarant geriausią modelį, iš visų devyniolikos kintamųjų buvo panaudojama net septyniolika. Du nepatekė į sąrašą yra: *gyventojų užimtumas* ir *valdžios sektoriaus išlaidos*. Vertinant pagal jautrumo analizės aspektus, keli svarbiausi kintamieji yra: *nedarbas*, *darbo užmokestis*, *dvylikos ir šešių mėnesių EURIBOR palūkanų normos*.

3.3.3. Estijos akcijų rinkos indeksas OMX Tallinn

Pirmiausia, kaip ir ankstesniais atvejais, atrenkami reikšmingi kintamieji, kurie galimai galėtų turėti didesnės ar mažesnės įtakos, prognozuojant Estijos indekso kitimą mėnesiu į priekį. 25 lentelėje jie matomi susisteminti pagal atskirą metodą ir duomenų paslinkimo lygį, o įvertinus 22 pav. pateiktą grafiką galima sužinoti, kaip dažnai kiekvienas kintamasis buvo išskiriamas, kaip turintis įtakos indekso reikšmių pasikeitimui.



22 pav. Dažniausiai atrinkti makroekonominiai rodikliai. OMX Tallinn indeksas

25 lentelė. Atrinkti reikšmingi makroekonominiai rodikliai pagal OMX Tallinn indekso atvejį

Eil. Nr.	Metodo pavadinimas	Atrinkti kintamieji		
		Nepaslinkta	Paslinkta 1 periodu	Paslinkta 2 periodais
1.	Rekursinis požymių pašalinimo metodas (RFE)	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, valiutos, EURIBOR_1/3/6/12, statybos_KI, produkcijos_KI</i>	<i>VKI, SVKI, skola, valiutos, EURIBOR_1/3/6/12, statybos_KI, produkcijos_KI</i>	<i>VKI, SVKI, skola, EURIBOR_3/6/12, produkcijos_KI</i>
2.	Koreliacija	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, išlaidos, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, išlaidos, YTM_1Y/3Y/10Y</i>	<i>VKI, SVKI, BVP, BNP, skola, nedarbas, užimtumas, EURIBOR_1/3/6/12, užmokestis, statybos_KI, produkcijos_KI, išlaidos, YTM_1Y/3Y/10Y</i>
3.	Regresinė analizė	<i>VKI, SVKI, BNP, nedarbas, užimtumas, valiutos, EURIBOR_1, produkcijos_KI, išlaidos, YTM_10Y</i>	<i>VKI, SVKI, BNP, nedarbas, užimtumas, valiutos, EURIBOR_1, užmokestis, produkcijos_KI, išlaidos</i>	<i>VKI, SVKI, BNP, nedarbas, užimtumas, valiutos, EURIBOR_1, užmokestis, produkcijos_KI</i>
4.	Dalinė koreliacija	<i>Valiutos, EURIBOR_3, YTM_1Y/3Y/10Y</i>	<i>Nedarbas, valiutos, EURIBOR_3/6/12, statybos_KI</i>	<i>Skola, EURIBOR_1/3, YTM_1Y/3Y</i>
5.	Granger'io priežastingumo ryšio nustatymo testas	<i>BVP, BNP, valiutos, EURIBOR_1_3, produkcijos_KI, išlaidos</i>	<i>BVP, BNP, valiutos, užmokestis, produkcijos_KI, išlaidos, YTM_1Y</i>	<i>BVP, BNP, EURIBOR_1_3, užmokestis, statybos_KI, išlaidos, YTM_1Y</i>

Kaip matoma grafike (žr. 22 pav.), dažniausiai pasikartojusiais rodikliais galima nurodyti *pramonės produkcijos kainų indeksą, vieno ir trijų mėnesių EURIBOR palūkanų normas, bendrąsias nacionalines pajamas*, o rečiausiai - *Vokietijos bazinės pajamingumo kreivės 3 ir 10 metų iki obligacijų išpirkimo duomenis, gyventojų užimtumą*. Tokie rezultatai kiek skiriasi nuo Lietuvos ir Latvijos atvejų. Pavyzdžiui, Lietuvos ir Latvijos *bendrosios nacionalinės pajamos* buvo išskiriamos retai kaip reikšmingos, o Estijos – net 10 kartų. *Gyventojų užimtumo* rodiklis Lietuvos atveju pasikartojė 10 kartų, Latvijos – 8, o Estijos – tik 6. Lietuvos ir Latvijos *valdžios sektoriaus išlaidų* išskyrimas buvo retas reiškinys (atitinkamai 3 ir 4 kartai), o Estijos – 8 kartai.

Įvertinus 25 lentelės įrašus pastebima, kad duomenų paslinkimas vienu ar dviem periodais į priekį vėl turėjo įtakos tam, kad kai kurie rodikliai būdavo nurodomi kaip reikšmingi, jei taip nebuvo anksčiau, ir atvirkščiai.

Kaip jau tapo įprasta, nustačius įtakos priklausomam kintamajam turinčius nepriklausomus kintamuosius pagal skirtingus kintamųjų atrinkimo metodus, atliekamas akcijų rinkos indekso kitimo prognozavimo vienu mėnesiu į priekį modelių sudarymas. Tam vėl pritaikomi tie patys anksčiau minėti penki klasifikavimo metodai. Gauti modelių kokybės vertinimo įverčiai AUC pateikiami 26, 27 ir 28 lentelėse pagal duomenų nepaslinkimo, paslinkimo vienu periodu ir paslinkimo dviem periodais į priekį kategorijas.

26 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys nepaslinkti

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,523	0,573	0,527	0,556	0,534	0,565
kNN	0,519	0,516	0,535	0,562	0,523	0,527
nnet	0,519	0,516	0,535	0,562	0,523	0,527
DT	0,503	0,526	0,510	0,503	0,503	0,503
SVM	0,503	0,532	0,515	0,505	0,525	0,522

27 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 1 periodu

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,541	0,560	0,520	0,536	0,556	0,543
kNN	0,516	0,512	0,503	0,507	0,503	0,503
nnet	0,516	0,512	0,503	0,507	0,503	0,503
DT	0,511	0,503	0,503	0,503	0,503	0,503
SVM	0,504	0,516	0,504	0,523	0,567	0,533

28 lentelė. OMX Tallinn indekso kitimo prognozavimo modelių AUC įverčiai. Duomenys paslinkti 2 periodais

Metodas	Visi kintamieji	RFE	Koreliacija	Regresija	Dalinė koreliacija	Granger testas
RF	0,545	0,569	0,537	0,550	0,547	0,556
kNN	0,526	0,512	0,503	0,526	0,503	0,550
nnet	0,526	0,512	0,503	0,526	0,503	0,550
DT	0,503	0,503	0,503	0,547	0,587	0,506
SVM	0,505	0,534	0,504	0,503	0,503	0,512

Įvertinus gautus rezultatus, pirmiausia, pastebima tai, jog čia AUC reikšmės, apskritai, yra žemesnės nei Lietuvos ir Latvijos akcijų rinkų indeksų kitimo prognozės atvejais. Vertinant pagal klasifikatorių, kaip ir dviem atvejais prieš tai, vyrauja atsitiktinių miškų metodas bei juo sudaryti modeliai. Vis dėlto aukščiausias AUC įvertis (0,587) gautas tada, kai taikytas sprendimų medžio klasifikatorius, o duomenys paslinkti dviem periodais į priekį. Reikšmingi kintamieji čia atrinkti pagal dalinės koreliacijos rezultatus. 26, 27 ir 28 lentelėse, kaip ir anksčiau, geltona spalva pažymėti ploteliai rodo, kuris klasifikatorius klasifikavo geriausiai, kai taikomi visi kintamieji. Žalia spalva apibūdina geriausią rezultatą, kai kintamieji yra atrenkami, o kartu ir tai, jog toks įvertis buvo geresnis nei tada, kai klasifikuojama su visais duomenimis. Pilka spalva pažymėtos tų geriausių modelių AUC reikšmės, kurios gautos klasifikuojant pagal skirtingą kintamųjų atrankos metodą. Lyginant, kai klasifikuojama su visais duomenimis, ir kai kintamieji yra atrenkami, pastarojo būdo pritaikymo metu rezultatai pagerėja keliais punktais. Vis dėlto tai guodžia mažai, kadangi pati klasifikavimo kokybė nėra itin gera: AUC reikšmės nė vienu atveju nesiekia net 0,600. Šį kartą duomenų paslinkimas turėjo įtakos, norint sudaryti bent kiek geresnį modelį, – toks gautas tada, kai jie buvo paslenkami dviem periodais į priekį. Taigi, apibendrinant gautus rezultatus galima teigti, jog geriausiai buvo klasifikuojama, kai kintamieji atrinkti pritaikius dalinę koreliaciją, o naudotas klasifikatorius - sprendimų medžio metodas (DT). Tai nulėmę kintamieji buvo: *valdžios sektoriaus skola, vieno ir trijų mėnesių EURIBOR palūkanų normos bei Vokietijos bazinės pajamingumo kreivės 1 ir 3 metų iki obligacijų išpirkimo duomenys*.

Nors sudarytų geriausių akcijų rinkų indeksų kitimo prognozavimo mėnesiu į priekį modelių AUC reikšmės ir nėra itin aukštos nė vienos šalies atveju, tačiau reikšmingų kintamųjų atrinkimas, pritaikant vieną įprastą kintamųjų atrankos ir keturis jautrumo analizės metodus, turėjo įtakos tam, jog klasifikavimo tikslumas pagerėdavo keliais procentais. Duomenų paslinkimas vienu ar dviem periodais taip pat neretai turėjo teigiamos įtakos. Aukščiausi AUC įverčiai gauti analizuojant Latvijos indeksą, o žemiausi – Estijos. Vertinant pagal mašininio mokymosi metodą, akivaizdžiai dominavo atsitiktinių miškų klasifikatorius, tačiau keli kiek galima geresni modeliai buvo sudaryti pritaikant atraminių vektorių ir sprendimų medžio metodus. Kaip minėta, pats klasifikavimo tikslumas nebuvo itin aukštas ir indeksų kitimo prognozėms, kuriomis galima pasitikėti, sudaryti modeliai nelabai tinkami. Vis dėlto galima išskirti, kurie rodikliai buvo atrenkami dažniausiai kaip turintys reikšmingos įtakos Baltijos šalių akcijų rinkų indeksų kintamumui. Tai yra: *darbo užmokestis, gyventojų užimtumas, valdžios sektoriaus skola* (Lietuva); *darbo užmokestis, statybos sąnaudų elementų kainų indeksas* (Latvija); *pramonės produkcijos kainų indeksas, vieno ir trijų mėnesių EURIBOR palūkanų normos, bendrosios nacionalinės pajamos* (Estija).

Išvados

1. Baltijos šalyse veikiančių įmonių akcijų kainų kitimo prognozavimas viena diena į priekį ar šių valstybių akcijų rinkų indeksų kitimo prognozavimas vienu mėnesiu į priekį yra neištirta sfera, kadangi tokie tyrimai minimų šalių atveju nėra įgyvendinti.
2. Atlikta literatūros analizė atskleidė, jog labiausiai įmonių akcijų kainų ar akcijų rinkų indeksų pokyčius lemia tokie makroekonominiai rodikliai kaip bendrasis vidaus produktas, piniginių lėšos, valiutų kursas, palūkanų norma ir infliacija. Apžvelgus su Baltijos šalių akcijų rinkomis atliktus tyrimus pastebėta, jog tyrėjai dažniausiai nustato tokius joms įtaką darančius makroekonominis rodiklius: bendrasis vidaus produktas, pinigų kiekis, valdžios sektoriaus skola, nedarbo lygis, valiutų kursas, tiesioginės užsienio investicijos. Kai atliekamas akcijų kainų kitimo prognozavimas viena diena į priekį, aukštas klasifikavimo modelių tikslumas pasiekiamas tada, kai kintamieji parinkti iš įvairių duomenų šaltinių. Taip pat didelę reikšmę, sudarant prognozavimo modelius, turi duomenys, susiję su naujienomis.
3. Atlikta kitų autorių įgyvendintų akcijų kitimo prognozavimo tyrimų apžvalga parodė, jog dažnu atveju tiksliausiai klasifikuoja/prognozuoja atraminių vektorių metodas, taip pat tokiuose tyrimuose pritaikomi sprendimų medžio, neuroninių tinklų klasifikatoriai. Nustatyta, jog ankstesniuose darbuose, siekiant atrinkti reikšmingus kintamuosius, kai prognozuojamas akcijų kitimas, jautrumo analizės metodai nenaudoti.
4. Šiame darbe, prieš atliekant akcijų kainų prognozavimo viena diena į priekį tyrimą, ištirta, jog dažniausiai kaip reikšmingi kintamieji atrenkami tiek techniniai indikatoriai, tiek akcijų prekybos duomenys: aukščiausia, žemiausia, koreguota, atidarymo, uždarymo kaina, eksponentinis slankusis vidurkis, paprastasis slankusis vidurkis, stochastinio osciliatoriaus linija K, Bollinger'io indikatoriaus apatinės, viršutinės juostų reikšmės bei procentiniai skaičiavimai. Kito tyrimo metu dažniausiai atrinkti tokie kintamieji, galimai, turintys reikšmingos įtakos, prognozuojant Baltijos šalių akcijų rinkų indeksų kitimą: darbo užmokestis, gyventojų užimtumas, valdžios sektoriaus skola (Lietuva); darbo užmokestis, statybos sąnaudų elementų kainų indeksas (Latvija); pramonės produkcijos kainų indeksas, vieno ir trijų mėnesių palūkanų normos, bendrosios nacionalinės pajamos (Estija).
5. Ištirta, kad tada, kai prognozuotas įmonių akcijų kainų kitimas, Lietuvos ir Estijos atvejais reikšmingų kintamųjų atrinkimas neturėjo įtakos, o Latvijos įmonės atveju buvo naudingas dalinės koreliacijos metodas – klasifikuota tiksliau nei taikant visus duomenis. Gauti aukšti geriausių modelių tikslumo (AUC) įverčiai (0,836, 0,824 ir 0,805) parodė, jog panaudojant vien tik istorinius akcijų kainos duomenis galima gauti gerus akcijų kainų prognozavimo viena diena į priekį rezultatus.
6. Prognozuojant akcijų rinkų indeksų kitimą mėnesiu į priekį nustatyta, jog sudarytų geriausių modelių tikslumo reikšmės nebuvo aukštos nė vienos šalies atveju. Tačiau reikšmingų kintamųjų atrinkimas, pritaikant jautrumo analizės metodus, turėjo įtakos tam, jog klasifikavimo tikslumas pagerėdavo bent keliais procentais. Duomenų paslinkimas vienu ar dviem periodais (mėnesiais) į priekį (kadangi iškelta prielaida, jog makroekonominiai rodikliai gali ne iš karto veikti indeksų pokyčius) taip pat dažnai turėjo teigiamos reikšmės tam, kad buvo klasifikuojama tiksliau. Aukščiausi tikslumo (AUC) įverčiai gauti analizuojant Latvijos indeksą, o žemiausi – Estijos. Kadangi klasifikavimo kokybė šiuo atveju nebuvo itin gera, galima daryti išvadą, jog indeksų kitimo prognozėms, kuriomis galima pasitikėti, sudaryti modeliai nelabai tinkami,

tačiau tikslumas gali bent kiek pagerėti, jei kintamieji atrenkami ir/ar paslenkami tam tikrą skaičių periodų į priekį.

7. Atlikus akcijų kainų kitimo prognozavimą nustatyta, jog geriausiai klasifikavo atraminių vektorių klasifikatorius, o indeksų kitimo prognozavimo metu akivaizdžiai dominavo atsitiktinių miškų metodas, tačiau keli kiek galima geresni modeliai sudaryti pritaikant atraminių vektorių ir sprendimų medžio metodus.
8. Investuotojams, atliekantiems vienų ar kitų įmonių akcijų analizę bei sprendžiantiems klausimą dėl šių akcijų pirkimo, pardavimo, laikymo, rekomenduojama prognozavimo modelius sudaryti su tokiais akcijų prekybos ir techninių indikatorių duomenimis, kokie taikyti šio tyrimo metu, kadangi prognozavimo tikslumas buvo aukštas. Jautrumo analizės metodus taip pat rekomenduojama panaudoti, kadangi, kaip parodė rezultatai, tokiu būdu galima gauti tikslesnes prognozes. Taikytinas klasifikatorius – atraminių vektorių. Jei atliekamas akcijų rinkos indekso kitimo prognozavimas, rekomenduojama įvesti papildomų makroekonominių kintamųjų ir, žinoma, pritaikyti jautrumo analizės metodus, nes taip galima gauti tikslesnius rezultatus.

Literatūros sąrašas

1. BalaGuru R., Kumar L. V. (2015). A Study on Stock Market Quotations and Indices. *Conference Paper* [interaktyvus] [žiūrėta 2019-02-21]. Prieiga per ResearchGate: https://www.researchgate.net/publication/317828249_A_STUDY_ON_STOCK_MARKET_QUOTATIONS_AND_INDICES_SPECIAL_REFERNECE_WITH_BSE_SENSEX;
2. Faerber E. (2008). *All About Stocks. The Easy Way to Get Started*. Third Edition. USA: The McGraw-Hill Companies, doi: 10.1036/0071494553;
3. Capital Market Information Center (2012). *An introduction to Stock Market Indices* [interaktyvus] [žiūrėta 2019-02-21]. Prieiga per: <http://www.cmhc.sec.gov.lk/wp-content/uploads/2012/09/13.-AN-INTRODUCTION-TO-STOCK-MARKET-INDICES.pdf>;
4. Hofer E. (1999). Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. *Computer Physics Communications* [interaktyvus], Vol. 117, No. 1-2, p. 21-34 [žiūrėta 2019-02-23]. Prieiga per ScienceDirect: <https://www.sciencedirect.com/science/article/pii/S0010465598001532>;
5. Kopustinskas V., Alzbutas R., Augutis J. (2007). Matematinų modelių parametų jautrumo ir rezultatų neapibrėžtumo statistiniai tyrimo metodai. *Energetika* [interaktyvus], Vol. 53, No. 3, p. 10-15 [žiūrėta 2019-02-23]. Prieiga per ResearchGate: https://www.researchgate.net/publication/271852949_Matematiniu_modeliu_parametru_jautrumo_ir_rezultatuu_neapibreztumo_statistiniai_tyrimo_metodai;
6. Cacuci D. G. (2003). *Sensitivity and Uncertainty Analysis. Theory* [interaktyvus], Vol. 1, p. 144, USA: Chapman & Hall/CRC [žiūrėta 2019-02-23]. Prieiga per: http://inis.jinr.ru/sl/tot_ra/0/0/3/Cacuci-Sensitiv.pdf;
7. Antony C., Jeevanand E. S. (2007). The Elasticity of the Price of a Stock and its Beta. *Journal of Applied Quantitative Methods*, Vol. 2, No. 3, p. 334-342, ISSN: 1842-4562;
8. Graham B., Dodd D. L. (ed., 1976). *Security Analysis*. The Classic 1934 Edition. Donnelley R. R., Indiana, ISBN 0-07-024496-0;
9. Bagdonas R., Klimašauskas D. (2005). Vertybinių popierių kainai įtaką darantys veiksniai. *Finansai* [interaktyvus], p. 24-31 [žiūrėta 2019-05-16]. Prieiga per: http://www.elibrary.lt/resursai/DB/StatistikosDep/LEA/2005_02/lea05_2_07.pdf;
10. Chen N. F., Roll R., Ross S. A. (1986). Economic Forces and the Stock Market. *The Journal of Business* [interaktyvus], Vol. 59, No. 3, p. 383-403 [žiūrėta 2019-02-15]. Prieiga per JSTOR: https://www.jstor.org/stable/2352710?seq=1#page_scan_tab_contents ;
11. Bodurtha J. N., Cho D. C., Senbet L. W. (1989). Economic Forces and the Stock Market: An International Perspective. *The Global Finance Journal*, Vol. 1, No. 1, p. 21-46, ISSN: 1044-0283;
12. Boreika P., Pilinkus D. (2009). Makroekonominų rodiklių ir akcijų kainų tarpusavio ryšys Baltijos šalyse. *Ekonomika ir vadyba*, 14, p. 692-697, ISSN: 1822-6515;
13. John I. J., Duke II J. (2013). Macroeconomic Factors that Influence Stock Market Development in Nigeria. *International Journal of Business and Management Review* [interaktyvus], Vol. 1, No. 2, p. 43-56 [žiūrėta 2019-02-15]. Prieiga per ResearchGate: https://www.researchgate.net/publication/267694631_MACROECONOMIC_FACTORS_THAT_INFLUENCE_STOCK_MARKET_DEVELOPMENT_IN_NIGERIA;
14. Mariappan R., Hari N., Jyotishi A. (2013). Impact of Macroeconomic Variables on Stock Market Indices Value: A Cross-Country Indices Study. *Ushus J B Mgt*, Vol. 12, No. 3, p. 65-76, ISSN: 0975-3311;
15. Hunjra A. I., Chani M. I., Shahzad M., et al. (2014). The Impact of Macroeconomic Variables on Stock Prices in Pakistan. *International Journal of Economics and Empirical Research* [interaktyvus], Vol. 2, No. 1, p. 13-21 [žiūrėta 2019-02-15]. Prieiga per ResearchGate: https://www.researchgate.net/publication/263620295_The_Impact_of_Macroeconomic_Variables_on_Stock_Prices_in_Pakistan;
16. Ouma W. N., Muriu P. (2014). The Impact of Macroeconomic Variables on Stock Market Returns in Kenya. *International Journal of Business and Commerce*, Vol. 3, No. 11, p. 1-31, ISSN: 2225-2436;
17. Jareno F., Negrut L. (2016). US Stock Market and Macroeconomic Factors. *The Journal of Applied Business Research* [interaktyvus], Vol. 32, No. 1, p. 325-340 [žiūrėta 2019-02-17]. Prieiga per ResearchGate: https://www.researchgate.net/publication/282292396_US_Stock_Market_and_Macroeconomic_Factors;
18. Duy V. Q. (2016). The Impact of Macroeconomic Factors on Stock Price Index, VN-Index. *International Journal of Innovative Science*, Vol. 3, No. 7, p. 69-84, ISSN: 2348-7968;

19. Ligocka M., Pražak T., Stavarek D. (2016). The Effect of Macroeconomic Factors on Stock Prices of Swiss Real Estate Companies. *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis* [interaktyvus], Vol. 64, No. 6, p. 2015-2024 [žiūrėta 2019-02-15]. Prieiga per ACTA: <https://acta.mendelu.cz/64/6/2015/>;
20. Kabeer M. A. (2017). The Influence of Macroeconomic Factors on Stock Markets Performance in Top SAARC Countries and China. *Journal of Business and Financial Affairs*, Vol. 6, No. 1, p. 1-10, ISSN: 2167-0234;
21. Chauque D. F. F., Rayappan P. AP (2018). The Impact of Macroeconomic Variables on Stock Market Performance: A Case of Malaysia. *Edelweiss Applied Science and Technology*, Vol. 2, No. 1, p.100-104, ISSN: 2576-8484;
22. Inumula K. M., Tadararla A., Deeppa K. (2018). Impact of Macroeconomic Factors on Stock Price: Application of MLR. *International Journal of Management and Applied Science*, Vol. 4, No. 3, p. 75-78, ISSN: 2394-7926;
23. Ndlovu B., Faisal F., Resatoglu N. G., et al. (2018). The Impact of Macroeconomic Variables on Stock Returns: A Case of the Johannesburg Stock Exchange. *Romanian Statistical Review* [interaktyvus], No. 2, p. 87-104 [žiūrėta 2019-02-15]. Prieiga per RevistadeStatistica: http://www.revistadestatistica.ro/wp-content/uploads/2018/06/RRS-2_2018_A61.pdf;
24. Chen M. -H., Kim W. G., Kim H. J. (2005). The impact of macroeconomic and non-macroeconomic forces on hotel stock returns. *International Journal of Hospitality Management*, Vol. 24, p. 243-258, doi: 10.1016/j.ijhm.2004.06.008;
25. Zopiatis A., Savva Ch. S., Lambertides N., et al. (2017). *Tourism Stocks in Times of Crises: An Econometric Investigation of Unexpected Non-macroeconomic Factors* [interaktyvus] [žiūrėta 2019-02-18]. Prieiga per: <https://ideas.repec.org/p/ems/eureir/100332.html>;
26. Cheng M. -Ch., Tzeng Z. -Ch., Kang W. -L. (2011). The Impact of Non-Macroeconomic Events on Taiwan Electronic Industry Stock Index Returns, *Global Economy and Finance Journal* [interaktyvus], Vol. 4, No. 1, p. 80-101 [žiūrėta 2019-02-18]. Prieiga per ResearchGate: https://www.researchgate.net/publication/265077467_The_Impact_of_Non-Macroeconomic_Events_on_Taiwan_Electronic_Industry_Stock_Index_Returns;
27. Isidore R., Christie P. (2018). Fundamental Analysis Versus Technical Analysis – A Comparative Review. *International Journal of Recent Scientific Research*, Vol. 9, No. 1(B), p. 23009-23013, ISSN: 0976-3031;
28. Tigent Brands Inc. (2010). *Guide To Fundamental and Technical Analysis* [interaktyvus], p. 60-61 [žiūrėta 2019-02-18]. Prieiga per: <https://s3-ap-southeast-1.amazonaws.com/tv-prod/documents%2F1871-Fundamental%26Technical+AnalysisManual.pdf>;
29. Petrusheva N., Jordanoski I. (2016). Comparative Analysis Between the Fundamental and Technical Analysis of Stocks. *International Journal of Process Management – New Technologies* [interaktyvus], Vol. 4, No. 2, p. 26-31 [žiūrėta 2019-02-18]. Prieiga per ResearchGate: https://www.researchgate.net/publication/302057912_Comparative_analysis_between_the_fundamental_and_technical_analysis_of_stocks;
30. Young M. G. (2011). *The Complete Guide to Selling Stocks Short: Everything You Need to Know Explained Simply*. Florida: Atlantic Publishing Group, p. 121-134;
31. Bollinger J. (2002). *Bollinger on Bollinger Bands*. USA: McGraw-Hill, ISBN: 0-07-137368-3;
32. Murphy J. J. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. USA: New York Institute of Finance, ISBN: 0-7352-0066-1;
33. Weng B. (2017). *Application of machine learning techniques for stock market prediction*. PhD Dissertation [interaktyvus]. Alabama: Auburn University [žiūrėta 2019-03-08]. Prieiga per AUETD: <https://etd.auburn.edu/bitstream/handle/10415/5652/Application%20of%20machine%20learning%20techniques%20for%20stock%20market%20prediction.pdf?sequence=2>;
34. Kim. K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, Vol. 55, No. 1, p. 307–319, doi: 10.1016/S0925-2312(03)00372-2;
35. Grigoryan H. (2015). Stock Market Prediction using Artificial Neural Networks. Case Study of TALIT, Nasdaq OMX Baltic Stock. *Database Systems Journal* [interaktyvus], Vol. 6, No. 2, p. 14-23 [žiūrėta 2019-05-12]. Prieiga per ResearchGate: https://www.researchgate.net/publication/301790635_Stock_Market_Prediction_using_Artificial_Neural_Networks_Case_Study_of_TALIT_Nasdaq OMX_Baltic_Stock;

36. Maksvytienė I., Safonovas G. (2016). Ekonometrinių modelių pritaikymas OMXV indekso pokyčių prognozavimui. *Taikomoji ekonomika: sisteminiai tyrimai*, Vol. 10, No. 1, p. 187-203, doi: 10.7220/AESR.2335.8742.2016.10.1.10;
37. Marčišauskienė J., Cibulskienė D. (2013). Baltijos šalių makroekonominių rodiklių ir akcijų rinkos kainų tarpusavio ryšio vertinimas. *Ekonomika ir vadyba: aktualijos ir perspektyvos*, Vol. 29, No. 1, p. 51-61, ISSN: 1648-9098;
38. Hsing Y. (2014). Impacts of Macroeconomic Factors on the Stock Market in Estonia. *Journal of Economics and Development Studies*, Vol. 2, No. 2, p. 23-31, ISSN: 2334-2390;
39. Vaitkė E., Martinkutė-Kaulienė R. (2018). Investuotojų lūkesčių įtakos Lietuvos vertybinių popierių rinkos plėtrai tyrimas. *Mokslas – Lietuvos ateitis/ Science – Future of Lithuania*, Vol. 10, Article ID: mla.2018.420, p. 1-11, doi: 10.3846/mla.2018.420;
40. Pilinkus D. (2010). Macroeconomic Indicators and Their Impact on Stock Market Performance in The Short and Long Run: The Case of The Baltic States. *Technological and Economic Development of Economy*, Vol. 16, No. 2, p. 291-304, doi: 10.3846/tede.2010.19;
41. Oli. Sigo M., Selvam M., Maniam B., et al. (2018). Big Data Analytics-Application of Artificial Neural Network in Forecasting Stock Price Trends in India. *Academy of Accounting and Financial Studies Journal*, Vol. 22, No. 3, p. 1-13, ISBN:1528-2635-22-3-225;
42. Kristoufek L., Vosvrda M. (2014). Measuring capital market efficiency: long-term memory, fractal dimension and approximate entropy. *The European Physical Journal B*, Vol. 87, No. 162, p. 1-9, doi: 10.1140/epjb/e2014-50113-6;
43. Nasdaq Baltic (2019). *Indeksai* [interaktyvus] [žiūrėta 2019-05-13]. Prieiga per: <https://nasdaqbaltic.com/lt/indeksai/>;
44. Suthaharan Sh. (2016). *Machine Learning Models and Algorithms for Big Data Classification*. Thinking with Examples for Effective Learning. New York: Springer, p. 273, doi: 10.1007/978-1-4899-7641-3;
45. Breiman L. (2001). Random Forests. *Machine Learning* [interaktyvus], Vol. 45, No. 1, p. 5-32 [žiūrėta 2019-05-04]. Prieiga per SpringerLink: <https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf>;
46. Han J., Kamber M., Pei J. (2012). *Data Mining Concepts and Techniques*. Third Edition. USA: Elsevier, ISBN: 978-0-12-381479-1;
47. Aggarwal Ch. C. (2015). *Data Mining: The Textbook*. Switzerland: Springer, p. 293-297, doi: 10.1007/978-3-319-14142-8;
48. Kudyba S. (2014). *Big Data, Mining, and Analytics: Components of Strategic Decision Making*. Boca Raton: Taylor & Francis Group, p. 89-90, ISBN: 978-1-4665-6871-6;
49. Witten I. H., Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Second Edition. San Francisco: Elsevier, ISBN: 0-12-088407-0;
50. Guyon I., Weston J., Barnhill S., et al. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* [interaktyvus], Vol. 46, p. 389-422 [žiūrėta 2019-05-09]. Prieiga per SpringerLink: <https://link.springer.com/content/pdf/10.1023%2FA%3A1012487302797.pdf>;
51. Gogtay NJ., Thatte UM. (2017). Principles of Correlation Analysis. *Journal of The Association of Physicians of India* [interaktyvus], Vol. 65, p. 78-81 [žiūrėta 2019-05-09]. Prieiga per JAPI: http://www.japi.org/march_2017/12_sfr_principles_of_correlation.pdf;
52. Dean J. (2014). *Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners*. New Jersey: John Wiley & Sons, ISBN: 978-1-118-92069-5;
53. Brown B. L., Hendrix S. B. (2005). Partial Correlation Coefficients. *Encyclopedia of Statistics in Behavioral Science*, doi: 10.1002/9781118445112.stat06488;
54. Granger C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* [interaktyvus], Vol. 37, No. 3, p. 424-438 [žiūrėta 2019-05-14]. Prieiga per <http://tyigit.bilkent.edu.tr/metrics2/read/Investigating%20%20Causal%20Relations%20by%20Econometric%20Models%20and%20Cross-Spectral%20Methods.pdf>;
55. Al-Radaideh Q. A., Assaf A. A., Alnagi E. (2013). Predicting stock prices using data mining techniques. *The International Arab Conference on Information Technology (ACIT'2013)* [interaktyvus] [žiūrėta 2019-05-11]. Prieiga per ResearchGate: https://www.researchgate.net/publication/281865047_Predicting_Stock_Prices_Using_Data_Mining_Techniques;
56. Leung C. K. –S., MacKinnon R. K., Wang Y. (2014). A Machine Learning Approach for Stock Price Prediction. *18th International Database Engineering & Applications Symposium (IDEAS'14)*, p. 274-277, doi: 10.1145/2628194.2628211;

57. Lee Y., Ryu H., Lee H. (2017). Stock Prediction and Prediction Accuracy Improvement using Sentiment Analysis and Machine Learning based on Online News. *International Conference on Industrial Engineering and Operations Management* [interaktyvus], p. 1338-1349 [žiūrėta 2019-05-11]. Prieiga per: <http://ieomsociety.org/ieom2017/papers/384.pdf>;
58. Tharwat A. (2018). Classification assessment methods. *Applied Computing and Informatics*, p. 1-13, doi: 10.1016/j.aci.2018.08.003;
59. Lunardon N., Menardi G., Torelli N. (2014). ROSE: A Package for Binary Imbalanced Learning. *Contributed Research Articles*, Vol. 6, No. 1, p. 79-89, ISSN: 2073-4859.

Informacinių šaltinių sąrašas

1. *Estijos statistikos duomenų bazė*, <http://andmebaas.stat.ee/Index.aspx?lang=en>;
2. *Europos centrinis bankas*, <https://www.ecb.europa.eu/home/html/index.en.html>;
3. *Europos pinigų rinkų institutas*, <https://www.emmi-benchmarks.eu/>;
4. *Latvijos statistikos duomenų bazė*, <https://data1.csb.gov.lv/pxweb/en/?rxid=0e60689c-3f71-449a-a3a1-3abbf9ff78f1>;
5. *Lietuvos statistikos duomenų bazė*, <https://osp.stat.gov.lt/web/guest/pradinis>;
6. *Nasdaq Baltic*, <https://www.nasdaqbaltic.com/market/>;
7. *Pasaulio bankas*, <http://www.worldbank.org>;
8. *Tarptautinių atsiskaitymų bankas*, <https://www.bis.org>.

Priedai

1 priedas. Techninė analizė parentos tyrimo dalies duomenų apdorojimo kodas

```
### Ikeliami dieniniai duomenys, braizomi grafikai ###

# Grigeo - LT
# Grindeks - LV
# Tallinna Kaubamaja Grupp - EST

rm(list = ls())

if ("pacman" %in% rownames(installed.packages()) == FALSE)
install.packages("pacman")
require(pacman)
pacman::p_load(quantmod, zoo, plyr, TTR, caret, fscaret, kernlab)
pacman::p_load(gmodels, C50, shiny, dplyr, mlbench, cran, corrplot, RColorBrewer, rquery
)

## Imones duomenys

X <- read.csv2("X_imone_R.csv", header=T)

X$date<-strptime(X$date, format="%Y-%m-%d")
Xxts<-xts(X[, -1], order.by=X$date)

sum(is.na(Xxts))
head(Cl(Xxts))
X_plot <- Xxts$Y.VS.Close
barChart(X_plot)

kint<-Xxts$Y.VS.Close # uždarymo kaina
plot(kint)

# Generuojamos dienos gražos

X_returns <- diff(log(kint))[-1] # ln(price(t+1)/price(t))
chartSeries(X_returns)
head(X_returns)
plot.zoo(X_returns, ylab="Įmonė", xlab="", main = "Dienos gražos" )
summary(X_returns)

### Sudaromi techniniai rodikliai ###

## Naudojami uždarymo kainos duomenys

colnames(Xxts)

# Simple Moving Average (SMA)

sma20 <- SMA(Xxts$Y.VS.Close, n=20)

# Exponential Moving Average (EMA)

ema20 <- EMA(Xxts$Y.VS.Close, n=20)

# Bollinger Bands indicator

bb20 <- BBands(Xxts$Y.VS.Close, n=20, sd=2.0)
XxtsPlusBB <- data.frame(X, bb20)
XxtsBB<-xts(XxtsPlusBB[, -1], order.by=XxtsPlusBB$date)
```

```

XxtsBB%>%Ad()%>%chartSeries()
XxtsBB%>%chartSeries(TA='addBBBands();addVo()',subset='2018',theme =
chartTheme("white"))

# Relative Strength Indicator (RSI)

rsi14 <- RSI(Xxts$Y.VS.Close,n=14)

# The MACD function

macd <- MACD(Xxts$Y.VS.Close,nFast=12, nSlow=26, nSig=9, maType=SMA)
tail(macd, n=10)

# Stochastic Oscillator / Stochastic Momentum Index

so14 <-stoch(Xxts$Y.VS.Close,nFastK=14,nFastD=3,nSlowD=3,
bounded=TRUE,smooth=1)

# Sukuriami target stulpeliai

target_close <- diff(Xxts$Y.VS.Close)
colnames(target_close) <- "Target_Close"

target_close[target_close<=0] <- 0      # 0 reiskia nekyla
target_close[target_close>0] <- 1     # 1 reiskia kyla

Target_X <- data.frame(target_close) # sudaroma duomeniu matrica

# Sujungiami visi duomenys

allX <- data.frame(X,sma20,ema20,bb20,rsi14,macd,Target_X)
allX <- data.frame(X,so14)
Xxtsall <- xts(allX[,-1], order.by=allX$Date)

write.csv2(allX, file='X_tech_so.csv')

### Nustatomos isskirty. Ju turi buti ne daugiau 5 proc. ###

Xkint <- read.csv2("X_tech_su_kaina.csv", header=T)
ts<- ts(Xkint[c(-1,-2)])
scaled.dat <- scale(ts,center = TRUE, scale = TRUE)
boxplot.stats(scaled.dat)$out
boxplot(scaled.dat, main="Išskirčių nustatymas")

### Taikant jautrumo analizes metodus atrenkami kintamieji ###

# Rekursinis pozymiu pasalinimo metodas (RFE)

set.seed(7)

colnames(Xkint)
sum(is.na(Xkint))

control <- rfeControl(functions=rFuncs, method="cv", number=10)
results <- rfe(scaled.dat, Xkint[,2], sizes=c(1:17),rfeControl=control)
print(results)
predictors(results)
plot(results, type=c("g", "o"))

# Koreliacine analize

Xcorr <- Xkint
corr_X <-cor(scaled.dat, Xcorr$Close, method = "spearman")

```

```
print(corr_X)

# Regresine analize

Xkint <- Xkint[,-1]
allX <- data.frame(scaled.dat, Xkint$Close)

library(konfound)
library(arm)

X_reg <- lm(Xkint.Close ~ ., data = allX)

summary(X_reg)
colnames(Xkint)
konfound(X_reg, SVKI)

# Doline koreliacija

library(boot)
library(sensitivity)

X_part_corr <- pcc(Xkint, Xkint$Close)
print(X_part_corr)
```

2 priedas. Įmonių akcijų kainų kitimo prognozavimo/klasifikavimo kodas²

```
### Klasifikavimas ###

rm(list = ls())

## Reikiamu paketu instaliavimas

if ("pacman" %in% rownames(installed.packages()) == FALSE)
install.packages("pacman")
require(pacman)
pacman::p_load(DRR, doParallel, caret, e1071, randomForest, MASS, FNN, RSNNS,
tidyr, precrec, OptimalCutpoints)
pacman::p_load(neuralnet, nnet)
pacman::p_load(zoo, tidyverse, reshape2, dplyr, rpart.plot, rpart)
cl <- makeCluster(max(1,detectCores()-1))
registerDoParallel(cl)
pacman::p_load(devtools)
#install_github("davidavdav/ROC")
library(ROC)

## Duomenų paruošimas

factorsNumeric <- function(d) modifyList(d, lapply(d[, sapply(d, is.factor)],
as.numeric))

data <- read.csv2('X_class_visi.csv', header=T)
colY <- "Target_Close"
idxY <- which(colnames(data) %in% colY)
data[,idxY] <- factor(data[,idxY],labels=c("False","True"))

sum(is.na(data))
table(data$Target_Close)
prop.table(table(data$Target_Close))

nmd <- names(data)
formulaLong <- as.formula(paste(paste(colY," ~",sep=""), paste(nmd[!nmd %in%
colY], collapse = " + ")))
formulaShort <- as.formula(paste(paste(colY,".",sep="~"))

# Atliekamas duomenų standartizavimas

procValues <- preProcess(factorsNumeric(data[, -idxY]), method = c("center",
"scale"))

## Klasinių balansavimas

library(ROSE)

data.rose <- ROSE(Target_Close ~ ., data = data, seed = 2)$data
#data.rose[,idxY] <- factor(data.rose[,idxY],labels=c("True","False"))

table(data.rose$Target_Close)
head(data.rose)

procValues2 <- preProcess(factorsNumeric(data.rose[, -idxY]), method =
c("center", "scale"))
```

² Kai kurios kodo dalys pritaikytos iš studijų programos „Didžiųjų verslo duomenų analitika“ paskaitų medžiagos

```

## Modeliu su caret sudarymas

# Atsitiktiniu miskiu metodo (RF) parametrai

tuneGrid_rf <- expand.grid(.mtry=c(floor(sqrt(ncol(data.rose))))))
control <- trainControl(method="cv", number=5, classProbs=T,
summaryFunction=twoClassSummary, allowParallel = T)
modellist <- list()

# Artimiausiu kaimynu metodo (k-NN) parametrai

tuneGrid_knn <- expand.grid(k=1:99)

# Neuroniniu tinku metodo (nnet) parametrai

tuneGrid_nnet <- expand.grid(size=seq(from = 10, to = 30, by = 10),
decay = seq(from = 0.1, to = 0.5, by = 0.1))

# Sprendimu medzio metodo (DF) parametrai

tuneGrid_dt <- expand.grid(maxdepth = 2:10)

# Atraminu vektoriu metodo (SVM) parametrai

sigmas <- sigest(formulaShort, data.rose, frac = 1, scaled = TRUE)
tuneGrid_svm <- expand.grid(sigma = round(seq(sigmas[1], sigmas[3], len=3),4), C
= 2^c(-3:2))

## Modeliu sudarymas

set.seed(7)
k <- 5 # isorinio kryzminio patikrinimo daliu skaicius
myFolds <- createFolds(data[,colY],k)
myResults <- NULL
for (i in 1:k) {

  tstInd <- myFolds[[i]]
  trnIdx <- as.logical(rep(1,1,nrow(data.rose)))
  trnIdx[tstInd] <- FALSE
  trnInd <- which(trnIdx)
  target <- as.logical(data[tstInd,idxY])

  trnDataProc <- predict(procValues2, factorsNumeric(data.rose[trnInd,-idxY]))
  tstDataProc <- predict(procValues, factorsNumeric(data[tstInd,-idxY]))

  cat(sprintf("\nCV fold %d out of %d / Random Forest\n", i, k))
  for(ntree in c(1000, 1500, 2000, 2500, 3000)){
    set.seed(7)
    rf_caret <- caret::train(formulaShort, data =
cbind(trnDataProc,Target_Close=data.rose[trnInd,idxY]),
method="rf", metric="ROC", tuneGrid = tuneGrid_rf,
trControl=control, ntree=ntree)
    key <- toString(ntree)
    modellist[[key]] <- rf_caret
  }
  results <- resamples(modellist)
  summary(results)
  print(rf_caret)
  model <- rep("RF",length(target))
  soft <- predict(rf_caret,data[tstInd,-idxY],type="prob")
  score <- soft[,2]
  myResults <- rbind(myResults,data.frame(tstInd,model,score,target))
  rm(rf_caret)
}

```

```

    cat(sprintf("\nCV fold %d out of %d / k-Nearest Neighbors\n", i, k))
    set.seed(7)
    knn_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target_Close=data.rose[trnInd,idxY]),
        method="knn", metric="ROC", tuneGrid = tuneGrid_knn,
trControl=control)
    print(knn_caret)
    model <- rep("kNN", length(target))
    soft <- predict(knn_caret, data[tstInd, -idxY], type="prob")
    score <- soft[,2]
    myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
    rm(knn_caret)

    cat(sprintf("\nCV fold %d out of %d / Neural Network\n", i, k))
    set.seed(7)
    nnet_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target_Close=data.rose[trnInd,idxY]),
        method="nnet", metric="ROC", tuneGrid =
tuneGrid_nnet, trControl=control)
    print(nnet_caret)
    model <- rep("nnet", length(target))
    score <- predict(nnet_caret, tstDataProc, type="prob")
    score <- soft[,2]
    myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
    rm(nnet_caret)

    cat(sprintf("\nCV fold %d out of %d / Decision tree\n", i, k))
    set.seed(7)
    dt_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target_Close=data.rose[trnInd,idxY]),
        method="rpart2", metric="ROC", tuneGrid =
tuneGrid_dt, trControl=control)
    print(dt_caret)
    model <- rep("DT", length(target))
    soft <- predict(dt_caret, tstDataProc, type="prob")
    score <- soft[,2]
    myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
    rm(dt_caret)

    cat(sprintf("\nCV fold %d out of %d / SVM with RBF kernel\n", i, k))
    set.seed(7)
    svm_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target_Close=data.rose[trnInd,idxY]),
        method = "svmRadialSigma", metric="ROC", tuneGrid =
tuneGrid_svm, trControl = control)
    print(svm_caret)
    model <- rep("SVM", length(target))
    soft <- predict(svm_caret, tstDataProc, type="prob")
    score <- soft[,2]
    myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
    rm(svm_caret)
}

myModels <- levels(myResults[, "model"])
myScores <- spread(myResults, model, score)

## Tikslumo tikrinimas

# Sumaisymu matricos sudarymas

myF <- NULL

```



```

for (i in 1:length(myModels)) {
  #opt.cut.result <- optimal.cutpoints(X = myModels[i], status = "target",
tag.healthy = 0, methods = "SpEqualSe", data = myScores, trace = F)
  #threshold <- opt.cut.result$SpEqualSe$Global$optimal.cutoff$cutoff
  threshold <- 0.5
  confusionMatrix <-
caret::confusionMatrix(as.factor(myScores[,myModels[i]]>=threshold),as.factor(my
Scores$target),positive="TRUE",mode="everything")
  cat(paste0(myModels[i],'\n'))
  print(confusionMatrix)
  myF <- c(myF,as.numeric(confusionMatrix$byClass['F1']))
}

# ROC kreives radimas

myModelNames <- NULL
i <- 1
performance <-
roc.plot(myResults[myResults["model"]==myModels[i],],i,traditional=TRUE)
myModelNames[i] <- sprintf('%s AUC=%5.3f',myModels[i],1-performance['pAUC'])
for (i in 2:length(myModels)) {
  performance <-
roc.plot(myResults[myResults["model"]==myModels[i],],i,traditional=TRUE)
  myModelNames[i] <- sprintf('%s AUC=%5.3f',myModels[i],1-performance['pAUC'])
}
legend("bottomright",box.lty=0,cex=0.7,ncol=1,pt.cex=3,text.width=0.25,myModelNa
mes,lty=rep(1,1,length(myModels)),col=1:length(myModels))

```

3 priedas. Fundamentaliąja analize paremtos tyrimo dalies duomenų apjungimo kodas

```
## Duomenys is dieniniu apjungiami i menesinius

rm(list = ls())
getwd()
#setwd()

library(quantmod)
library(zoo)
library(plyr)

indexBC <- read.csv2('X.csv', header=T)
rownames(indexBC) <- indexBC$date
indexBC$date <- NULL

sum(is.na(indexBC))
head(indexBC)
str(indexBC)
summary(indexBC)

#LT<- subset(indexBC, select="OMXV.Value")
#LV<- subset(indexBC, select="OMXR.Value")
#EE<- subset(indexBC, select="OMXT.Value")

#plot(as.zoo(indexBC), screens = 1, lty = 1:3, xlab = "Data", ylab = "Kaina",
col = as.list(1:3))
#legend("bottomright", c("OMXT", "OMXR", "OMXV"), cex = 0.5, fill = 1:3, ncol =
2)

# Apjungiami pagal menesi

indexBC_monthly <- apply.monthly(indexBC,FUN=colMeans)

#plot(as.zoo(indexBC_monthly), screens = 1, lty = 1:3, ylab="", xlab="", main =
"Baltijos šalių indeksai", col = as.list(1:3))
#legend("bottomright", c("OMXT", "OMXR", "OMXV"), cex = 0.5, fill = 1:3, ncol =
2)

summary(indexBC_monthly)
sum(is.na(indexBC_monthly))

write.csv2(indexBC_monthly, file='indeksu_menesiniai_duomenys.csv')
```

4 priedas. Fundamentaliąja analize paremtos tyrimo dalies duomenų apdorojimo kodas

```
### Ikeliami menesiniai duomenys ###

rm(list = ls())
#getwd()

## Reikiamu paketu instaliavimas

if ("pacman" %in% rownames(installed.packages()) == FALSE)
install.packages("pacman")
require(pacman)
pacman::p_load(quantmod, zoo, plyr, TTR, caret, fscaret, kernlab)
pacman::p_load(gmodels, C50, shiny, dplyr, mlbench, cran, corrplot, RColorBrewer, rquery
)

## Salies akciju rinkos indeksas

X <- read.csv2("X_makro_data.csv", header=T)

X$DATE<-strptime(X$DATE, format="%Y-%m-%d")
Xxts<-xts(X[,-1], order.by=X$DATE)

sum(is.na(Xxts))
head(Xxts)
colnames(Xxts)

# Generuojamas tikslo stulpelis

Target <- diff(Xxts$OMX)
colnames(Target) <- "Target"

Target[Target<=0] <- 0      # 0 reiskia nekyla
Target[Target>0] <- 1     # 1 reiskia kyla

Target_LT <- data.frame(Target)

# Sujungiami visi duomenys

AllX <- data.frame(X, Target_X)
Xxtsall <- xts(allX[,-1], order.by=allX$DATE)

write.csv2(allX, file='X_macro.csv')

### Nustatomos isskirtys. Ju turi buti ne daugiau 5 proc. ###

Xkint <- read.csv2("X_makro_data_su_kaina_0.csv", header=T)
ts<- ts(Xkint[c(-1,-2)])
scaled.dat <- scale(ts, center = TRUE, scale = TRUE)
boxplot.stats(scaled.dat)$out
boxplot(scaled.dat, main="Išskirčių nustatymas")

### Taikant jautrumo analizes metodus atrenkami kintamieji ###

# Rekursinis pozymiu pasalinimo metodas (RFE)

set.seed(7)

colnames(scaled.dat)
sum(is.na(scaled.dat))

control <- rfeControl(functions=rFuncs, method="cv", number=10)
results <- rfe(scaled.dat, Xkint[,2], sizes=c(1:19), rfeControl=control)
print(results)
```

```

predictors(results)
plot(results, type=c("g", "o"))

# Koreliacine analize

Xcorr <- Xkint
corr_X <- cor(scaled.dat, Xcorr$OMX, method = "spearman")
print(corr_X)

# Regresine analize

Xkint <- Xkint[,-1]
allX <- data.frame(scaled.dat, Xkint$OMX)

library(konfound)
library(arm)

X_reg <- lm(Xkint.OMX ~ ., data = allX)

summary(X_reg)
colnames(Xkint)
konfound(X_reg, SVKI)

# Daline koreliacija

library(boot)
library(sensitivity)

X_part_corr <- pcc(Xkint, Xkint$OMX)
print(X_part_corr)

```

5 priedas. Akcijų rinkų indeksų kitimo prognozavimo/klasifikavimo kodas³

```
### Klasifikavimas ###

rm(list = ls())

## Reikiamu paketu instaliavimas

if ("pacman" %in% rownames(installed.packages()) == FALSE)
install.packages("pacman")
require(pacman)
pacman::p_load(DRR, doParallel, caret, e1071, randomForest, MASS, FNN, RSNNs,
tidyr, precrec, OptimalCutpoints)
pacman::p_load(neuralnet, nnet)
pacman::p_load(zoo, tidyverse, reshape2, dplyr, rpart.plot, rpart)
cl <- makeCluster(max(1, detectCores()-1))
registerDoParallel(cl)
pacman::p_load(devtools)
install_github("davidavdav/ROC")
library(ROC)

## Duomenų paruosimas

factorsNumeric <- function(d) modifyList(d, lapply(d[, sapply(d, is.factor)],
as.numeric))

data <- read.csv2('X_macro_0_visi.csv', header=T)
colY <- "Target"
idxY <- which(colnames(data) %in% colY)
data[,idxY] <- factor(data[,idxY], labels=c("False", "True"))

sum(is.na(data))
table(data$Target)
prop.table(table(data$Target))

nmd <- names(data)
formulaLong <- as.formula(paste(paste(colY, " ~", sep=""), paste(nmd[!nmd %in%
colY], collapse = " + ")))
formulaShort <- as.formula(paste(paste(colY, ".", sep="~"))

# Duomenų standartizavimas

procValues <- preprocess(factorsNumeric(data[, -idxY]), method = c("center",
"scale"))

## Klasijų balansavimas

#install.packages("ROSE")
library(ROSE)

data.rose <- ROSE(Target ~ ., data = data, seed = 2)$data
#data.rose[,idxY] <- factor(data.rose[,idxY], labels=c("True", "False"))

table(data.rose$Target)

procValues2 <- preprocess(factorsNumeric(data.rose[, -idxY]), method =
c("center", "scale"))
```

³ Kai kurios kodo dalys pritaikytos iš studijų programos „Didžiųjų verslo duomenų analitika“ paskaitų medžiagos

```

## Modeliu su caret sudarymas

#names(getModelInfo())
#modelLookup("xgbTree")

# Atsitiktiniu miskiu metodo (RF) parametrai
tuneGrid_rf <- expand.grid(.mtry=c(floor(sqrt(ncol(data.rose))))))
control <- trainControl(method="cv", number=3, classProbs=T,
summaryFunction=twoClassSummary, allowParallel = T)
modellist <- list()

# Artimiausiu kaimynu metodo (k-NN) parametrai
tuneGrid_knn <- expand.grid(k=1:99)

# Neuroniniu tinku metodo (nnet) parametrai
tuneGrid_nnet <- expand.grid(size=seq(from = 10, to = 30, by = 10),
                             decay = seq(from = 0.1, to = 0.5, by = 0.1))

# Sprendimu medzio metodo (DF) parametrai
tuneGrid_dt <- expand.grid(maxdepth = 2:10)

# Atraminiu vektoriu metodo (SVM) parametrai
sigmas <- sigest(formulaShort, data.rose, frac = 1, scaled = TRUE)
tuneGrid_svm <- expand.grid(sigma = round(seq(sigmas[1], sigmas[3], len=3),4), C
= 2^c(-3:2))

# Modeliu sudarymas

set.seed(7)
k <- 3 # isorinio kryzminio patikrinimo daliu skaicius
myFolds <- createFolds(data[,colY],k)
myResults <- NULL
for (i in 1:k) {

  tstInd <- myFolds[[i]]
  trnIdx <- as.logical(rep(1,1,nrow(data.rose)))
  trnIdx[tstInd] <- FALSE
  trnInd <- which(trnIdx)
  target <- as.logical(data[tstInd,idxY])

  trnDataProc <- predict(procValues2, factorsNumeric(data.rose[trnInd,-idxY]))
  tstDataProc <- predict(procValues, factorsNumeric(data[tstInd,-idxY]))

  cat(sprintf("\nCV fold %d out of %d / Random Forest\n", i, k))
  for(ntree in c(1000, 1500, 2000, 2500, 3000)){
    set.seed(7)
    rf_caret <- caret::train(formulaShort, data =
cbind(trnDataProc,Target=data.rose[trnInd,idxY]),
                             method="rf", metric="ROC", tuneGrid = tuneGrid_rf,
trControl=control, ntree=ntree)
    key <- toString(ntree)
    modellist[[key]] <- rf_caret
  }
  results <- resamples(modellist)
  summary(results)
  print(rf_caret)
  model <- rep("RF",length(target))
  soft <- predict(rf_caret,data[tstInd,-idxY],type="prob")
  score <- soft[,2]
  myResults <- rbind(myResults,data.frame(tstInd,model,score,target))
}

```

```

rm(rf_caret)

cat(sprintf("\nCV fold %d out of %d / k-Nearest Neighbors\n", i, k))
set.seed(7)
knn_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target=data.rose[trnInd,idxY]),
      method="knn", metric="ROC", tuneGrid = tuneGrid knn,
trControl=control)
print(knn_caret)
model <- rep("kNN", length(target))
soft <- predict(knn_caret, data[tstInd, -idxY], type="prob")
score <- soft[,2]
myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
rm(knn_caret)

cat(sprintf("\nCV fold %d out of %d / Neural Network\n", i, k))
set.seed(7)
nnet_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target=data.rose[trnInd,idxY]),
      method="nnet", metric="ROC", tuneGrid =
tuneGrid_nnet, trControl=control)
print(nnet_caret)
model <- rep("nnet", length(target))
score <- predict(nnet_caret, tstDataProc, type="prob")
score <- soft[,2]
myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
rm(nnet_caret)

cat(sprintf("\nCV fold %d out of %d / Decision tree\n", i, k))
set.seed(7)
dt_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target=data.rose[trnInd,idxY]),
      method="rpart2", metric="ROC", tuneGrid =
tuneGrid_dt, trControl=control)
print(dt_caret)
model <- rep("DT", length(target))
soft <- predict(dt_caret, tstDataProc, type="prob")
score <- soft[,2]
myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
rm(dt_caret)

cat(sprintf("\nCV fold %d out of %d / SVM with RBF kernel\n", i, k))
set.seed(7)
svm_caret <- caret::train(formulaShort, data =
cbind(trnDataProc, Target=data.rose[trnInd,idxY]),
      method = "svmRadialSigma", metric="ROC", tuneGrid =
tuneGrid_svm, trControl = control)
print(svm_caret)
model <- rep("SVM", length(target))
soft <- predict(svm_caret, tstDataProc, type="prob")
score <- soft[,2]
myResults <- rbind(myResults, data.frame(tstInd, model, score, target))
rm(svm_caret)
}

myModels <- levels(myResults[, "model"])
myScores <- spread(myResults, model, score)

## Tikslumo tikrinimas

# Sumaisymu matricos sudarymas

myF <- NULL

```

```

for (i in 1:length(myModels)) {
  #opt.cut.result <- optimal.cutpoints(X = myModels[i], status = "target",
tag.healthy = 0, methods = "SpEqualSe", data = myScores, trace = F)
  #threshold <- opt.cut.result$SpEqualSe$Global$optimal.cutoff$cutoff
  threshold <- 0.5
  confusionMatrix <-
caret::confusionMatrix(as.factor(myScores[,myModels[i]]>=threshold),as.factor(my
Scores$target),positive="TRUE",mode="everything")
  cat(paste0(myModels[i],'\n'))
  print(confusionMatrix)
  myF <- c(myF,as.numeric(confusionMatrix$byClass['F1']))
}

# ROC kreives radimas

myModelNames <- NULL
i <- 1
performance <-
roc.plot(myResults[myResults["model"]==myModels[i],],i,traditional=TRUE)
myModelNames[i] <- sprintf('%s AUC=%5.3f',myModels[i],1-performance['pAUC'])
for (i in 2:length(myModels)) {
  performance <-
roc.plot(myResults[myResults["model"]==myModels[i],],i,traditional=TRUE)
  myModelNames[i] <- sprintf('%s AUC=%5.3f',myModels[i],1-performance['pAUC'])
}
legend("bottomright",box.lty=0,cex=0.7,ncol=1,pt.cex=3,text.width=0.25,myModelNa
mes,lty=rep(1,1,length(myModels)),col=1:length(myModels))

```