



**Kauno technologijos universitetas**

Elektros ir elektronikos fakultetas

# **Investicinio portfelio formavimo strategijos sukūrimas ir tyrimas**

Baigiamasis magistro projektas

---

**Robertas Jonaitis**

Projekto autorius

**doc. dr. Vygandas Vaitkus**

Vadovas

---

**Kaunas, 2019**



**Kauno technologijos universitetas**

Elektros ir elektronikos fakultetas

# **Investicinio portfelio formavimo strategijos sukūrimas ir tyrimas**

Baigiamasis magistro projektas

Valdymo technologijos (6211EX014)

---

**Robertas Jonaitis**

Projekto autorius

**doc. dr. Vygandas Vaitkus**

Vadovas

**prof. dr. Rimvydas Simutis**

Recenzentas

---

**Kaunas, 2019**



**Kauno technologijos universitetas**

Elektros ir elektronikos fakultetas

Robertas Jonaitis

## **Investicinio portfelio formavimo strategijos sukūrimas ir tyrimas**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Roberto Jonaičio, baigiamasis projektas tema „Investicinio portfelio formavimo strategijos sukūrimas ir tyrimas“ 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)

Jonaitis, Robertas. Investicinio portfelio formavimo strategijos sukūrimas ir tyrimas. Magistro baigiamasis projektas / vadovas doc. dr. Vygandas Vaitkus; Kauno technologijos universitetas, Elektros ir elektronikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): elektronikos inžinerija, inžinerijos mokslai

Reikšminiai žodžiai: investicinis portfelis, akcijų kainos prognozavimas, kolektyviniai metodai.

Kaunas, 2019. 43 p.

### **Santrauka**

Pelningo investicinio portfelio sudarymas reikalauja daug žinių ir rinkos analizės. Šį procesą gali palengvinti mašininio mokymosi algoritmai ir optimizavimo metodai.

Šiame darbe aprašomi fundamentalios analizės rodikliai, portfelio sudarymo ir vertinimo įrankiai. Išsamiau supažindinama su akcijų atrankos metodais. Naudojant atsitiktinio miško ir *Gradient Boosting* regresorius pagal įmonių fundamentalius duomenis prognozuojami ateinančio ketvirčio kainų pokyčiai. Iš potencialiai pelningų akcijų pagal įvairias strategijas formuojami investiciniai portfeliai. Akcijų svoriai portfelyje nustatomi pagal istorinę efektyvaus investicijų pasiskirstymo kreivę. Įvairūs metodai išbandomi dviejų metų laikotarpiu, kas ketvirtį perrenkant portfelius. Sudaryti portfeliai įvertinami pasitelkiant lyginamąjį indeksą ir įvairius kokybės kriterijus.

Jonaitis, Robertas. Development and research of investment portfolio selection strategy. Master's Final Degree Project / supervisor assoc. prof. dr. Vygandas Vaitkus; Faculty of Electrical and Electronics Engineering, Kaunas University of Technology.

Study field and area (study field group): electronics engineering, engineering sciences

Keywords: investment portfolio, stock market prediction, ensemble methods.

Kaunas, 2019. 43 p.

### **Summary**

Profitable investment portfolio development requires a lot of knowledge and market analysis. Machine learning algorithms and optimization methods can facilitate this process.

The paper describes financial ratios, portfolio development and evaluation tools. Approaches to asset selection are discussed in greater detail. Random Forest and Gradient Boosting regressors with the fundamental data are used to predict the changes of stock prices for the next quarter. Based on selected stocks, investment portfolios are developed utilizing various strategies. Portfolio weights are calculated using historical efficient frontier. Different strategies are tested in a time horizon of two years, revising portfolio in every quarter. The performance of portfolios is determined by the chosen benchmark and different quality metrics.

## Turinys

<b>Paveikslų sąrašas</b> .....	<b>7</b>
<b>Įvadas</b> .....	<b>8</b>
<b>1. Investicinio portfelio formavimo etapai</b> .....	<b>9</b>
1.1. Investicijų atranka .....	9
1.2. Portfelio valdymas .....	11
1.2.1. Portfelio kokybės vertinimas .....	13
1.3. Tinkamas laikas investuoti .....	14
<b>2. Akcijų atrankos metodai</b> .....	<b>16</b>
2.1. Netiesinis rangavimas .....	16
2.2. Neraiškioji logika .....	16
2.3. Rekurentiniai neuroniniai tinklai .....	17
2.4. Evoliucinis algoritmas .....	18
2.5. Kolektyviniai metodai .....	19
2.5.1. Atsitiktinis miškas .....	20
2.5.2. <i>Gradient Boosting</i> .....	21
<b>3. Metodologija</b> .....	<b>22</b>
3.1. Duomenys .....	22
3.1.1. Duomenų bazės apdorojimas .....	22
3.1.2. Naudojamų duomenų imtis ir prognozavimo objektas .....	25
3.2. Primityvi prognozė .....	26
3.3. Kryžminis patikrinimas .....	27
3.4. Modelio parametrai .....	27
3.5. Prognozių tikslumo rodikliai .....	28
3.6. Modelio apmokymas ir testavimas .....	29
3.7. Portfelio sudarymas .....	30
3.8. Portfelijų vertės rodikliai .....	32
<b>4. Rezultatai</b> .....	<b>33</b>
4.1. Primityvi prognozė .....	33
4.2. Lyginamasis indeksas .....	33
4.3. Atsitiktinio miško metodas .....	34
4.3.1. Modelio vertinimas .....	34
4.3.2. Portfelijų vertinimas .....	35
4.4. <i>Gradient Boosting</i> metodas .....	37
4.4.1. Modelio vertinimas .....	37
4.4.2. Portfelijų vertinimas .....	38
4.5. Rezultatų aptarimas .....	40
<b>Išvados</b> .....	<b>41</b>
<b>Literatūros sąrašas</b> .....	<b>42</b>

## Paveikslų sąrašas

<b>1 pav.</b> Grafikas parodo, kaip teoriškai gali būti derinamos dvi investicijos siekiant gauti didžiausią įmanomą grąžą esant priimtinaam rizikos lygiui [4].....	12
<b>2 pav.</b> Didžiausio vertės smukimo absoliutinė vertė.....	14
<b>3 pav.</b> Išskleistas rekurentinis neuroninis tinklas [13] .....	17
<b>4 pav.</b> Genetinio algoritmo kūrimo etapai.....	18
<b>5 pav.</b> Skirtumas tarp <i>bagging</i> ir <i>boosting</i> modelių sudarymo [17].....	19
<b>6 pav.</b> Principinė atsitiktinio miško modelio schema [18].....	20
<b>7 pav.</b> Svetainėje simfin.com patalpintas duomenų failas turi 1 stulpelį ir 16183 eilučių.....	23
<b>8 pav.</b> Apdoroti duomenys .....	25
<b>9 pav.</b> Kiekvieno fundamentalaus rodiklio reikšmingumo dalis atsitiktinio miško mokymuisi.....	26
<b>10 pav.</b> Penkių dalių kryžminio patikrinimo išskaidymas .....	27
<b>11 pav.</b> Modelio tikslumo priklausomybė nuo medžių skaičiaus .....	28
<b>12 pav.</b> Modelio apmokymo ir testavimo etapai .....	29
<b>13 pav.</b> 20000 įvairių portfelių 2016 metų grąžos ir standartinio nuokrypio santykiai .....	31
<b>14 pav.</b> 12 porfelių sudarymo strategijų.....	31
<b>15 pav.</b> Lyginamojo indekso <i>S&amp;P 500</i> kainos grafikas .....	33
<b>16 pav.</b> Akcijų kainos pokyčių krypties prognozės tikslumas tarp skirtingų sektorių, naudojant atsitiktinio miško modelį .....	34
<b>17 pav.</b> Portfelių, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (16 perspektyviausių), kainų grafikai.....	35
<b>18 pav.</b> Portfelių, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (po 2 iš sektoriaus), kainų grafikai.....	36
<b>19 pav.</b> Akcijų kainos pokyčių krypties prognozės tikslumas tarp skirtingų sektorių, naudojant <i>Gradient Boosting</i> modelį .....	37
<b>20 pav.</b> Portfelių, sudarytų iš <i>Gradient Boosting</i> modelio atrinktų akcijų (16 perspektyviausių), kainų grafikai.....	38
<b>21 pav.</b> Portfelių, sudarytų iš <i>Gradient Boosting</i> modelio atrinktų akcijų (po 2 iš sektoriaus), kainų grafikai.....	39

## Įvadas

Investavimas į vertybinius popierius pasaulyje tampa vis labiau prieinamas. Įmonių akcijos (nuosavybės vertybiniai popieriai) išpopuliarėjo tarp investuotojų, kurie yra pasirengę priimti didesnę riziką ir už tai tikėtis didesnės gražos. 2018 metų pradžioje vien JAV akcijų rinka įvertinta 30 trilijonų JAV dolerių [1].

Kiekvienas investuotojas turi savo strategiją ir požiūrį į atskirus procesus bei įvykius, taigi turėdami vienodą informaciją elgiasi skirtingai. Todėl didėjantis investuotojų skaičius į akcijų rinką įneša vis daugiau nepastovumo. Taip pat informacinė erdvė yra užpildyta naujienomis bei pranešimais apie įmones, jų ekonominę būklę. O ir pačių vertybinių popierių, į kuriuos būtų galima investuoti, pasirinkimas nemažėja.

Visos šios problemos iškyla, kai investuotojas nusprendžia sudaryti investicinį portfelį ir renkasi vertybinius popierius bei jų proporcijas portfelyje. Savarankiškai atrinkti akcijas tampa per sudėtinga užduotis, todėl tarp investuotojų populiarėja įvairios praeityje pasiteisinusios strategijos, investiciniai modeliai ir prekyboje taikomi algoritmai. Šios priemonės ne tik tiksliau prognozuoja akcijos kainos kitimą, bet ir sutaupo investuotojų laiką.

Tiksliai prognozuoti ateities kainą yra beveik neįmanoma, tačiau tai nėra būtina pelnui uždirbti. Užtenka atspėti akcijų kryptį šešis kartus iš dešimties ir tokia strategija jau bus pelninga. Algoritmui svarbu atrasti pagrindinius ryšius tarp turimų duomenų ir prognozuojamos kainos bei išvengti triukšmų, kurių yra pilna akcijų rinkoje. Tam svarbu turėti tokius duomenis, kurie labiausiai atspindi kainą ir jos galimus pokyčius. Akcinių bendrovių finansiniai rodikliai dažniausiai yra tas veiksnys, į kurį labiausiai atsižvelgia investuotojai analizuodami įmones.

Darbo tikslas – sudaryti investicinį portfelį pasitelkiant fundamentalius duomenis bei kolektyvinius metodus ir įvertinti sudaryto portfelio kokybę.

Darbo uždaviniai:

1. apžvelgti investicinio portfelio, sudaryto vien iš įmonių akcijų, formavimo etapus;
2. apžvelgti akcijų atrankos metodus;
3. apdoroti fundamentalius įmonių duomenis;
4. sudaryti modelius, naudojantis kolektyviniais metodais;
5. iš atrinktų akcijų sudaryti investicinį portfelį;
6. įvertinti sudarytus modelius ir investicinius portfelius.



## 1. Investicinio portfelio formavimo etapai

Investavimas paprastai susideda iš trijų pagrindinių etapų [2].

1. investicijų atranka, kurioje pasirenkamos potencialiai pelningiausios investicijos iš įvairių turto klasių.
2. portfelio valdymas, kuris padeda rasti mažai rizikingo, bet pelningo portfelio dedamųjų proporcijas.
3. tinkamas laikas investuoti, kuriuo siekiama pirkti akcijas tuomet, kai šios yra pigiausios, ir parduoti, kai yra brangiausios.

Toliau išsamiau aprašytas kiekvienas etapas.

### 1.1. Investicijų atranka

Pradedant investuoti vienas pirmųjų uždavinių yra pasirinkti investavimo objektus. Investicijų pasirinkimas gal būti įvairus, pradedant nuo investavimo krypties (vertybiniai popieriai, nekilnojamas turtas, sutelktinis finansavimas ar kita) ir baigiant konkrečios šalies obligacijomis. Investicinio portfelio dedamosios yra vienas pagrindinių pelningumą lemiančių veiksnių.

Šiais laikais beveik nė vienas investicinis portfelis neapsieina be bendrovių akcijų [3]. Akcijų rinka yra žinoma kaip rizikingas, tačiau pelningas investavimo objektas. Per laikotarpį nuo 1900 m. iki 2009 m. pasaulio akcijų rinkų vidutinė reali grąža (akcijų kainų augimas ir reinvestuoti dividendai) siekė 5,2 %. Tačiau atskirų biržų grąža buvo labai nevienoda – pavyzdžiui, Švedijoje vidutinis metinis rezultatas buvo 7,2 %, o Italijoje šis rodiklis tesiekė 1,9 % [4].

Ilguoju laikotarpiu akcijų vertė priklauso nuo įmonių finansinės būklės ir investuotojų lūkesčių dėl ateities pelno. Nors pastarojo įtaką sunku nuspėti, įmonių finansinę padėtį, o taip pat ir galimą augimą, dažnai nusako jos finansiniai rodikliai. Akcinė bendrovė, turinti mažai skolų ir pakankamai grynųjų pinigų, paprastai turi gerus finansinius duomenis. Įmonės finansinės būklės vertinimas, pasitelkiant finansinius rodiklius, vadinamas fundamentaliąja analize. Į ją įeina ir rinkos bei pramonės šakų (sektorių) analizės, tačiau šiame darbe investicijų atrankai pasitelkiama tik įmonių finansinių rodiklių analizė [4, 5].

Toliau išvardinti finansiniai rodikliai, kurie yra vieni populiariausių fundamentalios analizės įrankių.

**Rinkos kapitalizacija** (angl. *Market Capitalization*) – rodiklis, iš dalies parodantis, kiek verta įmonė. Šis dydis yra lygus įmonės akcijų kiekiui padaugintam iš akcijos rinkos kainos.

$$\text{Rinkos kapitalizacija} = \text{akcijos kaina} \times \text{akcijų kiekis}. \quad (1)$$

**Akcijos buhalterinės (apskaitinės) vertės ir kapitalizacijos santykis** (angl. *Book-to-Market Ratio*) – rinkos rodiklis, naudojamas įmonių akcijoms vertinti. Kompanijos buhalterinė vertė supaprastinus yra jos nuosavas kapitalas (turtas be įsipareigojimų). Paprastai, jei kompanijos kapitalizacija yra didesnė už jos buhalterinę vertę, jos akcijos laikomos pervertintomis.

$$\text{Buhalt. vertės ir kapitalizacijos santykis} = \frac{\text{Buhalterinė vertė}}{\text{Rinkos kapitalizacija}} \quad (2)$$

**Pelno marža** (angl. *Net Profit Margin*) – rinkos rodiklis, kuris parodo, kiek įmonė uždirba grynojo pelno lyginant su jos pajamomis. Paprastai pelno marža lyginama su kitų panašių kompanijų maržomis. Taip pat marža gali būti lyginama ir su tos pačios įmonės ankstesnio periodo rezultatais, įvertinant veiklos tendenciją.

$$\text{Pelno marža} = \frac{\text{Grynasis pelnas}}{\text{Pajamos}} \times 100 \% \quad (3)$$

**Grynasis pinigų srautas** (angl. *Free Cash Flow*) yra pagrindinės veiklos pinigų srautas be įsigytų investicijų (per tą patį ataskaitinį laikotarpį). Šis rodiklis parodo kiek įmonei realiai lieka pajamų iš pagrindinės veiklos.

$$\text{Grynasis pinigų srautas} = \text{Pagrindinis pinigų srautas} - \text{Investicijos į ilgalaikį turtą} \quad (4)$$

Grynųjų pinigų srauto analizė parodo kompanijos galimybes grąžinti skolas, išmokėti dividendus, supirkti akcijas ir vykdyti plėtrą. Grynasis pinigų srautas yra lyg pinigų perviršis, kurį kompanija gali naudoti kaip nori, nekenkdamas vykdomai veiklai. Kai kompanijos akcijos kaina yra žema, o grynųjų pinigų srautas auga, yra didelė tikimybė, jog pajamos ir akcijos kaina ateityje taip pat augs.

**Nuosavo kapitalo grąža** (angl. *Return On Equity - ROE*) – populiarus ir kartu lengvai suprantamas, bei paskaičiuojamas rodiklis, kuris parodo kiek efektyviai panaudojamas nuosavas įmonės kapitalas (įmonės savininkų investuoti pinigai ir turtas), t. y. kiek kiekvienam įmonės nuosavo kapitalo eurui tenka grynojo pelno.

$$\text{Nuosavo kapitalo grąža} = \frac{\text{Grynasis pelnas}}{\text{Nuosavas kapitalas}} \quad (5)$$

**Turto grąža** (angl. *Return on Assets – ROA*) rodo, ar kompanija veikia pelningai. ROA yra vienas pagrindinių kompanijos efektyvumo vertinimo rodiklių, parodantis kokią grąžą sugeneruoja kiekvienas į kompaniją investuotas euras. Šis rodiklis leidžia palyginti skirtingo dydžio kompanijų veiklos efektyvumą.

$$\text{Turto grąža} = \frac{\text{Grynasis pelnas} + \text{Sumokėtos palūkanos}}{\text{Visas kompanijos turtas}} \quad (6)$$

Įmonė su didesne turto grąža paprastai yra patrauklesnė investuotojui, nes tokia įmonė sugeba efektyviai didinti pelną, turint ribotą kiekį turto. Įmonėms su ganėtinai maža turto grąža, norint išlaikyti esamą pelno lygį, ateityje gali prireikti skolintis pinigų apyvartai.

**Einamojo likvidumo koeficientas** (angl. *Current Ratio*) rodo, koku laipsniu trumpalaikis turtas padengia trumpalaikius įsipareigojimus (ir kitas per vienerius metus mokėtinas sumas). Rodiklis vertina įmonės galimybę įvykdyti trumpalaikius įsipareigojimus, panaudojus turimą trumpalaikį turtą. Likvidumas priklauso nuo įmonės gebėjimo trumpalaikį turtą paversti pinigais ir taip apmokėti savo skolas bei padengti kitas mokėtinas sumas. Paprastai įvairiuose ekonomikos sektoriuose jis yra skirtingas.

$$\text{Einamojo likvidumo koeficientas} = \frac{\text{Trumpalaikis turtas}}{\text{Trumpalaikiai įsipareigojimai}} \quad (7)$$

**Įmonės vertės ir EBITDA santykis** (angl. *Enterprise Value to Earnings Before Interest, Taxes, Depreciation and Amortization – EV/EBITDA*) parodo, kiek įmonės akcijos kainuoja lyginant su jos EBITDA, atsižvelgiant į įmonės skolas. EBITDA – tai pelnas prieš palūkanas, mokesčius, nusidėvėjimą bei amortizaciją. EV/EBITDA rodiklio pranašumai labiausiai išryškėja lyginant didelį pinigų srautą generuojančias įmones arba bendroves, esančias skirtingose rinkose, kuriose galioja skirtingi apskaitos standartai, mokesčių tarifai, palūkanos už paskolas. EBITDA yra stabilesnis, ir mažiau įvairių veiksnių veikiamas rodiklis nei grynasis pelnas.

**Įmonės vertės ir pardavimų santykis** (angl. *Enterprise Value to Sales – EV/Sales*) parodo, kiek investuotojas pasirengęs sumokėti už vieną įmonės pardavimų eurą. Vienas pats rodiklis nėra labai informatyvus, ir paprastai yra naudojamas kartu su kitais santykiniais rinkos rodikliais. Tačiau daugiausia reikšmės EV/Sales turi, kai įmonė dirba nuostolingai, arba pelno marža yra labai žema ir nesiekia rinkos vidurkio.

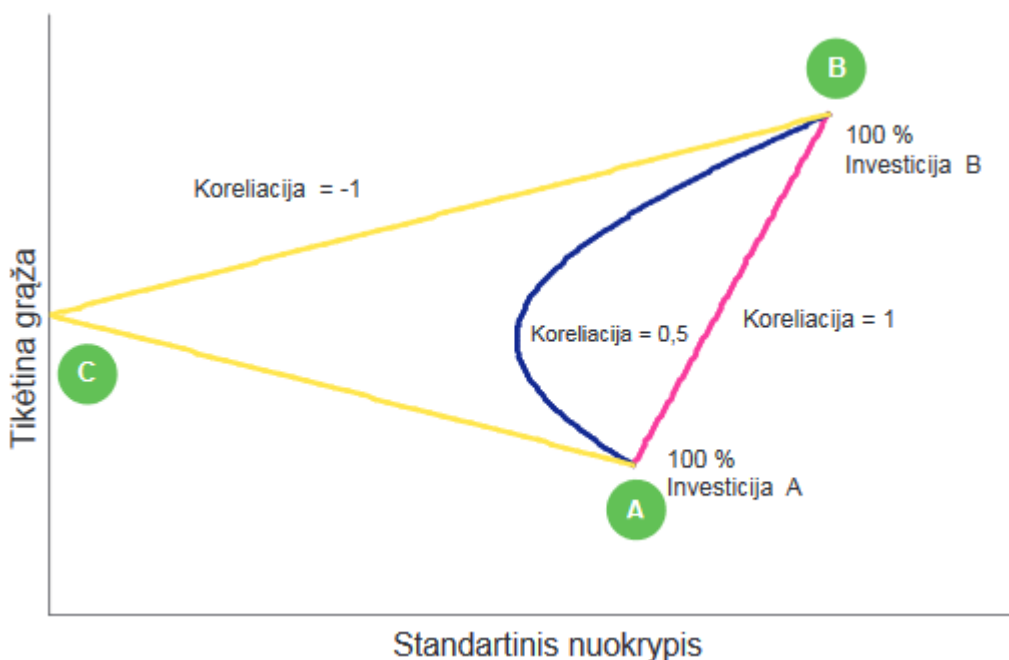
**Įsiskolinimo koeficientas** (angl. *Debt to Asset Ratio*) – finansinis rodiklis, lyginantis įmonės (finansines) skolas su jos visu turtu. Koeficientas parodo, kokia įmonės turto dalis yra finansuota iš skolintų lėšų. Paprastai laikoma, kad reikšmei esant žemiau 0,5 įmonės finansinė būklė yra gera, o reikšmei esant virš 0,7 – būklė nepatenkinama [6].

$$\text{Įsiskolinimo koeficientas} = \frac{\text{Grynoji skola}}{\text{Turtas}}. \quad (8)$$

## 1.2. Portfelio valdymas

Investavimo rizika kyla iš rinkos nepastovumo ir investicijų vertės svyravimų. Investicijų vertės svyravimas – tai vertybinių popierių arba rinkos indekso vertės svyravimo laipsnis apskaičiuotas pagal istorinius (statistinius) duomenis. Kitaip tariant, investicijų vertės svyravimas parodo neapibrėžtumo laipsnį arba riziką, kad investicijų vertė sumažės. Didesnis svyravimas reiškia, kad vertybinių popierių vertė gali kisti dideliame kainos intervale. Taigi per labai trumpą laiką investicijos vertė gali labai išaugti arba sumažėti. Mažas svyravimas reiškia, kad investicijos vertė per trumpą laiką labai nepasikeis ir ilguoju laikotarpiu vertės pokyčiai išliks vienodi [7].

**Modernioji portfelio teorija.** 1952 m. Haris Markovičius (*Harry Markowitz*) parašė straipsnį “Portfelio sudarymas”, kuriame pirmą kartą pateiktas optimalus investicinio portfelio nustatymo matematinis modelis, leidžiantis nustatyti ryšį tarp pajamingumo ir rizikos (vertės svyravimų). Portfelio laukiamam pelningumui įvertinti naudojamas portfelio gražos vidurkis, o rizikai – vidutinis standartinis nuokrypis arba dispersija. H. Markovičiaus teorija, dar vadinama moderniąja portfelio teorija, skirta pagrindinei investuotojų problemai – kaip investuoti kapitalą taip, kad esant tam tikram rizikos lygiui būtų gautas maksimalus pelnas, arba taip, kad norimas pelnas būtų pasiektas esant minimaliai rizikai [4, 7].



**1 pav.** Grafikas parodo, kaip teoriškai gali būti derinamos dvi investicijos siekiant gauti didžiausią įmanomą grąžą esant priimtinaam rizikos lygiui [4]

Aukščiau pateiktas grafikas rodo, kaip kinta portfelio grąža ir rizika, derinant dvi investicines skirtingų koreliacijų priemones. Taškas A parodo tik iš A investicijos sudaryto portfelio riziką ir tikėtiną grąžą, o taškas B – tuos pačius tik iš B investicijos sudaryto portfelio duomenis. Tiesiojoje A ir B taškus jungiančioje linijoje parodomos skirtingos procentinės A ir B investicijų dalys portfelyje, kai šių dviejų investicijų koreliacija yra ideali ir siekia +1. A ir B taškus jungianti kreivė parodo, kaip kintant investicijų proporcijoms keičiasi portfelio rizika ir tikėtina grąža, kai koreliacija tarp turto klasių yra 0,5. Kraštutiniu atveju, esant tobulai neigiamai koreliacijai, kreivė tampa A ir C bei C ir B taškus jungiančia tiese. Teoriškai tokiu atveju turėtų būti įmanoma suderinti dviejų klasių turto dalis taip, kad portfelio rizika būtų lygi 0, o tikėtiną grąžą parodytų C taškas vertikaloje ašyje.

Modernaus portfelio teorijoje šios kreivės vadinamos efektyvaus investicijų paskirstymo kreivėmis (angl. *efficient frontier*). Iš esmės tai yra toks portfelio paskirstymas tarp dviejų investicijų, kuris užtikrina didžiausią grąžą esant investuotojui priimtinos rizikos lygiui. Arba, kitaip tariant, jis nustato portfelio riziką, kurią investuotojui teks prisiimti siekiant nustatytos grąžos [4].

Modernioji portfelio teorija leidžia atrasti geriausią portfelio išdėstymą naudojant tik jau atrinktus vertybinius popierius. Tačiau investicinėse rinkose egzistuoja keli šimtai skirtingų vertybinių popierių, pvz., akcijos, obligacijos, valiutų rinka, opcionai, nekilnojamasis turtas, ateities sandoriai ir t.t. Šių vertybinių popierių kokybė svyruoja nuo labai gerų iki ypač prastų. Paprastai investuotojams yra sudėtinga atrasti potencialiai pelningus vertybinius popierius dėl nuolat kintančios informacijos apie juos ir jų kainų svyravimų. Todėl nėra protinga naudoti moderniąją portfelio teoriją su prastomis akcijomis. Tinkamesnis būdas būtų pirmiausiai atrinkti geros kokybės akcijas (naudojant fundamentalią analizę) ir tik tada optimizuoti portfelį, pasitelkiant H. Markovičiaus teoriją.

**Diversifikacija** – tai investavimo technika (gali būti investavimo strategijos dalis), kuri naudojama siekiant sumažinti investicijų portfelio riziką įtraukiant į jį didesnę kiekį skirtingų investicijų. Daugelis investuotojų žino, kad investicinį portfelį reikia diversifikuoti tarp skirtingų turto klasių, tačiau tik vienetai supranta, kodėl reikia taip elgtis. Investuojant į vieną turto klasę prisiimama nesisteminė rizika. Nesisteminė rizika – tai rizika, kuri būdinga vienai turto klasei, vienam regionui ar vienam pramonės sektoriui. Vienintelis būdas sumažinti nesisteminę riziką – diversifikuoti investicinį portfelį tarp skirtingų regionų, sektorių ir turto klasių [7].

Kai kurie įmonių sektoriai yra jautresni ekonomikos ciklams nei kiti. Pavyzdžiui, pramonės sektoriuje veikianti bendrovė labiau priklauso nuo bendro ekonomikos augimo nei farmacijos bendrovė. Mažmeninės prekybos maisto produktais sektoriuje veikiančios įmonės menkai reaguoja į ekonomikos ciklo pasikeitimus, o nekilnojamojo turto įmonių veiklai įtakos turi, pavyzdžiui, nedarbo lygis ir palūkanų normų pokyčiai. Tinkamu metu pasirenkant tinkamuose sektoriuose veikiančias investicijas, galima užtikrinti rinkos rezultatus viršijančią grąžą.

### 1.2.1. Portfelio kokybės vertinimas

Sudarytų portfelių kainos grafikams įvertinti pasirinktas *Standart & Poor's 500* (sutrumpintai *S&P 500*) lyginamasis indeksas. Jis yra sudarytas iš 505 didžiausios kapitalizacijos JAV bendrovių, kurios yra įtrauktos į *NYSE* ir *Nasdaq* biržų prekybos sąrašus. *S&P 500* indekso sudėtis yra nustatoma kiekvieną dieną ir apima apie 80% visos rinkos kapitalizacijos [8].

Be lyginamojo indekso portfelio kokybei įvertinti taip pat naudojami įvairūs vertinimo rodikliai. Jie padeda palyginti skirtingus portfelius ne tik pagal uždirbtą pelną, bet ir pagal rizikos lygį.

**Šarpo rodiklis** – vienas dažniausiai naudojamų grąžą ir riziką siejančių rodiklių. Jis leidžia palyginti perteklinę grąžą (faktinę grąžą be nerizikingos grąžos) su portfelio kintamumu (vertinamu pagal standartinį nuokrypį). Šarpo koeficientas leidžia apskaičiuoti grąžą, gaunamą už kiekvieną prisiimtą rizikos vienetą.

Pavyzdžiui, investuotojo portfelis per metus uždirbo 20 % grąžos, o standartinis nuokrypis buvo 40 procentų. Kitas investuotojas per tą patį laikotarpį uždirbo 15 %, tačiau jo standartinis nuokrypis buvo 10 %. Nerizikingų investicijų grąža siekė 5 %. Iš to galima daryti išvadą, kad pirmasis investuotojas sugebėjo uždirbti beveik pusę procentų grąžos kiekvienam jo prisiimtą rizikos procentui. Antrasis – 1 proc. grąžos kiekvienam prisiimtą rizikos procentui. Vadinas, nors antrojo investuotojo grąža buvo mažesnė, įvertinus prisiimtą riziką, jo efektyvumas buvo didesnis. Visgi, jeigu investuotojas yra pasirengęs prisiimti didesnę riziką, jis rinktųsi pirmąjį portfelį, nes tikėtina, kad jis uždirbs didesnę grąžą.

$$\text{Šarpo rodiklis} = \frac{R_p - R_f}{\sigma_p}; \quad (9)$$

čia  $R_p$  – vidutinė portfelio grąža;  
 $R_f$  – nerizikinga palūkanų norma;  
 $\sigma_p$  – portfelio standartinis nuokrypis.

**Sortino rodiklis** – kitas dažniau naudojamas metodas rizikai įvertinti. Skaičiavimo atžvilgiu labai panašus į Šarpo rodiklį, tik čia yra įvertinami ne visi fondo grąžos svyravimai, o tik neigiami arba tik tokie pokyčiai, kurie nebėra priimtini investuotojui.

$$\text{Sortino rodiklis} = \frac{R_p - R_f}{\sigma_d}; \quad (10)$$

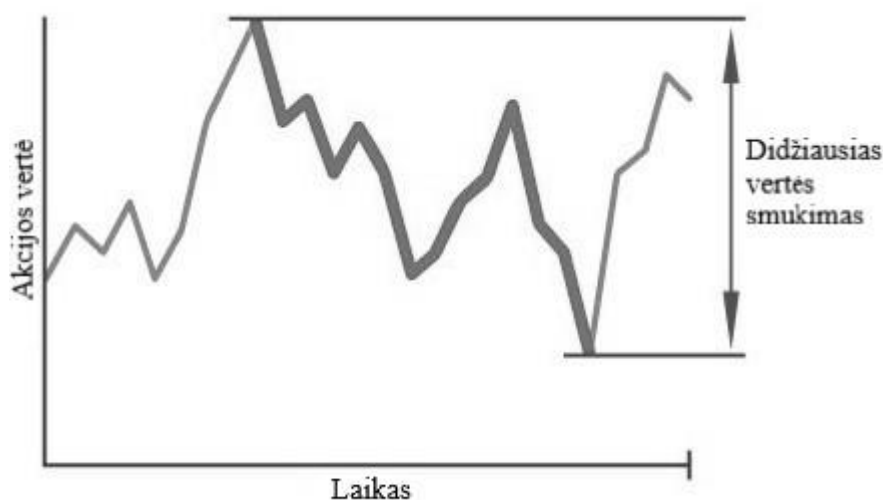
čia  $R_p$  – vidutinė portfelio grąža;

$R_f$  – nerizikinga palūkanų norma;

$\sigma_d$  – portfelio standartinis nuokrypis, atsižvelgiant tik į neigiamus (nepriimtinus) pokyčius.

**Didžiausias vertės smukimas** (angl. *Maximum Drawdown – MDD*) – yra didžiausias portfelio vertės kritimas nuo buvusios aukščiausios vertės per tam tikrą laikotarpį. Paprastai išreiškiamas procentais. Šis vertinimo rodiklis pabrėžia galimo turto praradimo dydį ir yra naudojamas įvairioms investavimo strategijoms palyginti [9].

$$MDD = \frac{\text{Mažiausia vertė} - \text{Didžiausia prieš tai buvusi vertė}}{\text{Didžiausia prieš tai buvusi vertė}}. \quad (11)$$



2 pav. Didžiausio vertės smukimo absoliutinė vertė

### 1.3. Tinkamas laikas investuoti

Bandant nuspėti tinkamą laiką pirkti akcijas, reikia žinoti kokiam laikotarpiui investuojama. Trumpalaikiam investavimui paprastai naudojama techninė analizė, kuria siekiama prognozuoti kainos svyravimus ateityje. Joje naudojami grafikai, vaizduojantys kainų dinamiką per tam tikrus laikotarpius. Techninė analizė remiasi tuo, kad kainoje atsispindi visi rinką įtakojantys faktoriai – ekonominiai, politiniai, psichologiniai ir kiti – visi jie sudaro bendrą kainos judesį.

Egzistuoja daugybė techninės analizės indikatorių, nuo stebinčių vien kainos grafiko formas iki nagrinėjančių ir akcijų pirkimo-pardavimo tendencijas. Vieni labiausiai naudojamų indikatorių yra šie [5, 10]:

- *Relative Strength Index (RSI)* – rodiklis, pranešantis apie vertybinio popieriaus per didelį pirkimą ar pardavimą.
- *Moving Average Convergence Divergence (MACD)* rodiklis praneša apie vertybinio popieriaus tendenciją ir inerciją (kaina kils ar kris, ir kiek smarkiai).
- *Stochastic Oscillator* yra plačiai naudojamas inercijos rodiklis. Jis įspėja apie užsitęsusią inerciją ir tendencijos krypties stiprumą.

Techninę analizę paprastai naudoja norintys uždirbti iš kainos svyravimų. Ji naudinga bandant prognozuoti kainos pokyčius kelioms dienoms ar vos valandoms. Ilgalaikiam investuotojui kainų svyravimai trumpuoju laikotarpiu sudaro tik nereikšmingą dalį investicijos pelno ar nuostolio.

Ilgalaikiam investavimui netinkamas laikas reiškia artėjančią kainų korekciją ar ekonominę nuosmukį. Tačiau tiksliai nuspėti šių reiškinių pradžią yra praktiškai neįmanoma. Todėl bendras patarimas yra investuoti periodiškai ir nuolatos. Investuojant tam tikrą sumą periodiškai sumažinama „blogo“ momento rizika, kai staiga nukritus akcijų kainoms iš karto prarandama dalis investuotos sumos.

Dėl šių priežasčių investavimo laiko pasirinkimas šiame darbe nebus nagrinėjamas.

## 2. Akcijų atrankos metodai

Atlikus fundamentalią įmonių analizę ir pasirinkus rodiklius, kurie geriausiai atspindi akcijos kainą arba kainos kitimą, atsiranda galimybė panaudoti šių kintamųjų reikšmes parenkant akcijas, kurias ruošiamasi įtraukti į investuotojo portfelį [11]. Jeigu tam tikro akcijos rodiklio koreliacija su akcijos kaina arba kainos didėjimu yra labai stipri, tai formuojant vertybinių popierių portfelį būtų tikslinga pasirinkti tokias akcijas, kurių šis rodiklis yra aukštas. Tačiau labai dažnai įvairių rodiklių reikšmės atskiroms akcijoms yra skirtingos. Todėl vertingoms akcijoms nustatyti yra naudojami įvairūs rangavimo būdai, mašininio mokymosi algoritmai bei kiti metodai.

### 2.1. Netiesinis rangavimas

Akcijos, į kurias verta investuoti, atrenkamos netiesiniu rangavimo metodu, kuris susideda iš dviejų etapų:

- Pirmajame etape parenkami rodikliai, kurie labiausiai atspindi vertybinių popierių kainą. Šiame etape gali būti panaudoti koreliacinės analizės rezultatai arba ekspertų nuomonė. Naudojant šiuos rodiklius atrenkami vertybiniai popieriai, kuriems vėliau atliekamas rangavimas.
- Antrajame etape vertybiniam popieriui už kiekvieną rodiklį skiriamas tam tikras balų skaičius, tam naudojant specialias netiesines funkcijas. Šių funkcijų analitinės ir grafines išraiškas, pasitelkęs savo patirtį, suformuoja pats investuotojas.

Pagal suminį balų už atskirus parametrus skaičių vertybiniam popieriui suteikiamas atitinkamas rangas. Aukšta vertybinių popierių rangavimo vieta yra neblogas ženklas, kad šias akcijas verta įtraukti į vertybinių popierių portfelį [10].

### 2.2. Neraiškioji logika

Pastaraisiais metais neraiškių aibių (angl. *fuzzy system*) metodai buvo sėkmingai naudojami sprendžiant valdymo uždavinius sudėtingose techninėse sistemose, ypač ten, kur klasikinių matematinių modelių sudarymas yra per daug komplikotas ir neefektyvus. Taip pat daugėja darbų, kur neraiškių aibių metodus bandoma panaudoti socialinėms ir ekonominėms sistemoms modeliuoti ir valdyti.

Neraiškiosios logikos taisyklės sudaromos „jeigu – tai“ teiginių pagalba. Šiuo metu neryškių aibių modeliai dažniausiai kuriami taikant tokius žingsnius:

1. Kintamųjų neraiškumo įvedimas. Šiame etape atrenkami proceso lingvistiniai kintamieji, jiems suteikiami lingvistinių termų vardai ir apibrėžiamos šių termų priklausomumo funkcijos.
2. Suformuojamos taisyklės, kurios aprašo, kaip lingvistinės išėjimo kintamųjų reikšmės priklauso nuo lingvistinių įėjimo kintamųjų reikšmių. Šiame etape taip pat nustatoma, kaip paveikiami sistemos išėjimo kintamieji, esant įvairiam taisyklių tenkinimo laipsniui.
3. Neraiškumo eliminavimas. Jo metu iš įvairių taisyklių tenkinimo laipsnių vaizdo pagal tam tikrus dėsnius suformuojamas galutinis sistemos išėjimas, kuris turi konkrečią kiekybinę išraišką [10].

Didžiausias neraiškiosios logikos trūkumas yra mokymosi galimybių stoka. Algoritmas efektyvus bus tiek, kiek į jį paduos vertingų taisyklių. Todėl pradedančiajam investuotojui, turinčiam mažai žinių apie vertybinių popierių rinką, neraiškiosios aibės metodas yra nerekomenduotinas.

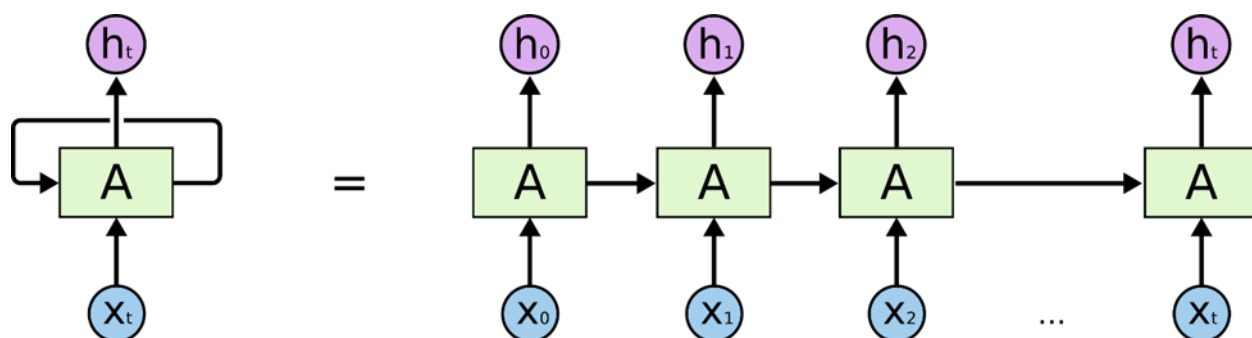


### 2.3. Rekurentiniai neuroniniai tinklai

Akcijų rinkos pobūdis yra netiesinis ir nepastovus, todėl daugybė tradicinių ir statistinių metodų yra taikoma akcijų rinkos prognozavimui. Skirtingų kombinacijų dirbtiniai neuroniniai tinklai (DNT) yra vienas dažniausiai naudojamų įrankių įvertinti rinkos pokyčius. Neuroninių tinklų gebėjimas mokytis bei apjungti netiesinę duomenų tendenciją puikiai tinka tokio pobūdžio uždavinių sprendimui. Be to, neuroniniai tinklai sugeba prisitaikyti prie duomenų modelio ir ryšio tarp įėjimų ir išėjimų, dėl to gaunamas kur kas tikslesnis spėjimas nei kitais tradiciniais metodais.[12]

Rekurentinis neuroninis tinklas (angl. *Recurrent Neural Network – RNN*), priklausantis dirbtinių neuroninių tinklų klasei, naudoja savo vidinę būseną (atmintį) įėjimų sekoms apdoroti. Tai leidžia šiuos tinklus taikyti rašysenos ar kalbos atpažinimui, bei užduotims, kurioms reikia atsižvelgti į istorinius duomenis.

Rekurentinį neuroninį tinklą galima išivaizduoti kaip daugybę to pačio tinklo kopijų, kur kiekvienas jų perduoda informaciją kitam. Tokia tinklo struktūra leidžia dirbti su duomenų sekomis ir sąrašais.



3 pav. Išskleistas rekurentinis neuroninis tinklas [13]

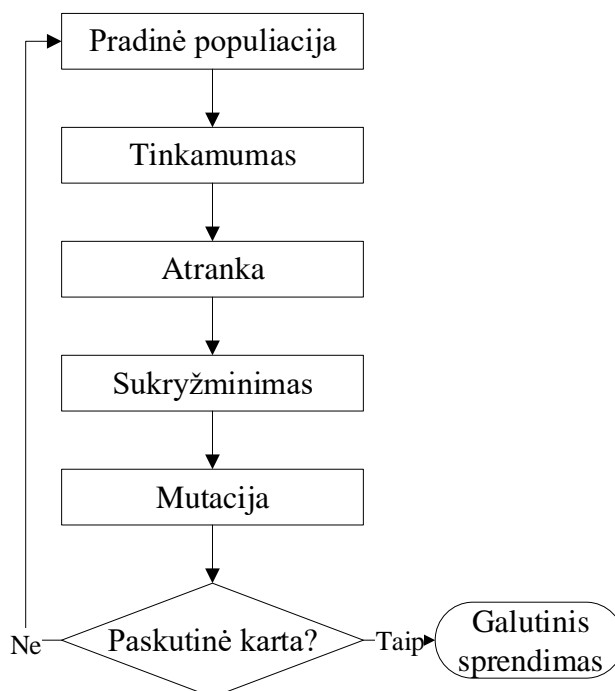
Ilgalaikės-trumpalaikės atminties (angl. *Long Short-Term Memory – LSTM*) tinklas yra rekurentinis neuroninis tinklas, kuris apmokomas naudojant atgalinį sklidimą laike (angl. *Backpropagation Through Time*) ir išsprendžia nykstančio gradiento problemą. Ši problema būdinga rekurentiniams tinklams, kai neuroninio tinklo svorius atnaujinantis gradientas tampa nykstamai mažas, todėl apmokymo metu svoriai nebėra atnaujinami ir gali nepasiekti optimalios konfigūracijos. [13, 14]

## 2.4. Evoliucinis algoritmas

Evoliucinis algoritmas (EA) yra algoritmas, paremtas biologijos žiniomis apie gyvybės evoliuciją. Šis algoritmas naudoja gamtoje egzistuojančius gyvybės evoliucinius mechanizmus: paveldėjimą, mutaciją, rekombinaciją ir natūraliąją atranką. Galimi optimizavimo problemos sprendimai yra tarsi individai tam tikroje populiacijoje, o algoritmo funkcija aprašo aplinką, kurioje sprendimai „gyvena“.

Evoliucinis algoritmas skirstomas yra keturis tipus: genetinis algoritmas (GA), genetinis programavimas (GP), evoliucinis programavimas (EP) ir evoliucinė strategija (ES). Iš visų šių tipų, populiariausias yra genetinis algoritmas.

GA ieško problemos sprendimo pasitelkiant skaičių eilutes (paprastai dvejetaines, nors geriausiai naudoti tokius, kurie labiausiai atspindėtų problemą). Populiacija susideda iš chromosomų, kuriose paprastai yra skirtingi genai. Pasirinkus pradinę populiaciją, evoliucija prasideda nuo visiškai atsitiktinių kitimų. Gavus naują populiacijos kartą (kandidatus) įvertinamas jos tinkamumas, atrenkamas tam tikras naujos kartos individų skaičius, pagal atrankos kriterijų. Atrinktieji individai pakeičiami darant mutacijas arba rekombinacijas ir sukuriama nauja populiacija. Vėliau viskas kartojama, atrenkant naujus tinkamiausius individus, sukuriant naują populiaciją. Ciklas kartojamas, kol gaunamas užduotį tenkinantis sprendimas.



4 pav. Genetinio algoritmo kūrimo etapai

Palyginus su kitais prekyboje taikomais algoritmais, genetinis algoritmas yra mažiau priklausomas nuo problemos ir suteikia didelę tikimybę pasiekti globalųjį optimumą. GA taip pat leidžia gauti nepilnai optimalų rezultatą. Kadangi finansiniame pasaulyje viskas greitai keičiasi, investuotojams, dėl laiko stokos, dažnai užtenka gauti neoptimalų, bet gana priimtina portfelio paskirstymo sprendimą [15, 16].

## 2.5. Kolektyviniai metodai

Pastaruoju metu rodiklių

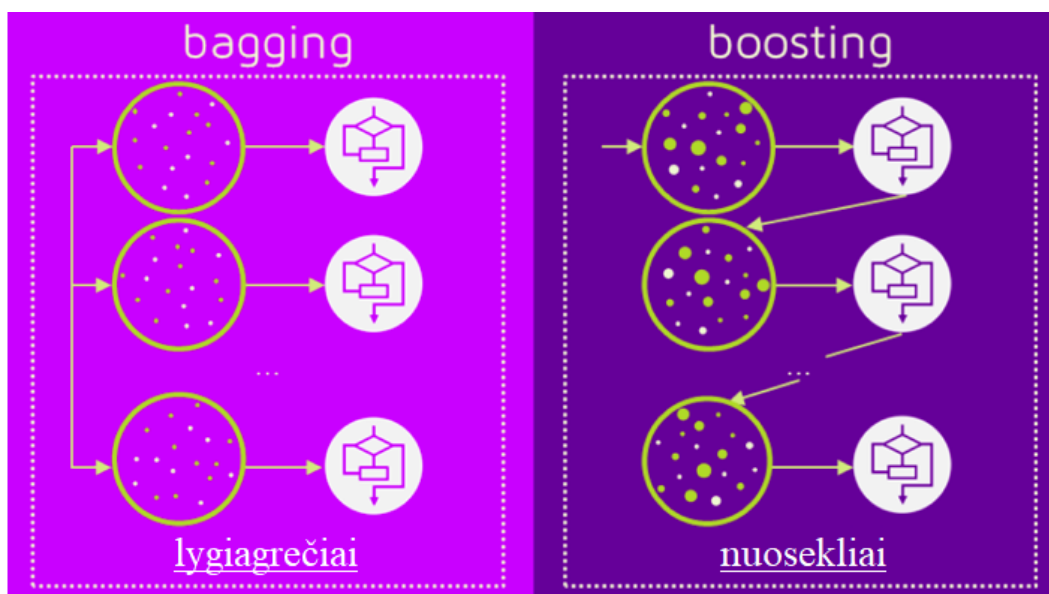
Dažniausiai paklaidos modelio mokymosi metu atsiranda iš trijų pagrindinių veiksnių: dispersijos (angl. *variance*), poslinkio (angl. *bias*) bei triukšmo. Su šiais veiksniais gerai susidoroja kolektyviniai metodai (angl. *ensemble methods*), kurie vietoj vieno modelio, naudoja jų grupę. Tokiu būdu sumažinama dispersija, pašalinami trukdžiai, atsirandantys iš pavienių modelių, užtikrinamas galutinio modelio stabilumas. Pagrindiniai kolektyviniai metodai yra tokie:

- *bagging* – sumažina dispersiją;
- *boosting* – sumažina poslinkį;
- *stacking* – padidina prognozavimo tikslumą.

Taikant *bagging* metodą sudaromas kolektyvas iš daugybės nepriklausomų modelių. Kiekvieno modelio apmokymui naudojama tik dalis visos duomenų imties ir tokiu būdu sudaromi skirtingi modeliai, kurie mažai koreliuoja tarpusavyje. Kombinuojant jų rezultatus gaunamas modelis su maža dispersija ir paklaidomis. Naudojant šį metodą, visi duomenys yra panaudojami vienodai dažnai, nėra persimokymo (angl. *overfitting*) problemos.

Taikant *boosting* metodą modeliai sudaromi paeiliui, kai kiekvienas jų mokosi iš prieš tai ėjusio klaidų. Dėl šios priežasties modelio apmokymas yra greitas, tačiau atsiranda persimokymo rizika. Taip pat duomenys yra panaudojami nevienodai dažnai, kadangi kiekvienas naujas modelis sutelkia daugiausia pastangų į sudėtingiausius duomenų pavyzdžius.

Taikant *stacking* metodą pirmiausiai apmokomi skirtingų tipų baziniai (pirmo lygmens) algoritmai, o jų rezultatai panaudojami antro lygmens algoritmų apmokymui. Taip modeliai „dėliojami“ vienas ant kito. Galiausiai, paskutinio lygmens algoritmų rezultatai panaudojami kaip pagrindinio modelio įėjimai. Šiuo metodu didinamas prognozavimo tikslumas, kadangi panaudojami heterogeniniai (skirtingų tipų) algoritmai [17].



5 pav. Skirtumas tarp *bagging* ir *boosting* modelių sudarymo [17]

Akcijų kainoms prognozuoti pasitelkiamas *bagging* metodu paremtas atsitiktinio miško modelis ir *boosting* metodu paremtas *Gradient Boosting* modelis.

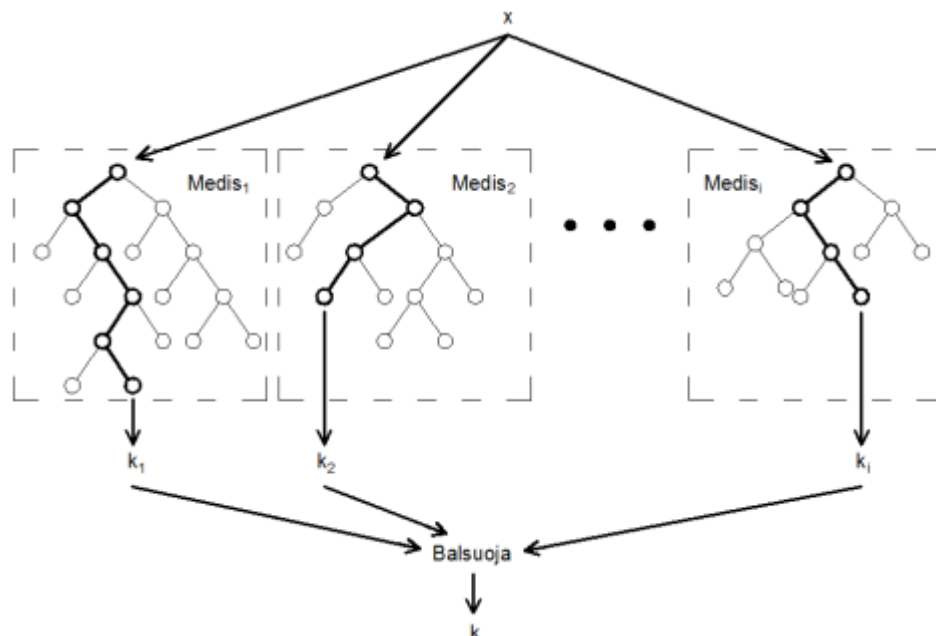
### 2.5.1. Atsitiktinis miškas

Atsitiktinis miškas (angl. *Random Forest – RF*) yra lankstus, gana lengvai naudojamas mašininio mokymosi algoritmas, paremtas Sprendimų medžių kolektyvu (angl. *ensemble of Decision Trees*). Atsitiktinis miškas sugeneruoja daug sprendimų medžių ir juos apjungia *bagging* metodu. Tokiu būdu gaunami tikslesni ir patikimesni rezultatai, nei naudojant vieną Sprendimų medį.

Iš pradžių yra generuojami sprendimų medžiai: kiekvieno klasifikavimo medžio šakos pasibaigia atitinkamai parinkta klase. Kiekvienas medis atlieka klasifikavimą ir galutiniame rezultate balsuoja už klases. Klasifikavimo atveju sprendimas gali būti priimamas skaičiuojant vidurkį arba pagal tai, kuri klasė surinko daugiausiai balsų, o regresijos atveju iš visų medžių gautų reikšmių skaičiuojamas vidurkis.

Didėjant medžių skaičiui atsitiktiniame miške, spėjimo tikslumas didėja iki tam tikros vertės. Mažas šališkumas ir maža koreliacija yra būtini norint gauti didelį tikslumą. Tam pasiekti medžiai yra auginami iki maksimalaus dydžio naudojant atsitiktines imtis:

- kiekvienas atskiras atsitiktinio miško medis yra sukuriamas panaudojant atsitiktinę imtį iš mokymui skirtų duomenų;
- sudarant medžius kiekviename mazge yra atsitiktinai parenkama  $n$  kintamųjų iš  $N$  galimų ir tik vienas kintamasis, lemiantis geriausią padalinimą, yra naudojamas iš  $n$  atrinktų.



6 pav. Principinė atsitiktinio miško modelio schema [18]

### 2.5.2. Gradient Boosting

*Gradient Boosting (GB)* – mašininio mokymosi algoritmas, kuris, kaip ir atsitiktinis miškas, paremtas Sprendimų medžių kolektyvu, tačiau veikiantis pagal *boosting* metodą. Algoritmas paremtas idėja, kad kitas geriausias modelis, sudarytas remiantis prieš tai buvusiais modeliais, sumažins prognozės paklaidą.

*Gradient Boosting* pradeda sprendimų medžio apmokymą, kiekvienam kintamajam suteikiant vienodus svorius. Kitame medyje, svoriai kintamiesiems priklauso nuo to, kaip smarkiai kiekvienas jų veikia bendrą prognozės paklaidą:

- Jei mažas kintamojo pokytis labai sumažinta paklaidą, tuomet kitame medyje šis kintamasis turės didelį svorį;
- Jei mažas kintamojo pokytis nepaveikia paklaidos, tuomet kitame medyje šio kintamojo svoris sumažės.

Tarkime, norint pagerinti algoritmo prognozės tikslumą, reikia mažinti vidutinę kvadratinę paklaidą (MSE):

$$MSE = \sum (y_i - y_i^p)^2; \quad (12)$$

čia  $y_i$  – i-oji tikroji reikšmė;  
 $y_i^p$  – i-oji prognozuojama reikšmė.

Naudojant gradientinio nusileidimo metodą ir atnaujinant prognozes nustatytu mokymosi greičiu, galima rasti reikšmes, su kuriomis MSE yra mažiausias:

$$y_i^p = y_i^p + \alpha * \frac{d \sum (y_i - y_i^p)^2}{d y_i^p}; \quad (13)$$

$$y_i^p = y_i^p - \alpha * 2 * \sum (y_i - y_i^p); \quad (14)$$

čia  $\alpha$  – mokymosi greitis;  
 $\sum (y_i - y_i^p)$  – absoliutinių paklaidų suma

Prognozės yra atnaujinamos taip, kad visų medžių daromų paklaidų suma artėtų prie 0 ir prognozuojama reikšmė artėtų prie tikrosios.

Įprastai dirbtinių neuroninių tinklų, ar kitų mašininio mokymosi algoritmų, kurie savo skaičiavimuose naudoja Euklidinį atstumą tarp dviejų taškų, apmokymui reikia pakeisti skirtingus kintamuosius taip, kad jie turėtų vienodas kitimo ribas arba išsiskaidymą. Kitu atveju, didesni skaičiai įgyja daug didesnius svorius nei mažesni. Atsitiktinio miško ir *Gradient Boosting* modeliai yra paremti Sprendimų medžiais, todėl jiems nėra būtina keisti duomenų mastelio [20, 21].

### 3. Metodologija

Tiriamąjį darbo tikslas yra sudaryti portfelį iš JAV biržose prekiaujamų įmonių akcijų, kuris pagal pelningumą ir kitus rodiklius sugebėtų aplenkti lyginamąjį indeksą – *S&P 500*. Šiai užduočiai spręsti sudaromas atsitiktinio miško ir *Gradient Boosting* metodais paremti regresorių modeliai. Prognozuojami akcijų kainos pokyčiai, pasitelkiant atrinktus įmonių finansinius duomenis. Iš atrinktų, potencialiai pelningiausių akcijų suformuojamas investicinis portfelis, kuris kas tam tikrą periodą performuojamas iš naujo. Portfelis įvertinimas pagal įvairius rodiklius.

#### 3.1. Duomenys

Tiksliam prognozavimui vienas esminių dalykų yra turėti duomenis, kurie labiausiai atspindi prognozuojamo rodiklio elgseną. Akcijos kainą ilguoju laikotarpiu labiausiai lemia įmonės finansiniai rezultatai: kapitalo pelningumas, pinigų srautas, pelno marža, įsiskolinimai ir kiti faktoriai, investuotojų vadinami fundamentaliais duomenimis.

Akcinės bendrovės, kurių akcijomis yra prekiaujama vertybinių popierių biržose, privalo viešai skelbti savo finansines ataskaitas, kuriose būtų pateikta informacija apie pelną, apyvartą, įsiskolinimus ir kitus rodiklius. Ir nors naujausi pranešimai yra laisvai prieinami, senesnių metų ataskaitas, pateiktas vienoje vietoje, gauti sudėtinga. Istoriniai, pavyzdžiui, daugiau nei metų senumo įmonių finansiniai rodikliai dažniausiai yra mokami (vieną išsamiausių duomenų bazių turi „Bloomberg Terminal“).

Tačiau po truputį atsiranda iniciatyvų, kurios sukaupiais finansiniais duomenimis dalinasi nemokamai. Viena iš tokių vietų yra svetainė [simfin.com](http://simfin.com), kurios autoriai automatizuotu būdu nuskaito įvairių JAV biržinių bendrovių finansines ataskaitas ir apdorotus rodiklius publikuoja viešai. Svetainėje įkeltas csv tipo failas, kuriame yra pateikta daugiau nei 1000 įmonių finansinės ataskaitos (ketvirčiais) nuo pat 2009 metų bei atitinkamos dienos akcijos kaina. Galima nagrinėti įmones net pagal 58 skirtingus rodiklius, nors dauguma ataskaitų nėra iki galo užpildytos.

##### 3.1.1. Duomenų bazės apdorojimas

Tačiau failas užima ypač daug vietos (450 MB), todėl naudoti jį modelio apmokymui, resursų ir greičio atžvilgiu, būtų neracionalu. Be to, faile įkelti duomenys yra sunkiai suprantami, todėl juos reikia apdoroti ir „išvalyti“.



Pasirinkti šie fundamentalūs rodikliai, kurie bus naudojami akcijų atrankai:

- rinkos kapitalizacija;
- akcijos buhalterinės vertės ir kapitalizacijos santykis;
- pelno marža;
- grynasis pinigų srautas;
- nuosavo kapitalo grąža;
- turto grąža;
- einamojo likvidumo koeficientas;
- įsiskolinimo koeficientas;
- įmonės vertės ir EBITDA santykis;
- įmonės vertės ir pardavimų santykis.

Dažnai dėl įvairių priežasčių finansinėse ataskaitose pateikti fundamentalūs rodikliai ne visai sutampa tarp skirtingų įmonių. Duomenų faile galima pamatyti, jog dauguma įmonių neturi vieno ar kito rodiklio visą ar dalį laikotarpio. Trūkstamų duomenų problema gali būti sprendžiama paprastais metodais, kai yra panaudojami esami duomenys, ar sudėtingesniais algoritmais, pavyzdžiui, Paslėpto Markovo modeliais. Tačiau kiekvienas užpildytas tuščias langelis yra tik apytikris spėjimas. Akcijų rinkoje dažnai pelną ar nuostolį lemia ir maža procento dalis, todėl kainoms prognozuoti geriausia taikyti tik esamus duomenis. Laiko eilutės su trūkstamais duomenimis yra pašalinamos.

Atlikus duomenų apdorojimą, finansiniai įmonių rodikliai yra paruošti mašininio mokymosi algoritmų apmokymui. 1 lentelėje parašyta, kiek įmonių priklauso kiekvienam sektoriui.

**1 lentelė.** Įmonių akcijų skaičius kiekviename sektoriuje

Sektoriaus pavadinimas	Įmonių skaičius duomenų bazėje
Technologijų/IT	319
Gamybinės produkcijos/transporto	260
Pramonės	218
Sveikatos apsaugos	230
Kasdieninio vartojimo prekių	97
Energetikos	91
Žaliavų	79
Nekilnojamojo turto	73



Time	Book to Market	Net Profit Margin	Free Cash Flow	Return on Equity	Return on Assets	Debt to Assets	Current Ratio	EV / EBITDA	EV / Sales	Market Cap	Share Price
2009-Q3	0.2724	0.2224	3845.0	0.0424	0.0212	0.2778	2.6687	40.1294	18.1511	97232.8	19.40
2009-Q4	0.284	0.2489	898.0	0.0522	0.0271	0.274	2.957	33.0816	15.8123	98276.58	19.62
2010-Q3	0.3313	0.1802	3878.0	0.0416	0.0207	0.2611	2.193	32.478	12.2561	98138.25	19.53
2010-Q4	0.2849	0.2179	1079.0	0.0537	0.0278	0.2533	2.4537	30.9457	13.3959	122335.22	24.22
2011-Q1	0.2481	0.2414	2863.0	0.0571	0.0313	0.218	2.8771	36.795	15.996	149350.11	29.51
2011-Q2	0.2582	0.2978	3767.0	0.0797	0.0436	0.2165	2.7603	27.7446	13.3083	155841.0	30.75
2011-Q3	0.3231	0.2197	5519.0	0.0446	0.0249	0.2002	3.0029	31.6386	13.269	127638.77	25.27
2011-Q4	0.297	0.2493	1153.0	0.0519	0.0301	0.2027	3.4756	32.6746	14.3787	142304.96	28.28
2012-Q1	0.3285	0.2764	3002.0	0.0577	0.0336	0.1987	3.2271	28.2291	12.9575	131694.55	26.45
2012-Q2	0.3748	0.3161	3587.0	0.0783	0.0441	0.2103	2.6009	19.0738	9.5107	117621.9	23.98
2012-Q3	0.3151	0.2486	5244.0	0.0465	0.0266	0.1929	2.7354	34.8421	14.9573	138794.63	28.73
2012-Q4	0.3124	0.2838	626.0	0.0595	0.0324	0.248	3.3311	29.8792	13.7962	138973.96	29.27
2013-Q1	0.2963	0.2795	2960.0	0.0572	0.0315	0.2486	3.4144	32.5638	15.0241	147799.08	31.32
2013-Q2	0.3145	0.3477	4198.0	0.0843	0.0465	0.2261	3.239	23.6159	11.9061	143561.4	30.90
2014-Q1	0.279	0.2756	3041.0	0.0567	0.0296	0.2795	3.3917	34.7103	16.0741	162105.28	36.16
2015-Q3	0.3176	0.2068	5523.0	0.0367	0.0159	0.3833	4.0363	39.6095	16.1242	149655.9	34.95
2015-Q4	0.2987	0.2443	417.0	0.0474	0.0206	0.3939	4.2619	39.4777	16.137	155106.88	36.86
2016-Q1	0.3191	0.2377	3267.0	0.0462	0.0204	0.3823	4.8261	35.7657	14.9818	145210.64	34.94
2016-Q2	0.3018	0.2657	3746.0	0.0589	0.0251	0.3909	3.7374	31.0337	13.8354	158341.23	38.33
2016-Q3	0.2954	0.2131	5454.0	0.0383	0.0149	0.4387	4.9169	45.0251	17.219	161940.64	39.44
2016-Q4	0.3083	0.2249	476.0	0.0416	0.0164	0.4395	3.96	41.8027	17.1467	158416.83	38.61
2017-Q1	0.3019	0.2432	2321.0	0.0439	0.0179	0.4304	4.1806	43.0967	17.8192	169003.05	41.07
2017-Q2	0.2983	0.2967	4497.0	0.0596	0.0239	0.429	3.0819	35.3246	15.9809	181862.52	43.96
2017-Q3	0.2762	0.2355	5886.0	0.038	0.016	0.3989	3.6258	53.3058	20.9675	204087.03	48.93
2017-Q4	0.2838	0.2309	408.0	0.0393	0.016	0.4372	5.0694	46.553	19.5941	198366.84	47.88
2018-Q1	0.2375	-0.4118	3566.0	-0.0834	-0.0291	0.4393	4.3261	46.1481	19.8506	203217.45	49.65
2018-Q2	0.2552	0.3071	4102.0	0.0745	0.0253	0.4397	3.9824	33.6035	15.6208	183702.12	45.96
2018-Q3	0.2109	0.2464	6305.0	0.0587	0.0176	0.4529	3.4255	50.3564	19.7306	182851.78	47.98
2018-Q4	0.1768	0.244	267.0	0.0751	0.0197	0.4905	2.8028	46.5802	19.3199	175607.2	48.35

## 8 pav. Apdoroti duomenys

### 3.1.2. Naudojamų duomenų imtis ir prognozavimo objektas

Kitas žingsnis yra nuspręsti, kiek toli į ateitį bus prognozuojama kaina. Paprastai investuotojai savo portfelį ir strategijas peržiūri bent kartą per metus, kuomet išleidžiamos finansinės ataskaitos. Kaip ir visose ateities prognozėse, taip ir investavime trumpesnis prognozuojamas laikotarpis leidžia tiksliau nuspėti kainos pokyčius. Šiame darbe pasirinkta prognozuoti būsimo ketvirčio grąžą (kainos pokytį per vieną laiko eilutę).

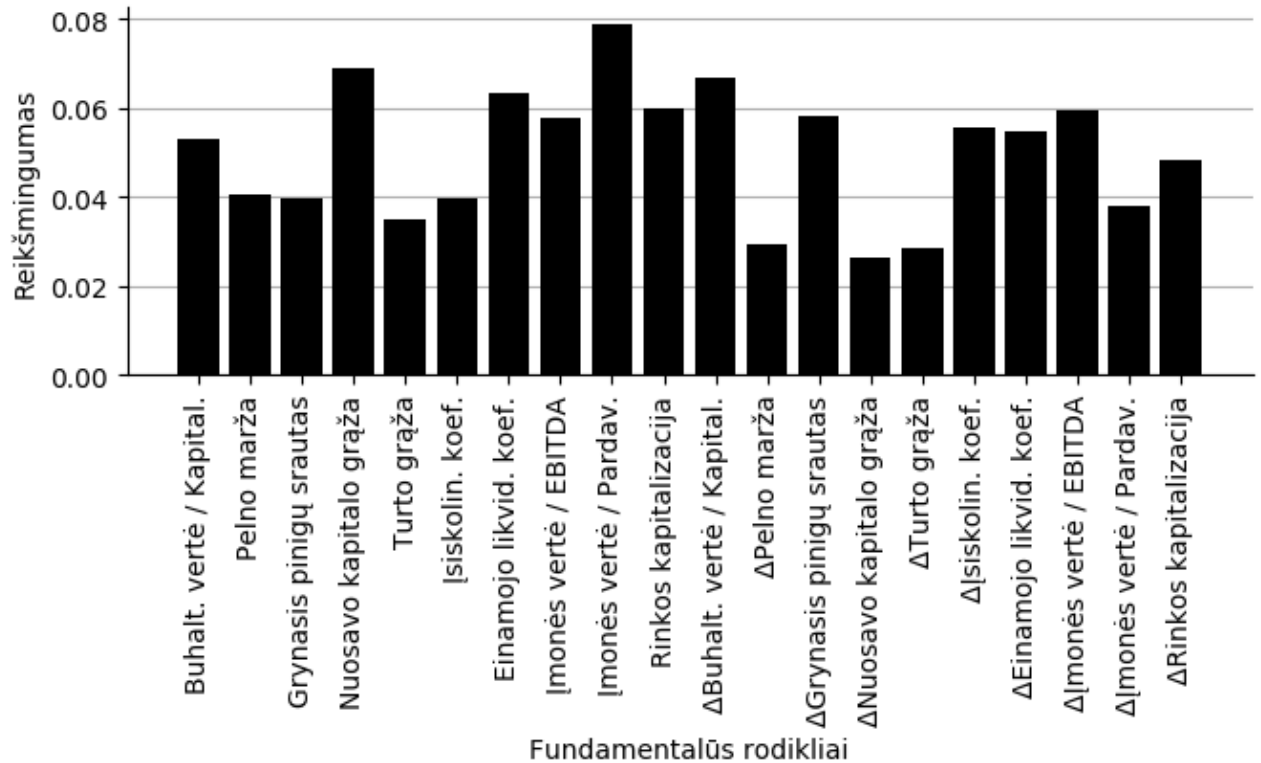
$$\text{Ketvirčio grąža} = \frac{\text{Dabartinė kaina} - \text{Kito ketvirčio kaina}}{\text{Dabartinė kaina}} \quad (15)$$

Tiksliai prognozei vien absoliutinių dydžių nepakanka. Modelio apmokymui papildomai panaudojami rodiklių metiniai pokyčiai (išvengiama sezoniškumo įtaka). Jie apskaičiuojami palyginus esamo ketvirčio rodiklių vertes su praeitų metų to paties ketvirčio rodiklių vertėmis. Pokyčiai nusako, kuria kryptimi ir koku greičiu juda rodiklio reikšmė. Net ir aukšta rodiklio reikšmė, jei ji sparčiai mažėja, gali lemti kainos kritimą [22].

Kiekviena duomenų laiko eilutė sudaryta iš 20 modeliui skirtų įėjimų (10 fundamentalių rodiklių reikšmių ir 10 rodiklių metinių pokyčių) bei vieno išėjimo (kito ketvirčio grąžą).

Naudojant atsitiktinio miško algoritmą, viena iš papildomų galimybių yra gauti kiekvienos duomenų grupės reikšmingumą atsitiktinio miško mokymuisi. Reikšmingumas atspindi, kiek tam tikras kintamasis pagerina prognozės tikslumą, lyginant su kitais kintamaisiais. Kuo didesnė ši reikšmė, tuo kintamasis daro daugiau įtakos galutiniam rezultatui.

Atskirų fundamentalių rodiklių įtaka kainos prognozavimui patikrinta su technologijų sektoriaus įmonėmis. Iš žemiau pateikto grafiko galima matyti, jog naudojamų rodiklių reikšmingumas nėra vienodas. Taip pat nustatyta, jog rodiklių įtaka tarp skirtingų sektorių yra labai nepastovi, kas patvirtina, kad kiekvienam sektoriui turi būti sudaryti atskiri modeliai.



9 pav. Kiekvieno fundamentalaus rodiklio reikšmingumo dalis atsitiktinio miško mokymuisi

### 3.2. Primityvi prognozė

Prieš pradėdant naudoti gilaus mokymosi modelius akcijų kainoms prognozuoti, dažnai pravartu sukurti primityvų, lengvai suprantamą modelį (angl. *naive*, *baseline model*). Šio modelio rezultatai būtų kaip atskaitos taškas, kurį reikia aplenksti, norint patikrinti pažangesnių mašininio mokymosi modelių efektyvumą. Tokie paprasti modeliai yra naudingi tuomet, kai bandoma išspręsti problemą, kuriai dar nėra žinomo sprendimo.

Kaip pavyzdys galėtų būti nesubalansuotos klasifikavimo užduotys, kuriose viena klasė yra daug dažniau pasitaikanti už kitas. Jeigu duomenų rinkinyje yra 90% grupės A atvejų ir 10% grupės B atvejų, tuomet primityvus metodas būtų visada prognozuoti grupę A. Tokio klasifikatoriaus tikslumas būtų 90%, todėl naudingas, mokymusi paremtas modelis turėtų prognozuoti tiksliau. Kartais tokius elementarius atskaitos taškus yra stebėtinai sudėtinga įveikti.

Šiame tiriamajame darbe kuriamas modelis, kuris prognozuotų kainos pokytį, todėl primityvi prognozė būtų tiesiog spėti, jog kitą ketvirtį kaina kis tiek pat, kiek kito per praėjusį ketvirtį.

### 3.3. Kryžminis patikrinimas

Modelio patikrinimui (validacijai) prie skirtingų modelio parametrų paprastai apmokymui skirti duomenys dalinami į apmokymo ir patikrinimo imtis. Tačiau kai duomenų imtis nėra didelė, patikrinimui skirta imtis yra labai maža. Todėl patikrinimo rezultatai tampa priklausomi nuo duomenų, kurie yra naudojami apmokymo ir patikrinimo imtyse. Tokia modelio validacija nėra patikima.

Geriausias sprendimas tokiai situacijai yra naudoti K-dalių kryžminį patikrinimą (angl. *K-fold cross validation*). Duomenų imtis padalinama į K dalis (paprastai į 4 ar 5). Tuomet sukuriama K skaičius modelių, kiekvienas iš jų apmokomas su  $K - 1$  skaičiumi dalių ir patikrinamas su likusia viena dalimi. Šio metodo privalumas – visa apmokymo imtis yra išnaudojama tiek apmokymui, tiek patikrinimui.

Atsitiktinio miško ir *Gradient Boosting* modeliai tikrinami naudojant 5 dalių kryžminį patikrinimą.

Duomenų imtis

Išskaidymas 1	Patikrinimas	Apmokymas	Apmokymas	Apmokymas	Apmokymas
Išskaidymas 2	Apmokymas	Patikrinimas	Apmokymas	Apmokymas	Apmokymas
Išskaidymas 3	Apmokymas	Apmokymas	Patikrinimas	Apmokymas	Apmokymas
Išskaidymas 4	Apmokymas	Apmokymas	Apmokymas	Patikrinimas	Apmokymas
Išskaidymas 5	Apmokymas	Apmokymas	Apmokymas	Apmokymas	Patikrinimas

10 pav. Penkių dalių kryžminio patikrinimo išskaidymas

### 3.4. Modelio parametrai

Algoritmo parametrų vertės nustatomos prieš pradedant modelio apmokymą. Naudojant atsitiktinio miško modelį, dažnai ir su numatytais parametrais galima tikėtis neblogų rezultatų. Šio modelio parametrų yra daugiau nei 10, tačiau esminiai yra du [23]:

- medžių skaičius modelyje – paprastai didesnis skaičius reiškia geresnį, tačiau, tuo pačiu, ir lėtesnį apmokymą duomenimis;
- didžiausias sprendimo medžio gylis – didesnis skaičius reiškia didesnę medžio išsišakojimą, kas leidžia modeliui gauti daugiau informacijos apie duomenis.

Pasiekti bent pakankamą rezultatą su *Gradient Boosting* modeliu yra kiek sunkiau nei su atsitiktiniu mišku. Jam derinami šie parametrai [24]:

- medžių skaičius modelyje;
- didžiausias sprendimo medžio gylis;
- nuostolių funkcija – funkcija, kurios rezultatai stengiamasi sumažinti;
- mokymosi greitis – rodiklis, kuriuo nurodoma, kokia dalis prieš tai buvusio medžio prognozės yra panaudojama kito medžio prognozei. Paprastai geresnė mažesnė rodiklio reikšmė, kadangi tai leidžia įsisavinti specifines medžių savybes, tačiau tuo pačiu reikia ir didesnio medžių skaičiaus modelyje.

### 3.5. Prognozių tikslumo rodikliai

Atsitiktinio miško ir *Gradient Boosting* modelių parametrai derinami pagal tai, kokios yra gaunamos paklaidos modelio patikrinimo metu. Naudojamos trijų tipų paklaidos.

- Vidutinė absoliutinė paklaida (MAE) yra vidutinis absoliutinis skirtumas tarp prognozės ir tikrosios vertės. Prognozuojant kainos pokytį, MAE parodo, kiek, vidutiniškai, procentų suklysta modelis.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^p|. \quad (16)$$

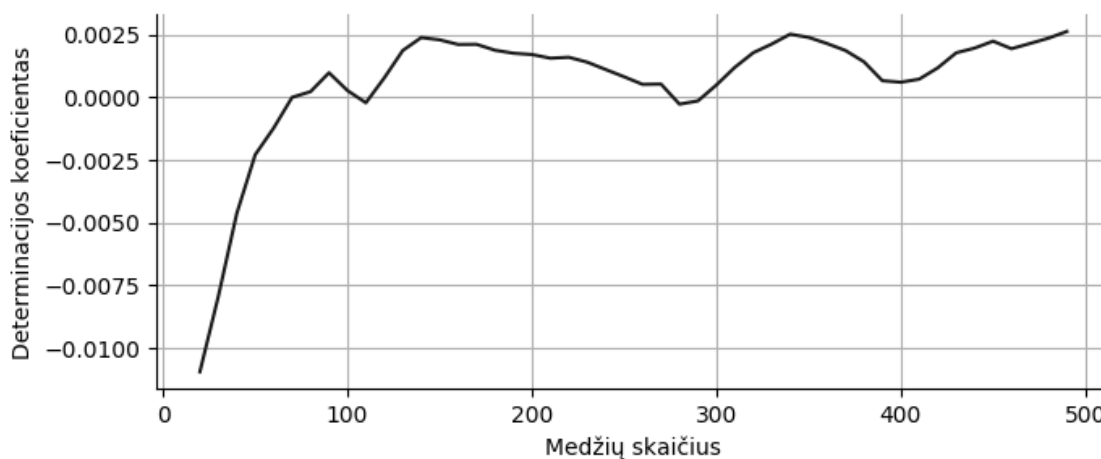
- Šaknis iš vidutinės kvadratinės paklaidos (RMSE) parodo vidutinį paklaidos dydį. Apskaičiuojamas, ištraukus šaknį iš vidutinio kvadratinio skirtumo tarp prognozės ir tikrosios vertės.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^p)^2}. \quad (17)$$

- Determinacijos koeficientas ( $R^2$ ) lygina skirtumus tarp  $y$  reikšmių, kai atsižvelgiama į regresijos modelį, su skirtumais tarp  $y$  reikšmių, kai į modelį neatsižvelgiama. Reikšmės kinta  $[-\infty, 1]$  ribose. Geriausias rezultatas yra 1, o neigiamos reikšmės rodo prastą modelio veikimą [25].

$$R^2 = 1 - \frac{\sum_i (y_i - y_i^p)^2}{\sum_i (y_i - y_{vid})^2}; \quad (18)$$

- čia  $n$  – iteracijų skaičius;  
 $y_i$  – tikroji reikšmė;  
 $y_i^p$  – prognozuojama reikšmė;  
 $y_{vid}$  – vidutinė tikroji reikšmė.



11 pav. Modelio tikslumo priklausomybė nuo medžių skaičiaus

Modeliui įvertinti taip pat naudojamas akcijos kainos pokyčio krypties prognozės tikslumas, t. y., kaip dažnai prognozuojamos reikšmės ženklas sutampa su tikrosios vertės ženklu. Visoms laiko eilutėms iš kiekvieno sektoriaus krypties prognozės tikslumas yra apskaičiuojamas ir sudedamas su prieš tai ėjusių laiko eilučių tikslumu. Vertinama gauta paskutinės laiko eilutės vidutinė visų sektorių reikšmė.

Krypties prognozės tikslumas naudojamas, nes dažnai užtenka atspėti, ar akcijos kaina kils, ar kris, tam, kad priimtum sprendimą dėl investavimo. Taip pat, prognozuoti kainos pokyčio kryptį yra daug lengviau nei prognozuoti pačios kainos pokytį.

### 3.6. Modelio apmokymas ir testavimas

Įprastai, naudojant mašininio mokymosi algoritmus, duomenys yra atsitiktinai išmaišomi ir padalinami į apmokymo ir testavimo duomenų imtis. Šiame darbe kintamieji yra priklausomi nuo laiko, todėl jų eiliškumas yra svarbus. Taip pat testavimui būtina naudoti tokią duomenų imties dalį, kuri pagal laiką eitų po apmokymo imties. Dėl šių priežasčių duomenys nėra išmaišomi.

Nors duomenų imtis tęsiasi iki pat 2018 metų, tačiau yra gana mažai duomenų iš šių metų. Todėl bendras laikotarpis apribojamas tarp 2009 ir 2017 metų.

Duomenys į apmokymo ir testavimo imtis atskirti pagal datą. Nustatyta, kad datą pasirinkus ties 2015 metų ketvirtuoju ketvirčiu, imtys pasidalina į mašininio mokymosi algoritms įprastines 65 % ir 35 % proporcijas. Tiriamajam darbui pasirinkta prognozuoti 2016-2017 metų kainos pokyčius – iš viso 8 laiko eilutes.

Atsitiktinis miškas apmokomas ir ištestuojamas etapais. Pirmiausiai modelis apmokomas ir patikrinamas, naudojant duomenis iš laikotarpio nuo 2009 metų pradžios iki 2015 metų trečio ketvirčio. Testavimui naudojami 2015 metų ketvirto ketvirčio duomenis ir prognozuojama akcijų grąža per 2016 metų sausio – kovo mėnesius. Kitame etape prie apmokymo duomenų imties pridama 2015 metų ketvirto ketvirčio laiko eilutė, testuojama ir prognozuojama vienu ketvirčiu į priekį.

Procesas kiekvienam sektoriui kartojamas 8 kartus, kol prognozuojama data pasieks 2018 metų sausio 1 dieną.

		Duomenų laiko eilutės								
		2009-Q1 – 2015-Q3	2015-Q4	2016-Q1	2016-Q2	2016-Q3	2016-Q4	2017-Q1	2017-Q2	2017-Q3
Prognozuojama data	2016-04-01									
	2016-07-01									
	2016-10-01									
	2017-01-01									
	2017-04-01									
	2017-07-01									
	2017-10-01									
	2018-01-01									

12 pav. Modelio apmokymo ir testavimo etapai

### 3.7. Portfelio sudarymas

Suderintas ir apmokytas modelis panaudojamas akcijų, kurių kaina, tikėtina, labiausiai augs, atrankai. Pradedant nuo 2016 metų, į apmokytą modelį yra paduodami vieno ketvirčio duomenys. Iš gautų rezultatų akcijos pasirenkamos dviem būdais:

- iš kiekvieno sektoriaus atrenkamos po 2 perspektyviausias įmonių akcijas (iš viso 16).
- iš visų testuojamų įmonių akcijų atrenkamos 16 perspektyviausių;

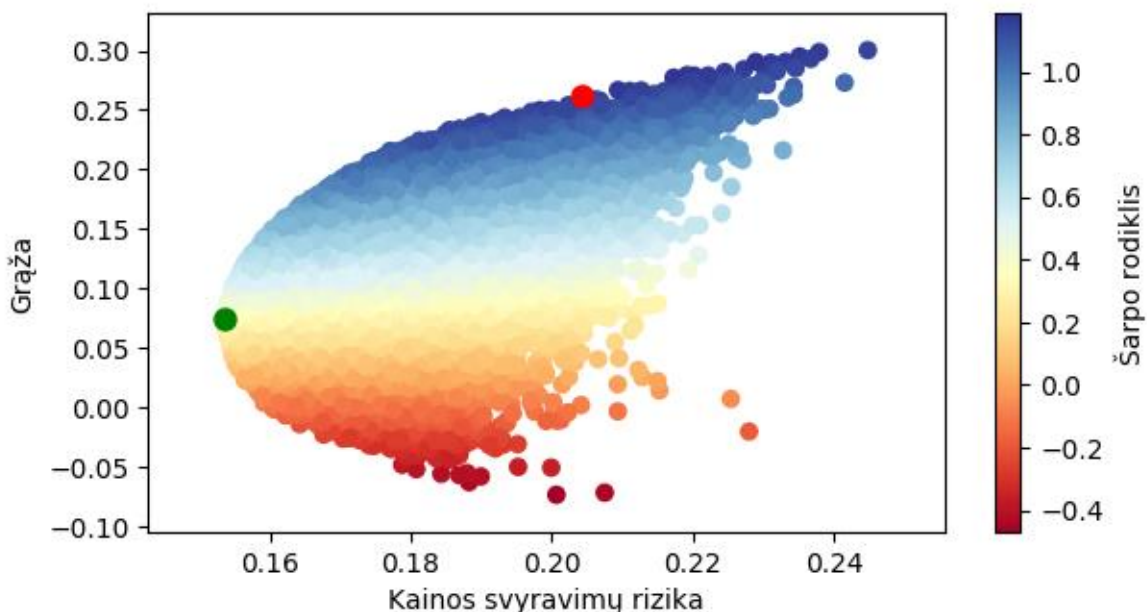
Atėjus kitam ketvirčiui, portfelis sudaromas iš naujo pagal naujas modelio prognozes.

Idealų skirtingų akcijų skaičių portfelyje yra labai sunku nustatyti. Tai priklauso nuo daugybės veiksnių, pavyzdžiui, makroekonominių rodiklių, investicinio laikotarpio, kaip dažnai atnaujinamas portfelis ir panašiai. Tačiau įvairūs tyrimai teigia [26], jog optimalus akcijų skaičius svyruoja tarp 5 ir 20. Didesnis skaičius jau ne tik nebedidina diversifikacijos, bet ir mažina galimą pelną. Todėl sudaromas diversifikuotas portfelis, atrenkant po 2 akcijas iš 8 sektorių, bei tokio paties dydžio nediversifikuotas portfelis, kurio akcijos atrenkamos neatsižvelgiant į sektorių. Pirmuoju būdu portfelis diversifikuojamas tarp skirtingų sektorių, taip sumažinant nesisteminę riziką. Atrinkus akcijas antruoju būdu, tikimasi pasiekti pelningiausią įmanomą portfelį.

Atrinkus akcijas, kitas žingsnis yra nustatyti optimalų vertybinių popierių pasiskirstymą investiciniame portfelyje. Norint gauti gerai diversifikuotą portfelį, reikia jį sudaryti iš daugiau mažai koreliuojančių ir mažai rizikingų akcijų. Tyrimo metu lyginami portfeliai, sudaryti iš taip pasiskirsčiusių akcijų:

- vienodų svorių – visos akcijos paskirstytos vienodai;
- didžiausias Šarpo – akcijos paskirstytos pagal istoriškai didžiausią Šarpo rodiklį turintį portfelį;
- mažiausiai rizikingas – akcijos paskirstytos pagal istoriškai mažiausiai rizikingą portfelį.

Pastarieji du portfeliai sudaromi pritaikant Moderniąja portfelio teoriją. Pasitelkus istorinius paskutinių vienerių metų duomenis iš atrinktų akcijų atsitiktiniu būdu yra sukuriama 20000 skirtingų portfelių, kuriuose akcijos turi įvairius svorius. Bendra svorių suma sudaro 1. Tuomet suskaičiuojama kiekvieno portfelio praėjusių metų grąža, standartinis nuokrypis (kainos svyravimų rizika) bei Šarpo rodiklis (nerizikinga grąža yra pastovi ir lygi 0 %). Visų portfelių rodikliai yra atvaizduojami viename grafike.



13 pav. 20000 įvairių portfelių 2016 metų grąžos ir standartinio nuokrypio santykiai

Kaip pavyzdys 13 paveiksle pavaizduoti 20000 portfelių, sudarytų iš „Microsoft Corp.“, „IBM Corp.“, „McDonald’s Corp.“, „Visa Inc.“ ir „AT&T Inc.“ įmonių akcijų. Akcijų pasiskirstymas visuose portfeliuose yra skirtingas. Linija, einanti palei figūros kraštą, yra efektyvaus investicijų pasiskirstymo kreivė. Kiekvienas taškas, esantis ant šio kreivės, atspindi optimalią akcijų kombinaciją turintį portfelį, kuris uždirba didžiausią grąžą prie esamo rizikos lygio. Žaliu tašku pažymėtas portfelis yra mažiausiai rizikingas (turintis mažiausiai svyravimų), o raudonu – portfelis, turintis didžiausią Šarpo rodiklio reikšmę.

Istorinė portfelių grąža ir vertės svyravimų rizika yra paskaičiuota iš laikotarpio, kuris atitinka apmokymo imties paskutinių vienerių metų laikotarpį. Tokiu būdu sudarant mažiausiai rizikingus ir didžiausią Šarpo rodiklį turinčius portfelius, naudojami paskutinių metų kainos kitimo duomenys.

Apjungus visus šiuos portfelio formavimo būdus, vienam laiko ketvirčiui gaunami 12 skirtingų portfelių: dviem metodais atrinkamos akcijos, iš atrinktų akcijų dviem būdais sudaromi portfeliai ir sudarytuose portfeliuose akcijos paskirstomos trimis būdais.

		Vienas laiko ketvirtis											
Akcijų atranka		Atsitiktinis miškas						Gradient Boosting					
Portfelio sudarymas		Pelningiausios			Diversifikuojant			Pelningiausios			Diversifikuojant		
Portfelio optimizavimas		VS	DŠ	MR	VS	DŠ	MR	VS	DŠ	MR	VS	DŠ	MR

14 pav. 12 portfelių sudarymo strategijų

### 3.8. Portfelijų vertės rodikliai

Sudarytų portfelijų kokybė vertinama pagal portfelio vertę viso laikotarpio pabaigoje (portfelio vertė laikotarpio pradžioje prilyginama 100 piniginių vienetų), vidutinę metinę grąžą, didžiausią vertės smukimą bei Šarpo ir Sortino rodiklius.

Kadangi sudaromi portfeliai yra skirti ilgalaikiam investavimui, Šarpo ir Sortinio rodiklių skaičiavimuose naudojama grąža yra mėnesinė. Paprastumo dėlei, nerizikinga palūkanų norma prilyginta 0 %. Skaičiuojant Sortino rodiklį, pageidautina grąža taip pat lygi 0 % (skaičiuojant svyravimus įtraukiami tik neigiami pokyčiai).

Portfeliai kuriami vien iš JAV kapitalo įmonių, todėl yra tikslinga kaip lyginamąjį indeksą pasirinkti *S&P 500*. Šis indeksas padės ne tik palyginti portfelijų rezultatus su JAV rinka, bet atskirti, kada nuostoliai patiriami dėl makroekonominių priežasčių, ir kada dėl prastai parinktų akcijų.



## 4. Rezultatai

### 4.1. Primityvi prognozė

Sudarius modelį, kuris dvejus metus prognozuoja, jog akcijos kaina kitą ketvirtį kis tiek pat, kiek kito praėjusį ketvirtį, gauta vidutinė absoliutinė paklaida lygi 0,203.

### 4.2. Lyginamasis indeksas

S&P 500 indekso dviejų metų rezultatai pateikti 15 paveiksle ir 2 lentelėje.



15 pav. Lyginamojo indekso S&P 500 kainos grafikas

2 lentelė. Lyginamojo indekso S&P 500 rezultatai

Vertinimo rodiklis	Lyginamasis indeksas
Portfelio vertė $V_{2018}$	137,45
Vidutinė metinė grąža	17,5 %
Didžiausias smukimas	-9,19 %
Šarpo rodiklis	0,88
Sortino rodiklis	4,92

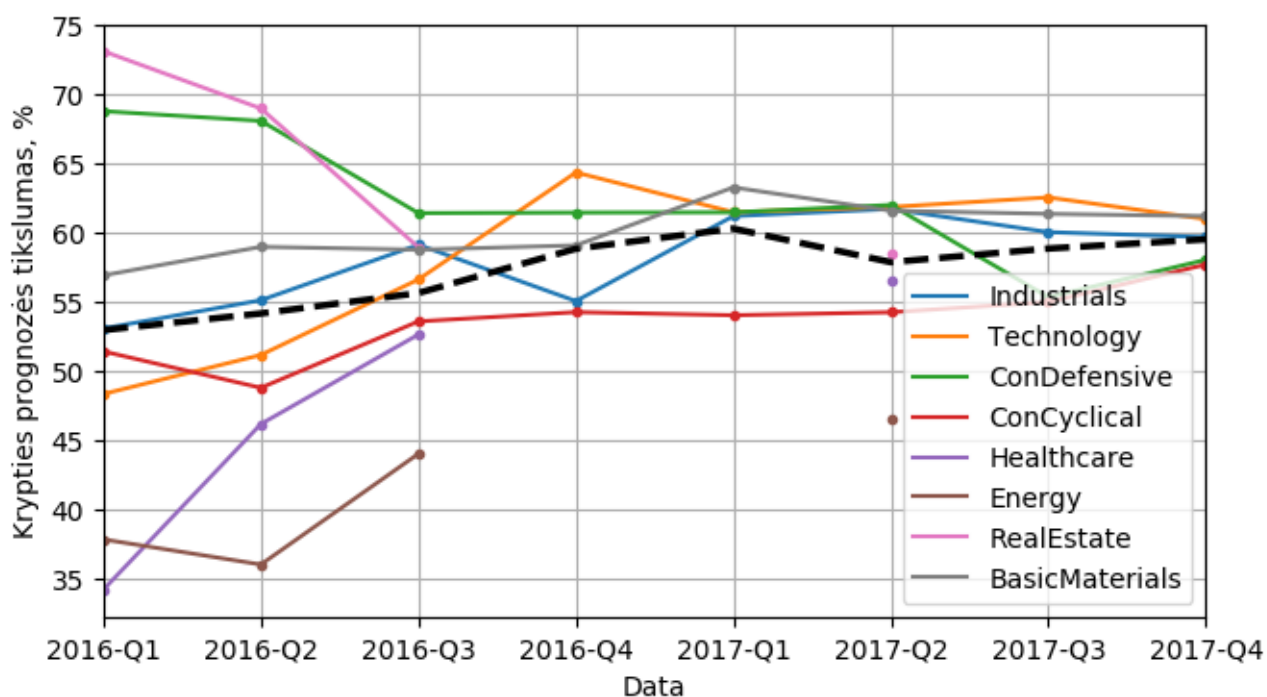
### 4.3. Atsitiktinio miško metodas

Modelio patikrinimo metu nustatyti geriausi parametrai pateikti 3 lentelėje.

**3 lentelė.** Geriausio atsitiktinio miško modelio parametrai

Medžių skaičius	150
Maksimalus gylis	Medis didinamas iki tol, kol pasiekia minimalų šakai reikalingą kintamųjų skaičių
Kokybės kriterijus	Vidutinė kvadratinė paklaida

#### 4.3.1. Modelio vertinimas



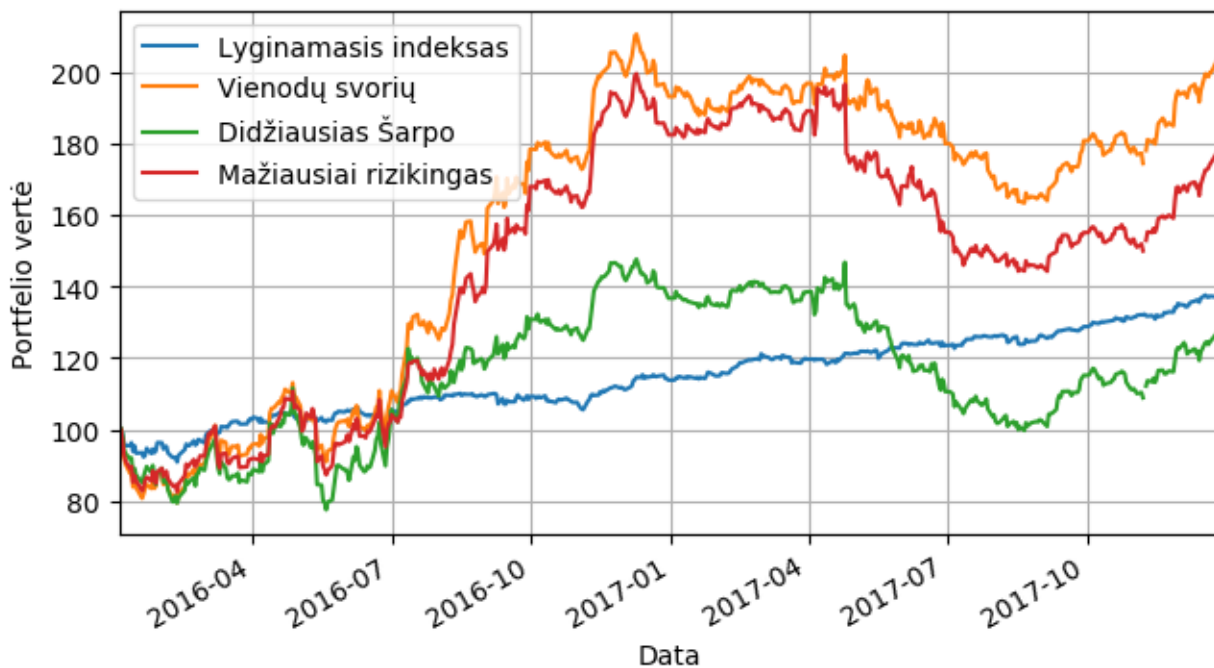
**16 pav.** Akcijų kainos pokyčių krypties prognozės tikslumas tarp skirtingų sektorių, naudojant atsitiktinio miško modelį

**4 lentelė.** Geriausio atsitiktinio miško modelio prognozės tikslumo rodikliai

Krypties prognozės vidutinis tikslumas	59,51 %
Vidutinė absoliutinė paklaida	0,13
Determinacijos koeficientas ( $R^2$ )	-0,52
Šaknis iš vidutinės kvadratinės paklaidos	0,23

### 4.3.2. Portfelijų vertinimas

Investicinių portfelijų, sudarytų iš atsitiktinio miško modelio atrinktų 16 perspektyviausių akcijų, rezultatai pateikti 17 paveiksle ir 5 lentelėje.

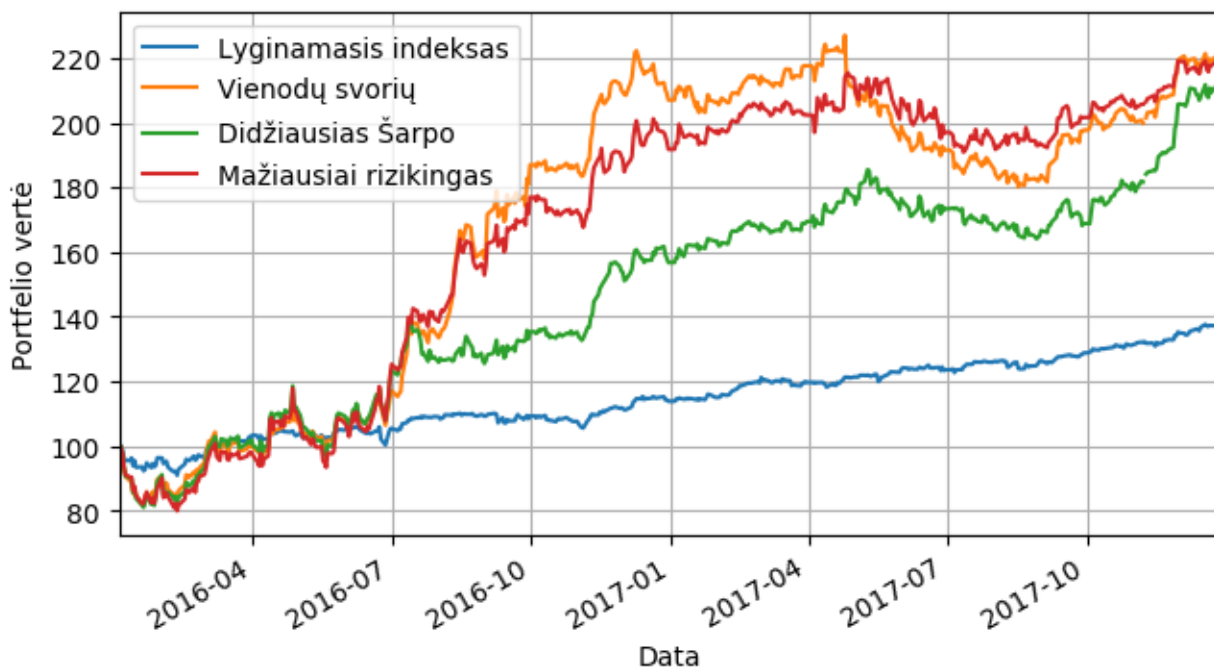


17 pav. Portfelijų, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (16 perspektyviausių), kainų grafikai

5 lentelė. Portfelijų, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (16 perspektyviausių), rezultatai

Vertinimo rodiklis	Vienodi svoriai	Didžiausias Šarpo	Mažiausiai rizikingas
Portfelio vertė $V_{2018}$	201,23	125,86	176,37
Vidutinė metinė grąža	49,4 %	14,93 %	40,17 %
Didžiausias smukimas	-22,42 %	-32,43 %	-27,69 %
Šarpo rodiklis	0,49	0,22	0,37
Sortino rodiklis	1,77	0,41	1,01

Investicinių portfelių, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (po 2 iš sektoriaus), rezultatai pateikti 18 paveiksle ir 6 lentelėje.



**18 pav.** Portfelių, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (po 2 iš sektoriaus), kainų grafikai

**6 lentelė.** Portfelių, sudarytų iš atsitiktinio miško modelio atrinktų akcijų (po 2 iš sektoriaus), rezultatai

Vertinimo rodiklis	Vienodi svoriai	Didžiausias Šarpo	Mažiausiai rizikingas
Portfelio vertė $V_{2018}$	221,15	212,66	220,26
Vidutinė metinė grąža	57,92 %	46,77 %	54,49 %
Didžiausias smukimas	-20,66 %	-18,91 %	-20,73 %
Šarpo rodiklis	0,61	0,8	0,74
Sortino rodiklis	2,2	4,39	2,95

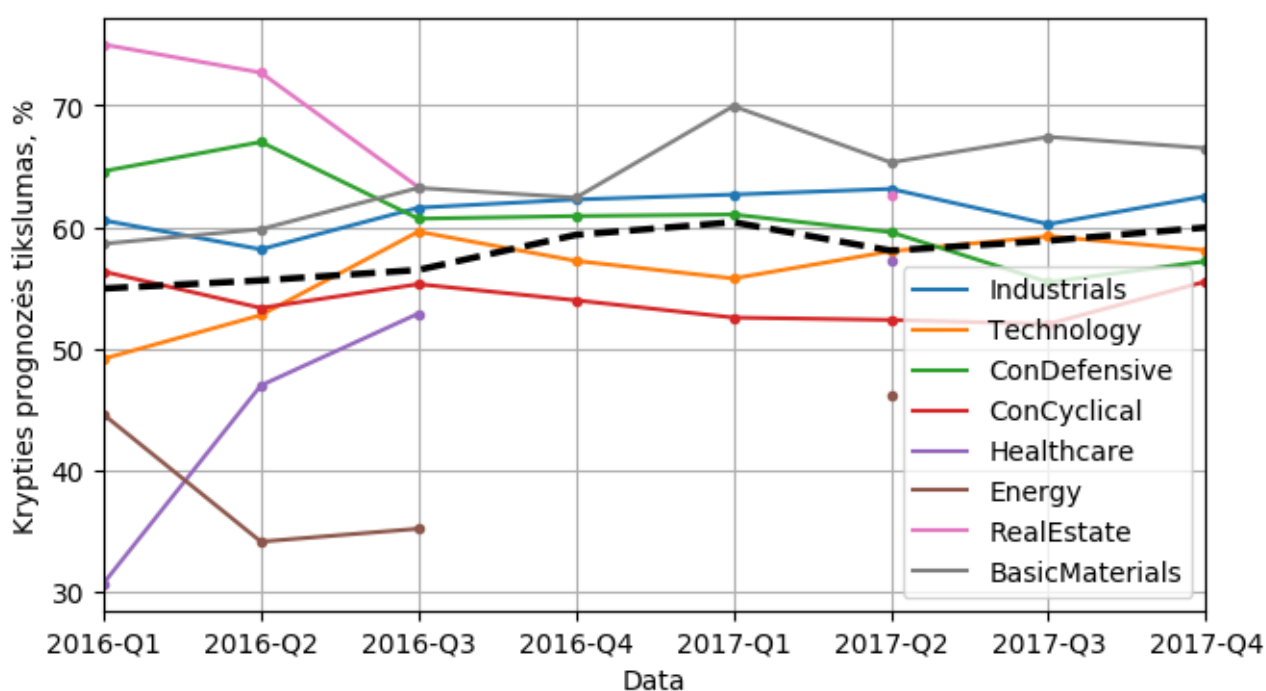
#### 4.4. Gradient Boosting metodas

Modelio patikrinimo metu nustatyti geriausi parametrai pateikti 7 lentelėje.

7 lentelė. Geriausi Gradient Boosting modelio parametrai

Medžių skaičius	150
Maksimalus gylis	2
Nuostolių funkcija	Mažiausias absoliutinis nuokrypis
Mokymosi greitis	0,01

##### 4.4.1. Modelio vertinimas



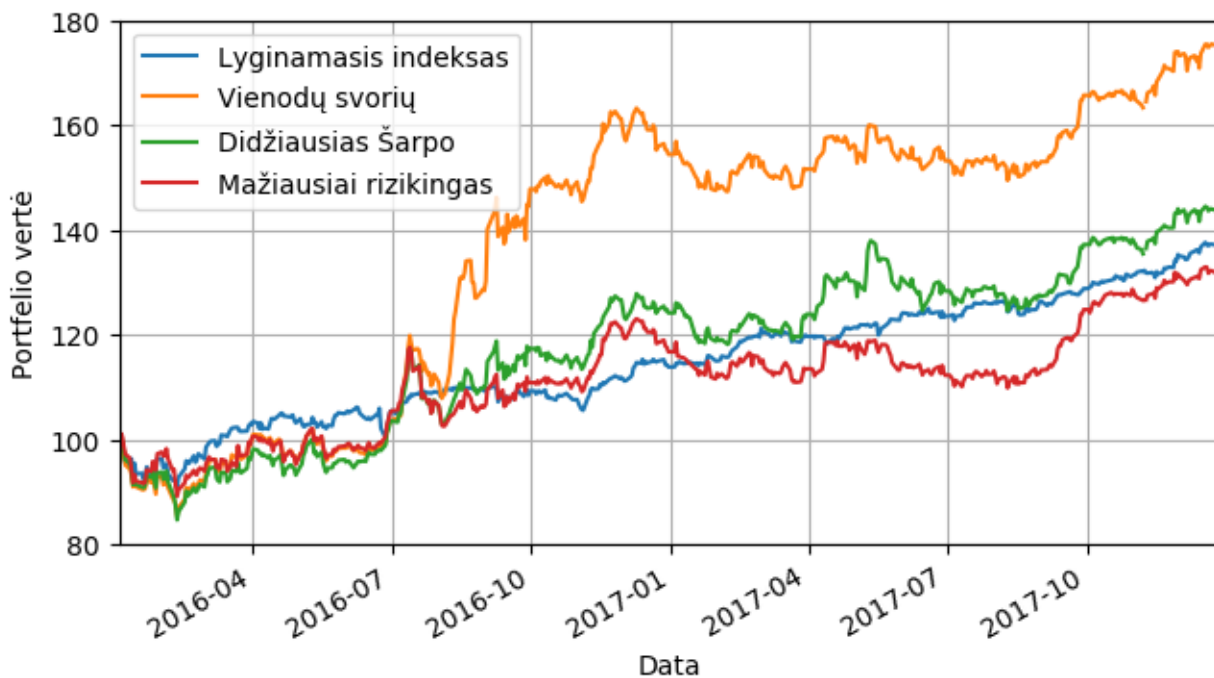
19 pav. Akcijų kainos pokyčių krypties prognozės tikslumas tarp skirtingų sektorių, naudojant Gradient Boosting modelį

8 lentelė. Geriausio Gradient Boosting modelio prognozės tikslumo rodikliai

Krypties prognozės vidutinis tikslumas	59,96 %
Vidutinė absoliutinė paklaida	0,124
Determinacijos koeficientas ( $R^2$ )	0,01
Šaknis iš vidutinės kvadratinės paklaidos	0,2

#### 4.4.2. Portfelijų vertinimas

Investicinių portfelijų, sudarytų iš *Gradient Boosting* modelio atrinktų 16 perspektyviausių akcijų, rezultatai pateikti 20 paveiksle ir 9 lentelėje.

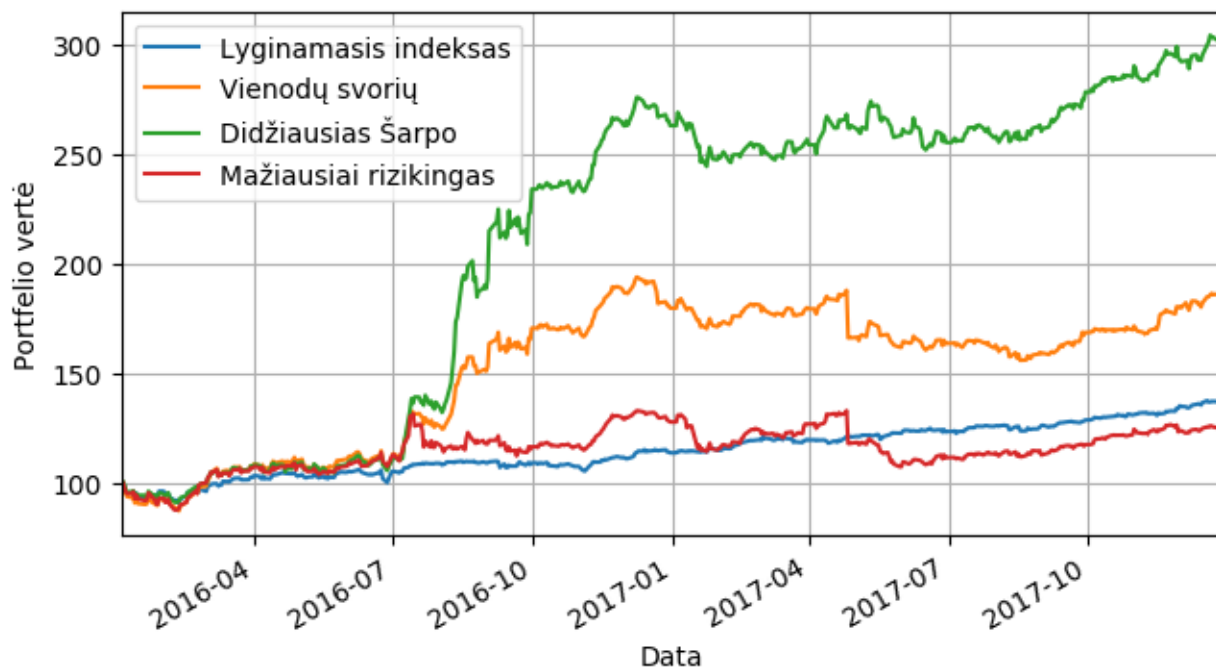


**20 pav.** Portfelijų, sudarytų iš *Gradient Boosting* modelio atrinktų akcijų (16 perspektyviausių), kainų grafikai

**9 lentelė.** Portfelijų, sudarytų iš *Gradient Boosting* modelio atrinktų akcijų (16 perspektyviausių), rezultatai

Vertinimo rodiklis	Vienodi svoriai	Didžiausias Šarpo	Mažiausiai rizikingas
Portfelio vertė $V_{2018}$	175,32	144,14	132,2
Vidutinė metinė grąža	34,37 %	20,45 %	15,41 %
Didžiausias smukimas	-14,96 %	-15,72 %	-12,8 %
Šarpo rodiklis	0,56	0,55	0,36
Sortino rodiklis	2,44	1,67	0,86

Investicinių portfelių, sudarytų iš *Gradient Boosting* modelio atrinktų akcijų (po 2 iš sektoriaus), rezultatai pateikti 21 paveiksle ir 10 lentelėje.



**21 pav.** Portfelių, sudarytų iš *Gradient Boosting* modelio atrinktų akcijų (po 2 iš sektoriaus), kainų grafikai

**10 lentelė.** Portfelių, sudarytų iš *Gradient Boosting* modelio atrinktų akcijų (po 2 iš sektoriaus), rezultatai

Vertinimo rodiklis	Vienodi svoriai	Didžiausias Šarpo	Mažiausiai rizikingas
Portfelio vertė $V_{2018}$	181,95	295,84	124,47
Vidutinė metinė grąža	40,67 %	87,91 %	12,36 %
Didžiausias smukimas	-19,56 %	-14,84 %	-19,22 %
Šarpo rodiklis	0,5	0,56	0,3
Sortino rodiklis	1,6	5,0	0,48

#### 4.5. Rezultatų aptarimas

Akcijų kainos pokyčių krypties prognozės tikslumo grafikuose matyti, jog trūksta kai kurių duomenų iš kelių sektorių. Šių sektorių akcijos tam tikrose laiko eilutėse neturi duomenų, todėl šių laikotarpių prognozės yra praleidžiamos. Tai padidina portfelių nesisteminę riziką, kadangi tais laikotarpiais akcijos buvo atrinktos iš ne visų sektorių.

Sudarant koreliacijas tarp skirtingų portfelių mėnesinės grąžos, gautos stiprios koreliacijos tarp visų portfelių. Tai rodo, jog tiek atsitiktinis miškas, tiek *Gradient Boosting* atrinko panašias potencialiai pelningas akcijas.

**11 lentelė.** Skirtingų modelių sudarytų modelių mėnesinės grąžos (vienodų svorių portfeliai) koreliacijos

Modelis	Atsitiktinis miškas pelningiausias	Atsitiktinis miškas diversifikuotas	<i>Gradient Boosting</i> pelningiausias	<i>Gradient Boosting</i> diversifikuotas
Atsitiktinis miškas pelningiausias	1	0,9570	0,8027	0,8754
Atsitiktinis miškas diversifikuotas	0,9570	1	0,8304	0,9061
<i>Gradient Boosting</i> pelningiausias	0,8027	0,8304	1	0,8214
<i>Gradient Boosting</i> diversifikuotas	0,8754	0,9061	0,8214	1

Iš abiejų modelių atrinktų akcijų sudarius investicinius portfelius, net 9 iš 12 pagal pelningumą sugebėjo aplenkti lyginamąjį indeksą. Tai rodo, jog ekonominio pakilimo metu (tiriamasis laikotarpis) sudaryti pelningą portfelį galima įvairiais metodais.

Sudarius portfelį vien iš potencialiai pelningiausių akcijų, nediversifikuojant tarp sektorių, gautas vidutinė metinė grąža lygi 29,12 %, o vidutinis Šarpo rodiklis – 0,425. Tuo tarpu diversifikuotų portfelių vidutinė metinė grąža lygi net 50,02 %, o vidutinis Šarpo rodiklis – 0,585. Diversifikuojant portfelį galima tikėtis ne tik geresnio portfelio, pagal Moderniąją portfelio teoriją, bet ir tikėtis daug didesnės grąžos.

Nors geriausiai pasirodė portfelis, kurio akcijos pasiskirsčiusios pagal istorinį didžiausią Šarpo rodiklį, tačiau toks metodas pasiskirstymui nustatyti negarantuoja optimalaus portfelio, nes kitų pagal tą patį metodą sudarytų portfelių grąža buvo tik vidutinė arba net mažesnė už lyginamąjį indeksą. Portfelis, sudarytas pagal istorinį mažiausią standartinę nuokrypį, nepadėjo išlaikyti nei mažų svyravimų, nei mažo didžiausio smukimo. Šie rezultatai rodo, jog istoriniai akcijų kainos svyravimai nėra ta priemonė, kurią reikėtų pasitelkti sudarant optimalų portfelį.

Mažas abiejų geriausių modelių medžių skaičius rodo, jog pasitelkus vien fundamentalius rodiklius, akcijos kainos pokyčių prognozuoti yra sudėtinga. Tam tikrą akcijos kainos dalį lemia investuotojų lūkesčiai, kurių pokyčius nuspėti yra labai sunku. Todėl mašininio mokymosi algoritmais paremtos portfelių sudarymo strategijos turėtų būti tik patarimojo pobūdžio.



## Išvados

1. Investicinio portfelio formavimo tyrimui buvo pasinaudota simfin.com svetainės teikiamomis finansinėmis ataskaitomis. Šie duomenys apdoroti taip, kad jie būtų lengvai skaitomi ir tinkami mašininio mokymosi algoritmams. Iš 58 skirtingų finansinių rodiklių atrinkti 10 labiausiai tinkamų fundamentaliai analizei. Atsitiktinio miško kintamųjų reikšmingumo nustatymas parodė, jog kartu su absoliutiniais rodikliais, kaip modelio įėjimus, verta panaudoti ir jų metinius pokyčius. Kaip modelio išėjimas pasirinktas procentinis akcijos kainos pokytis per ateinantį ketvirtį.
2. Akcijų atranka atlikta pasitelkiant atsitiktinio miško ir *Gradient Boosting* metodus. Duomenų imtis padalinta į du laikotarpius: modelių mokymosi ir testavimo. Mokymosi duomenų imtis panaudota ir modelio parametrų derinimui kryžminio patikrinimo būdu. Pagal abu metodus sukurti modeliai geriausią determinacijos koeficientą pasiekia, kai medžių skaičius yra 150. Didesnis medžių skaičius, nors ir neprastina rezultato, tačiau ilgina modelio mokymosi trukmę, todėl nėra racionalu naudoti daugiau nei 150 medžių.
3. Tiek atsitiktinio miško, tiek *Gradient Boosting* metodų gautos absoliutinės paklaidos (atitinkamai 13 % ir 12,4 %) yra daugiau nei 7 procentiniais punktais mažesnės už primityvaus modelio absoliutinę paklaidą (20,3 %). Tačiau modelių parametrų derinimo metu gauti geriausi -0,52 (atsitiktinio miško) ir 0,01 (*Gradient Boosting*) determinacijos koeficientai atspindi, jog modelio įėjimui naudojami duomenys turi mažai informacijos apie tai, koks bus modelio išėjimas. Atsižvelgus į didžiausius smukimus, rizikingesni yra atsitiktinio miško metodu sudaryti portfeliai.
4. Pagal sektorius diversifikuoti portfeliai turi ne tik geresnį Šarpo rodiklį, bet ir sugeba uždirbti didesnę pelną.
5. Iš gautų kainų grafikų matyti, jog tais pačiais akcijų pasiskirstymo metodais sudarytų portfelių rezultatai yra labai nepastovūs. Todėl naudoti vien istorinius kainų duomenis portfelio optimizavimui yra nerekomenduotina.
6. Portfelis, kurio akcijos yra atrenkamos iš skirtingų sektorių pagal *Gradient Boosting* metodą ir paskirstomos pagal didžiausią istorinį Šarpo rodiklį, pasiekė didžiausią vidutinę metinę grąžą – 87,91 % ir geriausią Sortino rodiklį – 5.

## Literatūros sąrašas

1. Vito J. Racanelli. The U.S. Stock Market Is Now Worth \$30 Trillion. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.barrons.com/articles/the-u-s-stock-market-is-now-worth-30-trillion-1516285704>
2. XingYu Fu, JinHong Du, YiFeng Guo, MingWen Liu, Tao Dong, XiuWen Duan. A Machine Learning Framework for Stock Selection. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://arxiv.org/pdf/1806.01743.pdf>
3. Bob French. How Are You Different From The Average Investor? (2018) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://retirementresearcher.com/how-are-you-different-from-the-average-investor/>
4. SEB. Investavimo menas. (2010) [žiūrėta 2019-05-07]. Prieiga per internetą: [https://www.seb.lt/sites/default/files/web/pdf/Investavimo\\_menas.pdf](https://www.seb.lt/sites/default/files/web/pdf/Investavimo_menas.pdf)
5. Veliota Drakopoulou. A Review of Fundamental and Technical Stock Analysis Techniques. (2015) [žiūrėta 2019-05-07]. Prieiga per internetą: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3204667](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3204667)
6. Vertybinių popierių birža NASDAQ OMX Vilnius. Įmonių finansinė analizė. Rodiklių skaičiavimo metodika. (2010) [žiūrėta 2019-05-07]. Prieiga per internetą: [https://www.nasdaqbaltic.com/files/vilnius/leidiniai/Rodikliu\\_skaiciavimo\\_metodika-final.pdf](https://www.nasdaqbaltic.com/files/vilnius/leidiniai/Rodikliu_skaiciavimo_metodika-final.pdf)
7. Investologija.lt. Investavimo rizika ir jos valdymas. (2016) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://investologija.lt/investavimas/kur-investuoti/investavimo-rizika-ir-jos-valdymas/?pdf=385>
8. Nauris Treigys. S&P 500 indeksas. Kaip į jį investuoti? (2019) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.aipt.lt/sp-500-indeksas/>
9. 8 Risk (Adjusted) Statistics Summarized. (2016) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.managedfuturesinvesting.com/8-risk-adjusted-statistics-summarized/>
10. Rimvydas Simutis. Valdymo metodai finansiniuose procesuose. E-konspektai. (2012)
11. John Alberg, Zachary C. Lipton. Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://arxiv.org/abs/1711.04837>
12. Crescenzo Gallo. Artificial Neural Networks in Finance Modelling. (2005) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.researchgate.net/publication/23743353>
13. Colah's blog. Understanding LSTM Networks. (2015) [žiūrėta 2019-05-07]. Prieiga per internetą: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
14. François Chollet. Deep Learning with Python. (2017)
15. Lean Yu, Shouyang Wang, Kin Keung Lai. Portfolio Optimization Using Evolutionary Algorithms. (2008) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.researchgate.net/publication/288554014>
16. Akash Shrivastava, Anugrah Singh. An Optimal Stock Portfolio Construction model using Genetic Algorithm. (2014) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://www.researchgate.net/publication/287564345>
17. Vadim Smolyakov. Ensemble Learning to Improve Machine Learning Results. (2017) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://blog.statsbot.co/ensemble-learning-d1dcd548e936>
18. Antanas Verikas, Adas Gelžinis, Marija Bačauskienė. Mining data with random forests: A survey and results of new tests. (2010) [žiūrėta 2019-05-07]. Prieiga per internetą:

- <https://www.researchgate.net/publication/222283704>
19. Krishni Hewa. A Beginners Guide to Random Forest Regression. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<https://medium.com/datadriveninvestor/random-forest-regression-9871bc9a25eb>
  20. Prince Grover. Gradient Boosting from scratch. (2017) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d>
  21. Harshdeep Singh. Understanding Gradient Boosting Machines. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>
  22. Andre Christoffer Andersen, Stian Mikelsen. A Novel Algorithmic Trading Framework Applying Evolution and Machine Learning for Portfolio Optimization. (2012) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<http://www.datascienceassn.org/sites/default/files/A%20Novel%20Algorithmic%20Trading%20Framework%20-%20Machine%20Learning%20for%20Portolio%20Optimization.pdf>
  23. Will Koehrsen. Hyperparameter Tuning the Random Forest in Python. (2018) [žiūrėta 2019-05-07]. Prieiga per internetą: <https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>
  24. Aarshay Jain. Complete Guide to Parameter Tuning in Gradient Boosting (GBM) in Python. (2016) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/>
  25. Vydas Čekanavičius. Taikomoji regresinė analizė socialiniuose tyrimuose. (2014) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<http://www.statistika.mif.vu.lt/wp-content/uploads/2014/04/regresine-analize.pdf>
  26. Keat-Yung Ng, Kim-Leng Goh. Effects of Size and Allocation Method on Stock Portfolio Performance: A Simulation Study. (2011) [žiūrėta 2019-05-07]. Prieiga per internetą:  
<http://www.ipedr.com/vol12/102-N10005.pdf>