



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

# **Didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose**

Baigiamasis magistro projektas

---

**Aurelija Kiltinavičiūtė**  
Projekto autorė

**Doc. dr. Audrius Kabašinskas**  
Vadovas

**Prof. dr. Lina Dagilienė**  
Vadovė

---

**Kaunas, 2018**



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

# **Didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose**

Baigiamasis magistro projektas  
Didžiųjų verslo duomenų analitika (621G12002)

---

**Aurelija Kiltinavičiūtė**  
Projekto autorė

**Doc. dr. Audrius Kabašinskas**  
Vadovas

**Doc. dr. Lina Dagilienė**  
Vadovė

**Prof. dr. Borisas Seminogovas**  
Recenzentas

**Dr. Tomas Iešmantas**  
Recenzentas

---

**Kaunas, 2018**



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas  
Aurelija Kiltinavičiūtė

## **Didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Aurelijos Kiltinavičiūtės, baigiamasis projektas tema „Didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

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

---

(vardą ir pavardę įrašyti ranka)

---

(parašas)

## Turinys

<b>Įvadas</b> .....	<b>10</b>
<b>1. Didieji verslo duomenys ir jų taikymas audite</b> .....	<b>12</b>
1.1. Didieji verslo duomenys .....	12
1.2. Audito rūšys ir tikslai .....	13
1.3. Audito pažanga didžiųjų duomenų bei analizės srityje.....	14
1.4. Priežastys stabdančios didžiųjų duomenų analitikos naudojimą audito procedūrose .....	17
1.5. Didžiųjų duomenų analitikos taikymas finansinių ataskaitų audito procedūrose .....	20
1.5.1. Didžiųjų duomenų analitikos teikiamos galimybės ir privalumai .....	21
1.5.2. Audito etapai ir juose taikomos analitinės procedūros .....	23
<b>2. Tyrimo metodai</b> .....	<b>29</b>
2.1. Tyrimo teorinė schema.....	29
2.2. Didžiųjų duomenų apdorojimas ir tyryba .....	30
2.3. Teksto tyryba.....	32
2.4. Teksto klasifikavimas.....	35
2.4.1. Naivaus Bajeso algoritmas .....	35
2.4.2. Atraminių vektorių algoritmas.....	36
2.5. Laiko eilučių analizė .....	37
2.5.1. Vienmačiai laiko eilučių prognozavimo metodai .....	41
<b>3. Tiriamoji dalis</b> .....	<b>46</b>
3.1. Tiriamieji objektai .....	46
3.2. Laiko eilučių analizė .....	47
3.3. Veiksnių darančių poveikį bendrajam pelningumui nustatymas .....	52
3.4. Finansinių rodiklių prognozavimas.....	55
3.5. Bendrojo žurnalo įrašų teksto tyryba .....	58
3.5.1. Bendrojo žurnalo įrašų klasifikavimas pagal operacijos aprašymą.....	65
<b>Išvados</b> .....	<b>68</b>
<b>Literatūros sąrašas</b> .....	<b>70</b>
<b>Priedai</b> .....	<b>76</b>

## Paveikslų sąrašas

pav. 1 Auditas pagal apimtį ir atlikėjus .....	13
pav. 2 Audito tikslai .....	17
pav. 3 Audito etapai pagal Cushingą ir Loebbecke (1986) .....	23
pav. 4 tyrimo teorinė schema.....	29
pav. 5. Didžiųjų duomenų apdorojimo schema [34] .....	30
pav. 6 Duomenų tyrybos uždaviniai pagal Weiss S.M. ir Indurkhya N. [40, (8psl)] .....	31
pav. 7 hiperplokštuma ir atraminiai vektoriai pagal [56] .....	37
pav. 8 UAB Duona ir UAB Sausainis pajamų kaita pamėnesiui.....	48
pav. 9 Mėnesinių pajamų palyginimas 2012 m.- 2017 m. laikotarpiu .....	49
pav. 10 <i>Box plot</i> procedūros grafikas .....	50
pav. 11 <i>UAB Duona</i> pardavimo pajamų laiko eilutės dekompozicija.....	50
pav. 12 <i>UAB Duona acf</i> ir <i>pacf</i> procedūrų grafikai .....	51
pav. 13 <i>UAB Duona random</i> komponentės <i>acf</i> grafikas .....	51
pav. 14 Koreliacinio ryšio stiprumo įvertinimo diagrama, <i>UAB Duona</i> atvejis.....	52
pav. 15 <i>Grangerio</i> testo rezultatai tarp bendrojo pelningumo ir vidutinio darbo užmokesčio .....	54
pav. 16 impulsų atsako grafikas .....	54
pav. 17 <i>FEVD</i> grafikas tarp BP ir DU, <i>UAB Duona</i> atvejis.....	55
pav. 18 <i>UAB Duona</i> pardavimo pajamų prognozė.....	56
pav. 19 <i>UAB Sausainis</i> pardavimo pajamų prognozė.....	58
pav. 20 <i>UAB Duona</i> žodžių debesis, 50 dažniausių žodžių.....	61
pav. 21 <i>UAB Sausainis</i> žodžių debesis, 50 dažniausių žodžių.....	64
pav. 22 savikainos kaita pamėnesiui.....	76
pav. 23 <i>UAB Sausainis</i> pajamų laiko eilutės dekompozicija.....	77
pav. 24 <i>UAB Sausainis acf</i> ir <i>pacf</i> procedūrų grafikai .....	77
pav. 25 <i>UAB Sausainis random</i> komponentės <i>ACF</i> grafikas .....	78
pav. 26 Koreliacinio ryšio stiprumo įvertinimo diagrama, <i>UAB Sausainis</i> atvejis .....	78
pav. 27 <i>Grangerio</i> testo rezultatai tarp bendrojo pelningumo ir sviesto kainos.....	79
pav. 28 <i>Grangerio</i> testo rezultatai tarp bendrojo pelningumo ir cukraus kainos .....	80
pav. 29 <i>Grangerio</i> testo rezultatai tarp bendrojo pelningumo ir kviečių kainos.....	80
pav. 30 <i>UAB Duona</i> bendrojo pelningumo prognozė.....	84
pav. 31 <i>UAB Sausainis</i> bendrojo pelningumo prognozė .....	85
pav. 32 <i>UAB Duona</i> 50 dažniausių išlaidų bendrojo žurnalo žodžių .....	86
pav. 33 <i>UAB Sausainis</i> 50 dažniausių išlaidų bendrojo žurnalo žodžių.....	87

## Lentelių sąrašas

lentelė 1. Galimos duomenų tyrybos technikos susipažinimo su klientu etape.....	24
lentelė 2. Galimos duomenų tyrybos technikos planavimo etape .....	25
lentelė 3. Galimos duomenų tyrybos technikos pagrindinių procedūrų ir testavimo etape.....	26
lentelė 4. Galimos duomenų tyrybos technikos vertinimo ir peržiūros etape .....	27
lentelė 5. Galimos duomenų tyrybos technikos audito užbaigimo etape .....	28
lentelė 6 tyrime naudojamų didžiųjų duomenų technikų teorinė schema .....	29
lentelė 7. Nagrinėjamų įmonių finansiniai duomenys 2017.12.31 datai .....	46
lentelė 8. Pardavimo pajamų ir savikainos metinių duomenų analizė.....	47
lentelė 9. Pardavimo pajamų ir savikainos mėnesinių duomenų analizė .....	48
lentelė 10 koreliacinės analizės ir <i>Grangerio</i> testo rezultatai, <i>UAB Duona</i> atvejis .....	53
lentelė 11 <i>UAB Duona</i> pajamų prognozės klaidų įverčiai.....	56
lentelė 12 <i>UAB Sausainis</i> pajamų prognozės klaidų įverčiai .....	57
lentelė 13 Klasifikavimo tikslumas, <i>UAB Duona</i> atvejis .....	65
lentelė 14 <i>UAB Duona</i> sumaišymų matrica.....	66
lentelė 15 Klasifikavimo tikslumas, <i>UAB Sausainis</i> atvejis.....	67
lentelė 16 <i>UAB Sausainis</i> sumaišymų matrica .....	67
lentelė 17 koreliacinės analizės ir <i>Grangerio</i> testo rezultatai, <i>UAB Sausainis</i> atvejis .....	79
lentelė 18 <i>UAB Duona</i> prдавimo pajamų prognozė 2018 metams.....	81
lentelė 19 <i>UAB Duona</i> prдавimo pajamų prognozės pasikliautiniai intervalai.....	82
lentelė 20 <i>UAB Sausainis</i> prдавimo pajamų prognozė 2018 metams .....	82
lentelė 21 <i>UAB Sausainis</i> prдавimo pajamų prognozės pasikliautiniai intervalai .....	83
lentelė 22 <i>UAB Duona</i> bendrojo pelningumo prognozės klaidų įverčiai .....	83
lentelė 23 <i>UAB Duona</i> bendrojo pelningumo prognozė.....	84
lentelė 24 <i>UAB Sausainis</i> bendrojo pelningumo prognozės klaidų įverčiai.....	85
lentelė 25 <i>UAB Sausainis</i> bendrojo pelningumo prognozė .....	86

Kiltinavičiūtė, Aurelija. Didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose. Magistro baigiamasis projektas vadovai doc. dr. Audrius Kabašinskas ir prof. dr. Lina Dagilienė; Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų kryptių grupė): fiziniai mokslai, matematika (01 P).

Reikšminiai žodžiai: didieji duomenys, analitiniai metodai, finansinių ataskaitų auditas.

Kaunas, 2018. 87 p.

## Santrauka

Didžiųjų duomenų analitikos teikiamais privalumais jau naudojamosi daugybėje sričių, tuo tarpu audito procedūros praktiškai nekinta jau 50-60 metų [7]. Nuomonė, kad didžiųjų duomenų analitika audite yra ne taip plačiai paplitusi kaip kitose srityse vyrauja didžiojoje dalyje šaltinių nagrinėjančių šią tematiką [1, 6, 8, 9, 10]. Duomenų analizė pabrėžiama, kaip veiksminga audito užduočių priemonė, tačiau auditorių supratimas ir duomenų analizės naudojimas audito užduočių srityje vis dar yra ribotas. Audito srities atstovai taip pat susiduria su daugybe neatsakytų klausimų galvojant apie galimas didžiųjų duomenų analitikos naudojimo pasekmes [10].

Šiame darbe siekiama pritaikyti didžiųjų duomenų analitikos metodus finansinių ataskaitų audito procedūrose, kai audituojamas subjektas yra Lietuvoje veikianti uždarojo tipo akcinė bendrovė. Tyrimui buvo pasirinktas 2012 – 2017 metų laikotarpis. Tiriamieji subjektai – dvi gamyba užsiimančios įmonės.

Atlikus tyrimą nustatyta, kad didžiųjų duomenų analitikos metodų taikymas finansinių ataskaitų audito procedūrose gali padidinti audito aprėptį, pasitelkiant teksto analizę galima analizuoti, ne dalį, bet 100 % bendrojo žurnalo įrašų. Tyrimo rezultatai rodo, kad teksto klasifikavimas gali sumažinti tikrinamą imtį net 99 %, taip sumažinant auditui skirtą laiką. Taikant įvairius prognozavimo metodus galima įvertinti riziką susijusią su veiklos tęstinumo prielaida. Nustatyta, kad prognozuojant tiriamų įmonių pardavimų pajamas geriausius paklaidų įverčius gavome taikydami *Holt-Winters* ir *BATS* modelius. Atlikus *Grangerio* priežastingumo testą tarp įmonių bendrojo pelningumo ir keturių išorės rodiklių nustatyta, kad iš tiriamų išorės rodiklių tik vidutinis darbo užmokestis turi priežastinį ryšį su *UAB Duona* bendroju pelningumu, tačiau bendrasis pelningumas yra ir priežastis ir pasekmė. Taip pat nustatyta, kad *UAB Sausainis* bendrąjį pelningumą sieja priežastinis ryšys su sviesto, cukraus bei kviečių kainomis, tačiau visais atvejais bendrasis pelningumas yra priežastis.

Kiltinavičiūtė, Aurelija. Application of Big data in procedures of financial statements audit. Master's Final Degree Project supervisors Assoc. Prof. Dr. Audrius Kabašinskas and Prof. Dr. Lina Dagilienė; The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): natural sciences, mathematics (01 P).

Keywords: big data, analytical methods, financial statements audit.

Kaunas, 2018. 87 pages.

### Summary

The benefits of big data analytics are already being used in many areas, while audit procedures are practically unchanged about 50-60 years [7]. Opinion that big data analytics is not as widespread in audit field as in other areas predominates in most of the sources exploring this topic [1, 6, 8, 9, 10]. Although data analytics are believed to improve audit effectiveness, understanding of using data analytics in audit engagements is still limited. Audit field representatives also faces many unanswered questions about the potential consequences of using big data analytics [10].

The aim of this work is to adapt big data analytics methods in procedures of financial statements audit, when the audited entity is a closed joint stock company operating in Lithuania. Years 2012-2017 investigation period was chosen for the study. Empirical research is based on two manufacturing companies.

The investigation found that the application of big data analytics methods in the financial statements audit procedures may increase audit scope, with the help of the text mining, it is possible to analyze not part, but 100 % of the journal entries. The research results indicate that the text classification can reduce the sample size that being examined by 99 %, thus reducing the time spent on the audit. The application of various forecasting methods can measure the risks associated with the going concern assumption. It was found that in forecasting the sales revenue of the investigated companies, we obtained the best error estimates using *Holt-Winters* and *BATS* models. After the *Granger* causality test among the companies gross profit margin and four external indicators, it was found that only the average wages in the country have a causal relationship with the gross profit margin of *UAB Duona*, but the gross profit is both a cause and a consequence. It was also found that *UAB Sausainis* gross profit margin has a causal relationship to the butter, sugar and wheat prices, but in all cases the gross profit margin is a cause.



## Santrumpos

- AAA (angl. American Accounting Association) - Amerikos apskaitos asociacija;
- AHP (angl. Analytic Heirarchy Process) - analitinės hierarchijos procesas;
- ANN (angl. Artificial Neural Networks) - dirbtiniai neuroniniai tinklai;
- ARIMA (angl. Autoregressive integrated moving average) - autoregresyvus integruotas slenkantis vidurkis;
- BBN (angl. Bayesian Belief Networks) – Bajeso tinklai;
- CAATS (angl. Computer Assisted Audit Techniques) - kompiuterizuoti audito metodai;
- DAWG (angl. Data Analytics Working Group) - tarptautinių audito ir užtikrinimo užduočių standartų valdybos įsteigta duomenų analitikos darbo grupė;
- Didysis ketvertas (angl. Big4) – keturios didžiosios tarptautinės audito ir konsultacijų bendrovės;
- DK – didžioji knyga, buhalterinėje apskaitoje naudojamas dokumentas, kuriame surenkami duomenys iš registrų ir apskaičiuojami sąskaitų likučiai.;
- EY – viena iš didžiojo ketvirto (angl. Big4) audito kompanija Ernst and Young;
- IAASB (angl. International Auditing and Assurance Standards Board) – tarptautinių audito ir užtikrinimo užduočių standartų valdyba;
- IFAC (angl. International Federation of Accountants) – tarptautinė buhalterijų federacija;
- KPMG – vienas iš didžiojo ketvirto (angl. Big4) audito kompanija;
- PwC - viena iš didžiojo ketvirto (angl. Big4) audito kompanija PricewaterhouseCoopers;
- SVM (angl. Support Vector Machines) – atraminių vektorių metodas;
- TAS (angl. International Standards on Auditing) – tarptautiniai audito standartai.

## Įvadas

Šiandienos verslo aplinka pasižymi kompleksišku, pastovia kaita, bei neapibrėžtumu. Už valdymą atsakingi asmenys susiduria su daugiau iššūkių nei bet kada anksčiau siekiant sukurti korporacinę kultūrą su patikimu rizikos valdymu ir kontrole. Norėdamos neatsilikti nuo vis sudėtingesne tampančios įmonių bei rizikos valdymo aplinkos audito įmonės turi galvoti apie technologijas galinčias atvesti pokyčius į audito procedūras [6].

Audito procedūros praktiškai nekinta jau 50-60 metų [7]. Dėl informacinių technologijų pažangos, didėjančio duomenų bei informacijos vartotojų kiekio audito srityje vis garsiau kalbama apie didžiųjų duomenų analizės naudojimą audito procese [1]. Informacija bei įžvalgos, kurios gali būti naudingos įmonei, išplito toli už tradicinių finansinių operacijų didžiojoje knygoje. Tad natūralu, kad keičiantis informacijos pobūdžiui bei apimčiai turi keistis ir audito procedūros.

Vyrauja nuomonė, kad didžiųjų duomenų metodai finansinių ataskaitų audito procese gali būti naudojami siekiant padidinti audito aprėptį, ištirti vidaus kontrolės veiksmingumą, sandorių tikslumą, identifikuoti rizikas, koreliacijas, svyravimus, išskirtis, prognozuoti finansines nesėkmes, taip pagerinant analitines procedūras visose audito etapuose ir tuo pačiu gerokai sumažinant audito išlaidas, išteklius ir laiką [15]. Didžiųjų duomenų naudojimas auditorių veikloje padėtų perkelti profesiją į vertės grandinę ir tapti tikrais verslo partneriais, o ne tik sandorio paslaugų teikėjais [8].

Galima rasti nemažai straipsnių aprašančių, kokius didžiųjų duomenų analitikos metodus būtų galima naudoti audite, tačiau jų praktinis taikymas yra tyrinėtas gana mažai. Iš nagrinėtos literatūros tik keletas straipsnių parodo, kaip šiuos metodus realizuoti praktiškai [25, 32], be to dauguma šių realizacijų pritaikyti banko arba draudimo sektoriaus audito procedūroms. Pastebima, kad literatūros šaltiniuose daugiausia kalbama apie dideles, viešo intereso įmones, o Lietuvos audituojamų įmonių rinka pakankamai siaura, jau nekalbant apie kompanijas, kurių akcijomis būtų prekiaujama biržoje, kurių yra labai mažai. Dėl šios priežasties aktualu ištirti, kaip pritaikyti didžiųjų duomenų analitikos metodus finansinių ataskaitų audito procedūrose, kai audituojama įmonė nėra viešo intereso, kadangi šis tyrimas gali būti aktualus visoms audito įmonėms audituojančioms uždarojo tipo akcinės bendrovės.

**Darbo tema:** didžiųjų verslo duomenų taikymas finansinių ataskaitų audito procedūrose.

**Darbo tikslas:** teoriškai pagrįsti, kokie didžiųjų duomenų analitikos metodai taikomi finansinių ataskaitų audito etapuose bei pasirinktus metodus pritaikyti empiriškai Lietuvos uždarojo tipo akcinių bendrovių atveju, taip siekiant pagerinti audito procedūrų efektyvumą, našumą bei kokybę.

**Darbo uždaviniai:**

1. išnagrinėti mokslinės bei metodinės literatūros šaltinius ir nustatyti finansinių ataskaitų audito procedūrose rekomenduojamus taikyti didžiųjų duomenų metodus;
2. sudaryti teorinę tyrimo schemą ir aprašyti tyrime naudojamus metodus;
3. nustatyti ar egzistuoja priežastinis ryšys tarp nagrinėjamų įmonių bendrojo pelningumo ir pasirinktų išorės rodiklių;
4. atlikti įmonių pardavimo pajamų laiko eilučių analizę bei jų prognozę, siekiant paremti veiklos tęstinumo prielaidą;
5. atlikti įmonių bendrojo žurnalo įrašų tekstinės informacijos analizę bei klasifikavimą, taikant SVM ir Naivaus Bajeso metodus.
6. apibendrinti tyrimo rezultatus ir pateikti rekomendacijas.

## 1. Didieji verslo duomenys ir jų taikymas audite

Šiame skyriuje bus nagrinėjama didžiųjų duomenų analitikos taikymo audite tema. Supažindinama su pagrindinėmis sąvokomis bei idėjomis. Pateikiama nagrinėtoje literatūroje atliktų tyrimų, tiesiogiai susijusių su nagrinėjama tema, rezultatų apžvalga bei analizė.

### 1.1. Didieji verslo duomenys

Sunku paneigti, kad mes gyvename didžiųjų duomenų (angl. Big data) eroje. Didžiulis duomenų kiekis iš įvairių informacijos šaltinių bei vis augantis jų atsiradimo tempas pritraukė ne tik informacinių technologijų ekspertų, bet ir daugybės organizacijų dėmesį. Istoriskai duomenys buvo kažkas, kas priklauso mums – struktūrizuoti ir žmogaus sukurti. Tačiau technologijų tendencijos išplėtė šį apibrėžimą ir įtraukė duomenis, kurie yra nestruktūrizuoti, generuojami mašinų bei gyvenantys už verslo ribų [1]. Informacija bei išvalgos, kurios gali būti naudingos įmonei, išplito toli už tradicinių finansinių operacijų didžiojoje knygoje. Šiandien duomenys yra gaunami praktiškai kiekviename žingsnyje iš internetinių tinklalapių, mobiliųjų prietaisų, automobilių, elektroninių laiškų, žiniasklaidos, vaizdo, balso įrašų ar tiesiog dokumentų. Klaidinga galvoti, kad terminas didieji duomenys susijęs tik su dideliu duomenų kiekiu, didieji duomenys taip pat reiškia ir didelio duomenų kiekio surinkimą, apdorojimą bei analizę, kadangi vien tik duomenų kaupimas ar milžiniško jų kiekio turėjimas žinių nesuteikia – reikia sugebėti šiuos duomenis paversti pridėtinę vertę kuriančia informacija [2].

Atsiradus terminui didieji duomenys, jis anglų kalba buvo apibrėžiamas trimis „v“ raidėmis [3]:

- *volume* – apimtys, t.y. didelis kintamųjų, stebėjimų bei duomenų rinkinių skaičius;
- *velocity* – sparta, t.y. didelė duomenų atsiradimo sparta bei dažnis;
- *variety* – įvairovė, nusako kad duomenų gali būti pačių įvairiausių tipų. Jie gali būti struktūrizuoti, nestruktūrizuoti, skaitiniai, tekstiniai, vaizdiniai, susidaryti iš datų ar bet kokio kito formato.

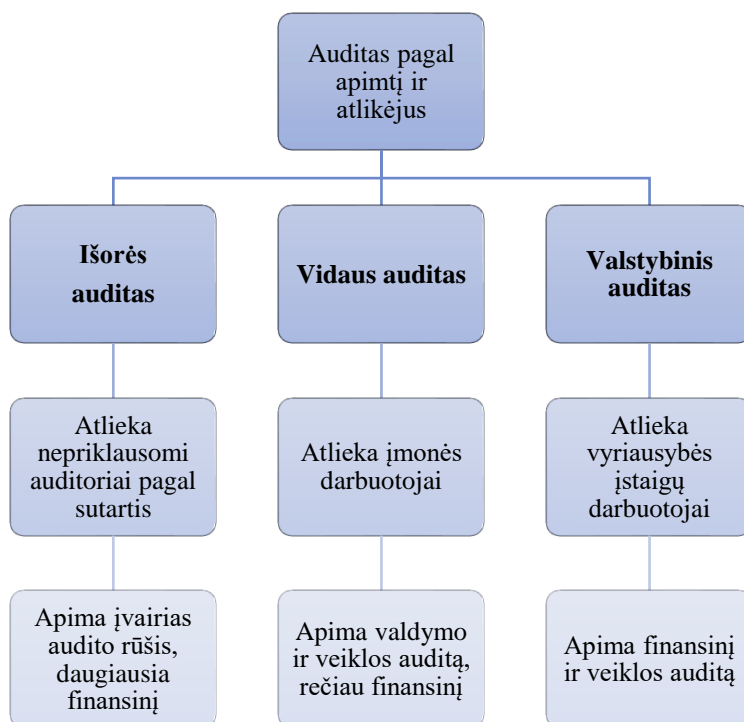
Vėliau į apibrėžimą buvo įtrauktos dar dvi „v“: *value* – vertė, kuri signalizuoja, kad duomenyse slypi potencialiai naudingos žinios, bei *veracity* – tikslumas, sufleruojantis, kad duomenų šaltiniuose gali būti ir neteisingos informacijos. Šiandieną šių „v“ skaičius yra gerokai išaugęs ir pranašaujama, kad jis dar augs [3].

Didžiųjų duomenų analitika (angl. Big data analytics) apibrėžiama kaip duomenų nagrinėjimo, valymo, transformavimo bei modeliavimo procesas, kuriuo siekiama išgauti naudingą informaciją, struktūras, modelius bei pasiūlyti vertingas išvadas ar paremti sprendimų priėmimą [3]. Ji prisideda

prie organizacijų sėkmės, nes suteikia tvirtą pagrindą efektyvumui ir našumui didinti. Didžiųjų duomenų analizės teikiamais privalumais jau naudojamosi daugybėje sričių, tokių kaip rinkodara, logistika, pardavimai, bankininkystė, gamyba [3]. Tai, kad pažangioji analitika (angl. advanced analytics) gali būti naudinga, neišvengimai pastebėjo ir audito srities profesionalai [1].

## 1.2. Audito rūšys ir tikslai

Sudėtingėjant verslo sąlygoms, intensyvėjant konkurencijai, išsami informacija apie subjekto veiklą, jos kokybę ir finansinę būklę tampa vis reikšmingesnė. Pateikta teisinga ir visapusiška informacija, padeda priimti veiksmingesnius sprendimus, todėl vis didesnę vaidmenį verslo ir viešojo sektoriaus organizacijų veikloje įgyja auditas. Auditas apibrėžiamas kaip procesas, kuriuo siekiama sumažinti informacinę vartotojų riziką. Šiandien auditas pagal atlikėjus ir audito apimtį yra skirstomas į tris pagrindines rūšis: išorinį, vidinį auditą bei vyriausybės arba valstybės auditą (žr. 1 pav.) [4].



pav. 1 Auditas pagal apimtį ir atlikėjus

Toliau šiame darbe bus nagrinėjamas didžiųjų duomenų taikymas išorės, tiksliau, finansinių ataskaitų audite.

Išorės auditas – tai auditas, atliekamas organizacijos, nesančios audituojamo subjekto dalimi ir nuo jos nepriklausomas. Šio audito tikslas – nustatyti, ar finansinės ataskaitos visais reikšmingais atvejais tikrai ir teisingai parodo audituojamos įmonės finansinę būklę, veiklos rezultatus ir pinigų srautus pagal taikomus finansinės atskaitomybės reikalavimus [5].

Norėdamas atlikti šią užduotį, auditorius turi suplanuoti ir atlikti tokias audito procedūras, kad gautų pakankamai tinkamų įrodymų, be to, audito standartai reikalauja, kad auditoriai surinktų fizinius įrodymus, kaip rizikos vertinimo proceso dalį. Atlikęs auditą, auditorius pareiškia nuomonę apie finansines ataskaitas. Ši nuomonė dar vadinama nepriklausomo auditoriaus išvada.

Finansinių ataskaitų auditą reglamentuoja LR Finansinių ataskaitų audito įstatymas, Europos Parlamento ir Tarybos reglamentas, Tarptautiniai audito standartai ir Buhalterijų profesionalų etikos kodeksas [5].

### **1.3. Audito pažanga didžiųjų duomenų bei analizės srityje**

Šiandienos verslo aplinka pasižymi kompleksiskumu, pastovia kaita, bei neapibrėžtumu. Už valdymą atsakingi asmenys susiduria su daugiau iššūkių nei bet kada anksčiau, siekdami sukurti korporacinę kultūrą su patikimu rizikos valdymu ir kontrole. Norėdamos neatsilikti nuo vis sudėtingesne tampačios įmonių bei rizikos valdymo aplinkos, audito įmonės turi galvoti apie technologijas, galinčias atvesti pokyčius į audito procedūras [6]. Informacinių technologijų pažanga bei nuolatos didėjantis duomenų srautas bei greitis palengvina, bet kartu ir apsunkina audito procedūras, kadangi tai sukuria daug įvairių naujų pavojų – sunkina reguliavo aplinką, daro ją kompleksine. Tad natūralu, kad keičiantis informacijos pobūdžiui bei apimčiai, turi keistis ir audito procedūros. Tačiau šis procesas vyksta ne taip sparčiai kaip norėtusi, kadangi pereiti nuo tradicinių audito metodų prie integruoto didžiųjų duomenų ir analitikos procedūrų naudojimo yra didžiulis šuolis griežtais standartais reglamentuotos profesijos atstovams. Puikus pagrindas nagrinėjamos temos aktualumui atskleisti yra audito įmonės *Grant Thornton* vadovaujančio partnerio D. Baskino išsakyta mintis: „Audito įrankių rinkinys atrodo taip pat, kaip jis atrodė prieš 50 ar 60 metų. Jeigu mes būtume gydytojai, tai būtų gana bauginanti mintis“ [7]. Nuomonė, kad didžiųjų duomenų analitika audite yra ne taip plačiai paplitusi kaip kitose srityse vyrauja didžiojoje dalyje šaltinių, nagrinėjančių šią tematiką [1, 6, 8, 9, 10].

Dar 2012 metais vienos pirmaujančių mokslinių tyrimų ir konsultacijų bendrovės *Gartner* atliktas tyrimas pranašavo, kad 2013 metais įmonių investicijos į didžiuosius duomenis sieks 34 mlrd. JAV dolerių, o per 2016 metus investicijos išaugs iki 232 mlrd. JAV dolerių [9]. Pagal naujausius 2016 metų *Gartner* tyrimus, investicijų apimtis į didžiuosius duomenis ir toliau didėja. Apklausa atskleidė, kad net 48 % kompanijų 2016 metais investavo į didžiuosius duomenis [11]. Tuo tarpu didžiojo ketverto (angl. Big4) įmonės siekia būti didžiųjų duomenų technologijų įgyvendinimo priešakyje, kadangi jie mato didžiuosius duomenis kaip vis svarbesnę jų atliekamų audito procedūrų dalį [9]. Allesas ir Grėjus taip pat pritaria, kad didieji duomenys gali būti labai vertingi finansinių ataskaitų auditoriams, tačiau ši prielaida neturėtų būti laikoma kaip faktas be atsakymų į klausimus, profesinio

skepticizmo bei tolesnių tyrimų, kadangi didžiųjų duomenų taikymas audite yra daug mažiau aiškus negu kitose srityse [9]. Tokios pat nuomonės yra ir Wang bei Cuthbertson, kurie teigia kad: „Duomenų analizė pabrėžiama, kaip veiksminga audito užduočių priemonė. Tačiau mūsų supratimas ir duomenų analizės naudojimas audito užduočių srityje vis dar yra ribotas“. Autoriai taip pat susiduria su daugybe neatsakytų klausimų apie galimas didžiųjų duomenų analitikos naudojimo pasekmes. Pavyzdžiui, ar duomenų analitikos naudojimas sumažintų audito atlikimo valandas, ar padėtų geriau suprasti klientų sandorius, ar suteiktų daugiau galimybių atskleisti ir užkirsti kelią sukčiavimams, ar duomenų analizės naudojimas gali pakeisti esamas audito procedūras, ar didžiųjų duomenų naudojimas turėtų įtakos įdarbinimo sprendimams ir begale kitų klausimų [10].

Audito kompanijos *Ernst & Young* (toliau EY) partneris R. Ramlukanas tai pat pripažįsta, kad nors jau yra pasiekta didelė pažanga ir didžiųjų duomenų bei analizės privalumai audito srityje jau matomi, vis tik tai yra kelionė. Norėdamas apibūdinti, kur auditas kaip profesija yra šiandien, jis lygina ją su televizijos ir filmų paslaugas teikiančia bendrove *Netflix*. Kai bendrovė *Netflix* 1997 metais pradėjo „filmai paštu“ (angl. "DVD-by-mail") modelį, išsiųsdavo filmus savo klientams, kurie po tam tikro laiko juos grąžindavo. *Netflix* visada žinojo, kad ateitis buvo filmų transliacija internetu, tačiau tuo metu technologija dar nebuvo parengta. Ramlukano nuomone, šiandien audito profesijos atstovai užsiima „filmai paštu“ ekvivalentu, perduodami duomenis iš klientų auditoriams. Bet tai, ką, pasak Ramlukano, *EY* siekia turėti yra pažangus audito įrankis, kuris būtų įdiegtas klientų duomenų centruose ir perduotų analizės rezultatus realiu laiku tiesiai audito komandoms [1].

Auditorių tiesioginis prisijungimas prie klientų duomenų bazių yra priešaringai vertinamas audito įmonių teisininkų. Pastarieji bijo, kad tiesioginis prisijungimas gali turėti neigiamos įtakos klientų duomenų sistemos veikimui ar net gali sugadinti duomenis. Teisininkai taip pat turi pastabų dėl 100 % duomenų testavimo, kadangi testuojant visus duomenis būtų sunku apsiginti, jeigu auditoriai vis tik nepastebėtų apgaulės ar klaidos. Tuo tarpu atliekant įprastinę atranką galima argumentuoti praleistą klaidą tuo, kad ji nepapuoletų atrinktą testuoti imtį. Nepaisydama visų įspėjimų, audito įmonės *PwC* partnerė Mary Grace Davenport teigia, kad: „Neįmanoma, kad auditoriai nenaudotų didžiųjų duomenų, kai jų klientų rinkodaros padaliniai taip stipriai remiasi jais“ [9].

Zhangas, Yangas ir Appelbaum plėtoja savo požiūrį apie didžiųjų duomenų naudojimą audite, remdamiesi keturių „v“ (apimtis, sparta, įvairovė, tikslumas) apibrėžimu. Autoriai teigia, kad būtent tikslumas yra didžiausias iššūkis auditoriams, norintiems naudoti didžiuosius duomenis. O norint prisitaikyti prie šio iššūkio reikėtų į audito įrodymų standartą (500-asis TAS) įtraukti įrodymus, gautus iš didžiųjų duomenų analitikos procedūrų [12]. Kragelis ir Titera taip pat teigia, kad audito standartai nesikeitė kartu su technologijų pažanga ir vis dar pabrėžia atrankas [8]. Su tuo, kad TAS

turėtų pasipildyti didžiųjų duomenų sukurtais įrodymais bei procedūromis, sutinka dauguma autorių, rašančių šia tematika [1, 6, 7, 8, 9, 10, 12].

Reaguodama į besikeičiančią technologinę aplinką bei duomenų analizės panaudojimo audito procedūrose keliamus klausimus, tarptautinių audito ir užtikrinimo užduočių standartų valdyba (toliau IAASB) 2015 metais įsteigė duomenų analitikos darbo grupę (toliau DAWG) tam, kad informuotų valdybą kaip ir kada būtų efektyviausia reaguoti į technologijų plėtrą visuomenės interesų labui. DAWG bendradarbiauja su įvairiomis suinteresuotomis šalimis, įskaitant apskaitos įmones, nacionalinių audito standartų leidėjus, audito reguliavimo ir priežiūros tarnybas bei tarptautinės buhalterijų federacijos (toliau IFAC) valdybos narius. DAWG veikla taip pat apima stebėjimą ir informacijos pateikimą apie įvairius duomenų analizės panaudojimo atvejus, susijusius su finansinių ataskaitų auditu. 2016 metų rugsėjį pasirodžiusiame leidinyje DAWG aptaria didžiųjų duomenų analitikos privalumus, trūkumus, teikiamas galimybes, atsako į klausimą, kodėl nėra TAS pakeitimų, įtraukiančių duomenų analizės naudojimą į audito procedūras. DAWG neneigia, kad duomenų analizės naudojimas finansinių ataskaitų audite gali pagerinti audito kokybę. Priešingai, sako, kad naudojant technologijas ir duomenų analizę, auditoriui suteikiamos galimybės efektyviau ir patikimiau suprasti audituojamą subjektą ir jį supančią aplinką, taip pagerinant rizikos vertinimo ir atsako į ją procedūras. Kaip rašoma leidinyje, DAWG į duomenų analizę žiūri optimistiškai, tačiau teigia, kad dar yra daug neatsakytų klausimų, tokių kaip: didžiųjų duomenų prieiga, teisiniai reikalavimai, duomenų analizės žinių trūkumas, audito priežiūros ir reguliavimo institucijų didžiųjų duomenų naudojimo patirties stoka, investicijos į auditorių paruošimą duomenų analizės naudojimui. DAWG pabrėžia, kad IAASB turi būti labai atsargūs, kad per anksti nepradėtų standartų, susijusių su duomenų analize keitimo, ypač jeigu tai gali turėti nenumatytų pasekmių [13].

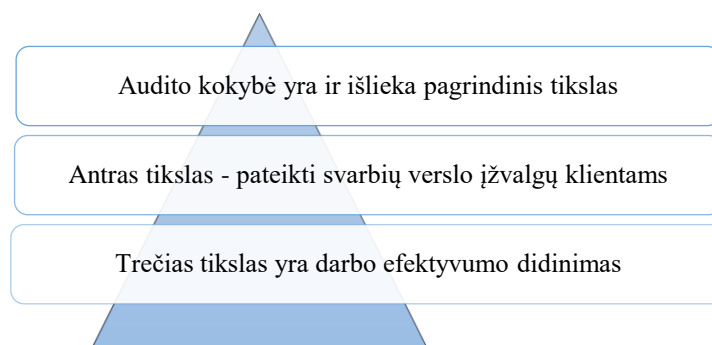
Didžiųjų duomenų naudojimas auditorių veikloje būtų naudingas ne tik todėl, kad išgaunama geresnė informacija, bet tai padėtų perkelti profesiją į vertės grandinę ir tapti tikrais verslo partneriais, o ne tik sandorio paslaugų teikėjais [8]. Darius Vaicekaskas savo daktaro disertacijoje teigia, kad: „Padidėjęs dėmesys auditorių darbo kokybei užtikrinant teisingą ir patikimą finansinę informaciją išorės vartotojams, padidino ir audito klientų lūkesčius, svarstant ką auditas turi atlikti ir kokią naudą turi suteikti. Atliekami įvairūs tyrimai ir autoriaus praktiniai stebėjimai atskleidžia, jog klientai vis dažniau mato auditorius kaip konsultantus, galinčius savo rekomendacijomis prisidėti prie klientų verslo stiprinimo. Tai suponuoja vis labiau audito paslaugų rinkoje įsitvirtinančią mintį, jog kokybiškas auditas turi būti papildomai naudingas ir klientui, kurdamas jam pridėtinę vertę“ [14].

Pridėtinės vertės kūrimas klientams neturėtų kurti papildomų audito procedūrų poreikio, kadangi tai mažintų darbo efektyvumą. Tai sufleruoja, kad norint vienu metu padėti kliento vadovybei priimti tinkamus verslo sprendimus ir nemažinti pačių darbo efektyvumo, bent dalis audito procedūrų turi



keistis. Technologiniai sprendimai, įskaitant algoritmus ir kitus didžiųjų duomenų analitikos įrankius, turėtų sutrumpinti rankinei analizei skirtą laiką ir leisti auditoriui daugiau laiko skirti analizei, kuri reikalauja profesinio sprendimo [15].

Didžiųjų verslo duomenų įtraukimas į audito procedūras šiek tiek pakeistų bendruosius audito tikslus, pridėtų naujų dedamųjų. Apibendrinant nagrinėtą literatūrą galima išskirti tris pagrindinius audito tikslus, kuriuos pasiekti galėtų padėti didžiųjų duomenų analitikos naudojimas audito procedūrose (žr. 2 pav.)[1, 6, 19, 21]:



pav. 2 Audito tikslai

Pagal 2 pav. matome, kad esminis ir nekintantis audito tikslas – visada buvo ir bus užtikrinimo užduoties atlikimo kokybė, kadangi nors paslaugą perka audituojamasis subjektas, šios paslaugos rezultatais naudojasi ir išorės informacijos vartotojai, kurie kartu su audituotos įmonės finansinėmis ataskaitomis gauna ir auditorių suteiktą užtikrinimą, kad finansinėse ataskaitose pateikiama patikima informacija [14]. Įžvalgų teikimas klientams ir darbo efektyvumo didinimas iš esmės labai svarbūs pačiai audito įmonei, kaip verslo organizacijai, kadangi tai ne tik padeda palaikyti našumą bei konkurencingumą, bet ir potencialiai galėtų padėti padidinti užimamą rinkos dalį [15].

#### **1.4. Priežastys stabdančios didžiųjų duomenų analitikos naudojimą audito procedūrose**

Audito srityje vis dažiau kalbama apie tai, kad didžiųjų duomenų atsiradimas, bei dažnėjantis duomenų analizės naudojimas ne tik atvėrė naujas galimybes, bet ir sukėlė nemažai naujų rūpesčių bei klausimų audito akademinei ir profesinei bendruomenei. Neatnaujinti profesiniai standartai, išaugę duomenų apsaugos reikalavimai bei darbuotojų turinčių žinių tiek audito, tiek didžiųjų duomenų analitikos srityje stoka yra tik keletas iš daugelio veiksnių keliančių susirūpinimą audito įmonėms pradedančioms naudoti pažangiąją analitiką.

Remiantis mokslinės ir metodinės literatūros šaltiniais išskiriamos pagrindinės priežastys stabdančios sėkmingą didžiųjų duomenų analitikos naudojimą audito procedūrose:

**Duomenų pateikimas netinkamu formatu.** Tai viena dažniausiai audito praktikoje pasitaikančių problemų, kuri neabejotinai stabdo duomenų analizės naudojimą. Duomenų pateikimas netinkamu formatu trukdo ar net užkerta kelią bet kokiai analizei, o informacijos konvertavimas į norimą pavidalą atima daug auditoriaus laiko taip mažindamas darbo efektyvumą [1, 15]. Galiausiai net sugaištas laikas tvarkant duomenų rinkinį neužtikrina, kad analizė galės būti atlikta, kadangi dažnai konvertavimo metu pakinta duomenų struktūra ir toks duomenų rinkinys tampa nepatikimu. Žinoma duomenų kokybė visuomet yra problema, kai kalbama apie didesnius jų kiekius [9, 10].

**Duomenų apsauga.** Duomenų surinkimo procese ir po jo didelį vaidmenį atlieka duomenų saugumas. Audito įmonės kasdiena sprendžia klausimus, susijusius su iš klientų gautų duomenų apsauga bei privatumu. Atsiradus duomenų pagrindu veikiančioms technologijoms šie klausimai tampa dar svarbesni [9, 15]. Didžiųjų duomenų eroje duomenis iš vieno šaltinio galima lengvai susieti su duomenimis iš kitų šaltinių. Tai reiškia, kad, kai nuteka dalis slaptų duomenų, jie gali greitai skliti ir susisieti su daugybe susijusių duomenų, kurie gali kelti pavojų daugelio verslo subjektų saugumui ir privatumui [16].

Ne tik auditoriai, bet ir jų audituojamos įmonės investuoja didelius pinigus į savo ir klientų duomenų apsaugą. Naudojami daugiasluksniai patvirtinimo procesai, saugomi duomenys yra šifruojami, diegiamos antivirusinės programos bei kiti saugumo sprendimai. Dėl šių priežasčių duomenų teikimo auditoriams procesas gali užimti daug laiko. Bendrovės net gali atsisakyti teikti duomenis argumentuodamos su saugumu susijusias problemas [1].

**Detalesnių duomenų išgavimas.** Teoriškai auditoriai turėtų turėti prieigą prie visos reikalingos įmonės informacijos. Praktiškai surinktų audito įrodymų apimtis ir kokybė priklauso nuo technologijų (t.y. ar įrodymai yra kompiuterizuota forma), kainos ir naudos suvaržymų bei socialinės sąveikos su klientais [17]. Šiandieną apskaitos duomenų išgavimas iš apskaitos programos dažniausiai yra susijęs su didžiosios knygos operacijomis. Norint audito procedūrose pasitelkti didžiuosius duomenis reikėtų kur kas detalesnių ataskaitų, detalizuojančių iš ko susideda kiekvienas įmonės apskaitos ciklas, toks kaip pajamos ar atsargos. Auditui paprašius platesnės informacijos susiduriama su tuo, kad audituojamos įmonės atstovai gali nemokėti suformuoti prašomos ataskaitos, arba programa net neturi tokių funkcijų [1]. Nustatyti su kuria iš minėtų problemų susidurta yra ganėtinai sudėtinga, kadangi klientai šiandien naudoja daugybę skirtingų tiek tarptautinių, tiek nacionalinių apskaitos programų (dažnai didesnės įmonės turi jų net ne po vieną), tad auditoriui žinoti visų programų funkcionalumus yra praktiškai neįmanoma. Be abejo, daugumai programų galima pridėti papildomų funkcionalumų, tačiau, kadangi įmonės dažniausiai perka bazinius programų paketus, kiekvienas programos papildymas daug kainuoja, o kol tai rūpi tik auditui, įmonės investuoti savo kaštų nenori.

**Dideli, sunkiai apdorjami duomenų kiekiai.** Detalesni duomenys lygiagrečiai reiškia ir didesnius jų kiekius. Įprastai pageidaujama, kad klientai auditui finansinius duomenis pateiktų *Microsoft Excel* formatu, kadangi tai yra patogiausias įrankis norint įvairiais pjūviais analizuoti duomenis. Čia susiduriama su problema, kad *Microsoft Excel* tapo nepajėgus nuskaityti bei analizuoti tiek duomenų eilučių, kiek jų savo veikloje per metus sugeneruoja įmonės, dažnai jame netelpa net mėnesio operacijos. Be to detalių ataskaitų generavimas dažnai reikalauja labai daug laiko, o tai trikdo audituojamų įmonių darbą.

Iš kitos pusės susiduriame su techniniais klausimais. Ar audito įmonės turi pakankamą technologinę infrastruktūrą saugoti bei apdoroti šiuos duomenis atsižvelgiant į jų dydį bei apimtį [15]?

**Analizės rezultatų, kaip audito įrodymų panaudojimas.** 500-asis Tarptautinis audito standartas (toliau TAS) *Audito įrodymai* teigia, kad: „Audito įrodymai – informacija, kuria remdamasis auditorius daro išvadas, kuriomis grindžiama auditoriaus nuomonė. Audito įrodymus sudaro apskaitos dokumentuose esanti informacija, kuria remiantis rengiamos finansinės ataskaitos, ir kita informacija“, standarte tai pat pabrėžiama jog: „Rengdamas ir atlikdamas audito procedūras, auditorius turi įvertinti informacijos, kuri yra naudotina kaip audito įrodymai, tinkamumą ir patikimumą“ [18]. Naudoti aprašomąją statistiką įmonės galimų rizikos sričių supratimui yra gana įprasta ir paprasta, tačiau naudoti analitiką atsakui į nustatytą riziką ir priimti tai kaip patikimus audito įrodymus yra kur kas sudėtingiau. Viena iš problemų yra tai, kad dauguma didžiųjų duomenų analitikos metodų bei algoritmų veikia *juodosios dėžės* principu, tad naudojant analizės rezultatus, kaip audito įrodymus būtų sunku įrodyti jų tinkamumą bei patikimumą [1]. TAS taip pat egzistuoja tokia sąvoka, kaip audito įrodymų pakankamumas. Audito įrodymų pakankamumas ir tinkamumas tarpusavyje susiję. Iš esmės, jei pagrindinė informacija nėra patikima ar pakankamai stipri ir jos kilmė nėra patikrinama, kaip kai kurių analitikos metodų atveju, reikės surinkti ir peržiūrėti daugiau įrodymų. Tačiau pagal TAS nepakankamos kokybės įrodymų negalima kompensuoti renkant didesni jų kieki [18].

**Neatnaujinti audito standartai.** Ypatingai svarbi detalė su kuria susiduria audito įmonės galvodamos apie pažangios analitikos naudojimą yra tai, kad audito standartai buvo parašyti prieš išplintant prieigai prie automatizuotų technologijų. Priklausomai nuo pasirinktų leistinių nuokrypių automatizuotais duomenų tyrybos metodais gaunama daug įvairių rezultatų. Šiuo metu auditoriams kelia rūpestį šiais metodais gautų išimčių bei nuokrypių gausa, kadangi remiantis galiojančiais audito standartais visi iškilę klausimai turi būti apsvarstyti, dokumentuoti ir išspręsti [8, 9, 10, 19]. D. Baskinas taip pat pastebi, kad audito standartai yra grįsti tuo, ką buvo galima daryti prieš 50 metų,

tad net jeigu auditorius gali analizuoti 100 % transakcijų, standartai vis tiek reikalauja atlikti atranką, o tai iššaukia dubliavimą bei neefektyvumą [7].

Pagal TAS siekiant padaryti pagrįstas išvadas, kuriomis grindžiama auditoriaus nuomonė, renkant audito įrodymus galima atlikti tiek detaliuosius testus, tiek pagrindines analitines procedūras. Detalių įrodymų rinkimas gali būti gana brangus, tačiau laikomas labiau patikimumu pagal standartus, o analitinės procedūros yra laikomos pigesnėmis, bet reguliuojančių institucijų nuomone, yra mažiau patikimos. Abu įrodymų rinkimo būdai yra leidžiami pagal standartus, tačiau jų panaudojimo laipsnis priklauso nuo auditoriaus profesinio sprendimo. Sandorių ir likučių detalių testų reikalavimas yra gana apibrėžtas, tačiau analitinių procedūrų reikalavimas yra visiškai neapibrėžtas, išskyrus tai, kad jos turėtų būti taikomos finansiniams duomenims. Nesant išsamių rekomendacijų apie tai kokios analitinės procedūros turi būti atliekamos sukelia didelių diskusijų didžiųjų duomenų panaudojimo audite srityje. Kyla elementarūs klausimai: kaip galima išmatuoti audito įrodymų kiekį, kuriuos analitikai pateikia didžiųjų duomenų kontekste? Kaip metodologiškai pagrįstu būdu šiuos įrodymus galima sujungti su kitų tipų audito įrodymais? Kaip ir kokios kiekybinės priemonės gali būti naudojamos siekiant padėti auditoriui priimti sprendimą dėl audito įrodymų pakankamumo [20]? Kol audito standartai nebus atnaujinti vargu ar pavyks į šiuos klausimus atsakyti.

**Žmonių išteklių trūkumas.** Kvalifikuotų specialistų, turinčių žinių tiek audito, tiek didžiųjų duomenų analitikos srityje stoka tai pat išskiriama kaip priežastis stabdanti duomenų mokslo naudojimą audito procedūrose [11]. Naujos kartos auditoriams informacinių technologijų žinių reikės taip pat kaip ir tradicinių kompetencijų finansų srityje. Norint sėkmingai naudoti duomenų analitiką kasdieninėje veikloje audito įmonės turi ruošti kitokio tipo auditorius, jie turi tobulinti analitinę mąstymą, įgyti kompetencijų duomenų surinkimo, agregavimo ir analizės srityse [21]. Natūralu, kad norit naudoti pažangios analitikos rezultatus, kaip audito įrodymus, auditorius pats turi labai gerai išmanyti atliekamų analizių techniką bei ypatumus, kadangi šioje srityje remtis klaidingai atliktos analizės rezultatais yra ypač pavojinga. Įmonių aplinka ir kultūra tai pat atlieka svarbų vaidmenį, skatinant audito darbuotojus tobulinti dabartinių audito procedūrų efektyvumą ir dalinantis žiniomis bei patirtimi su kolegomis visoje organizacijoje [15].

### **1.5. Didžiųjų duomenų analitikos taikymas finansinių ataskaitų audito procedūrose**

Šiame poskyryje bus nagrinėjamas didžiųjų duomenų taikymas finansinių ataskaitų audito procedūrose. Pirmiausia bus pristatomos didžiųjų duomenų analizės teikiamos galimybės bei privalumai. Įvardijus teikiamą naudą bus nagrinėjami atskiri audito etapai ir juose galimos naudoti duomenų tyrybos technikos, kurios padėtų šias galimybes realizuoti.

### 1.5.1. Didžiųjų duomenų analitikos teikiamos galimybės ir privalumai

Didžiųjų duomenų analitikos naudojimas suteikia daug pranašumų ir galimybių audito įmonėms bei jų klientams. Šias galimybes pagal jų pobūdį galima klasifikuoti į įtakojančias audito veiksmingumą, efektyvumą bei kokybę ir prisidedančias prie rizikos vertinimo bei valdymo [15].

Spartus IT sektoriaus augimas veikia bendrovių finansinių ataskaitų auditą dvejopai. Pirma, auditorius privalo įvertinti kliento IT struktūrų funkcionalumą ir efektyvumą atliekant auditą. Antra, įvertinti ar pats auditorius yra pajėgus taikyti sudėtingus IT įrankius efektyvumo audito procesuose didinimui [23]. Duomenų, su kuriais auditoriai turi susidoroti, kiekis ir įvairovė yra gerokai išaugęs. Tačiau dauguma didžiųjų audito kompanijų jau turi statistinės analizės programinius įrankius orientuotus į audito procedūras, leidžiančius atlikti įvairias analizes, skaičiavimus su pakankamai dideliu duomenų kiekiu. Šie ir kiti IT technologijų sprendimai turi didžiulį potencialą pagerinti audito užduočių našumą bei efektyvumą. Be to, pažangiosios analitikos naudojimas padeda auditoriams išlaikyti profesinį skepticizmą, o tai pagal TAS yra būtina norint kritiškai vertinti audito įrodymus [18]. Didžiųjų duomenų metodai audito procese gali būti naudojami siekiant padidinti audito aprėptį, iširti vidaus kontrolės veiksmingumą ir sandorių tikslumą, tuo pačiu metu gerokai sumažinant audito išlaidas, išteklius ir laiką [15].

Duomenų mokslas ir susijusios technologijos pastaraisiais metais labai pasistūmėjo, įtraukė metodus ir programines įrangas iš daugelio sričių: duomenų analizės, verslo įžvalgos, matematikos, tikimybių teorijos, statistikos, pritaikė modelių atpažinimą, duomenų vizualizaciją, teksto ir procesų tyrybą. Duomenų mokslas gali būti taikomas siekiant atlikti kokybiškesnius auditus ir pateikti naujų formų audito įrodymus. Pasitelkę didžiųjų duomenų analitikos įrankius auditoriai galėtų analizuoti ne tik duomenų suvestines, bet visus finansinę informaciją sudarančius duomenis. Identifikuoti rizikas, koreliacijas, svyravimus, išskirtis, ieškoti anomalijų, kurios galėtų parodyti potencialias problemas ar net apgaulę [9]. Tai leistų pagerinti analitines procedūras visose audito etapuose. Duomenų tyrybos bei prognozavimo metodai taip pat gali būti naudingi siekiant nustatyti ar prognozuoti finansines nesėkmes bei paremti veiklos tęstinumo prielaidą. Tinkamai pritaikius šias technologijas būtų galima galvoti ne apie dalies finansinių operacijų tikrinimą, kas atsitiktinės atrankos pagalba audito praktikoje vyksta šiandien, bet visų finansinių operacijų analizę [12, 21].

Nors audito įmonės vis dažniau remiasi didžiųjų duomenų analitika ir kitomis technologijomis, automatizuoja vis daugiau procesų, auditas vis dar labai priklausomas nuo auditorių profesinės kompetencijos ir kitų asmeninių bruožų [9, 7, 15, 26]. Žinoma, didžiųjų duomenų analitikos naudojimas, atliekant finansinės atskaitomybės auditą, nepakeis auditoriaus profesinių sprendimų ir profesinio skepticizmo. Tai reiškia, kad auditoriai vis tiek turės gerai suprasti audituojamą subjektą

ir jo aplinką. Auditoriaus gebėjimas analizuoti informaciją, pateiktą finansinėse ataskaitose, naudojant pažangiąją analitiką, leistų auditoriui įgyti gilesnį supratimą apie tai, kas faktiškai įvyko finansinės atskaitomybės sistemoje, o tai turėtų teigiamą poveikį audito kokybei ir vertės kūrimui [15].

315-ajame TAS rizikos vertinimo procedūros apibrėžiamos, kaip audito procedūros, atliekamos siekiant susipažinti su įmone ir jos aplinka, įskaitant įmonės vidaus kontrolę, taip pat nustatyti ir įvertinti reikšmingo iškraipymo riziką dėl apgaulės ar klaidos finansinių ataskaitų ir tvirtinimų lygmeniu [18]. Naudojant didžiųjų duomenų analitiką auditoriai gali veiksmingiau ir efektyviau analizuoti didelius duomenų rinkinius, taip įgyjant galimybę nustatyti audito rizikas ankstyvame audito etape. Be to, duomenų tyrybą sustiprina rizikos vertinimo ir atsako į riziką procedūras, kadangi ji suteikia auditoriams galimybę gauti patikimesnį supratimą apie audituojamą įmonę ir jos aplinką [9]. Ši platesnė ir gilesnė auditorių įžvalga suteikia daugiau vertingos informacijos apie audituojamo subjekto vidinį rizikos vertinimą bei verslo operacijas, lyginant su dabartiniu, daugiausia rankiniu būdu atliktu auditu [15].

Galiausiai, atsižvelgiant į rizika grįstą audito pobūdį, reikia nepamiršti audito rizikos susijusios su pakankamų, tinkamų ir patikimų audito įrodymų gavimu. Didžiųjų duomenų eroje rizikos vertinimo procesas gali tapti dar sudėtingesnis, nes didžiųjų duomenų nestruktūrinis pobūdis didina dviprasmybę. Kita vertus, pilnas finansinių duomenų rinkinio testavimas gali sumažinti su tam tikrais objektais susijusią riziką iki nulio. Taigi auditoriams tenka iššūkis susijęs su tuo, kaip jie gali gauti naudos iš padidėjusio informacijos kiekio, su kuria jie susiduria, ir kaip užtikrinti, kad audito sprendimai būtų pagrįsti kokybiška informacija, kuri yra svarbi ir patikima [24]. Vienas iš patikimų rizikos vertinimo procedūrų, naudojant duomenų analizės priemones, pavyzdžių galėtų būti IT incidentų duomenų analizė. Analizės metu galima įvertinti incidentų sunkumą ir dažnumą, nustatyti informacijos koregavimus, šalinimus, įprastuose sandoriuose atsiradusius pažeidimus ar anomalijas, taip nustatant galimus sukčiavimo požymius [25].

Akivaizdu, kad didžiųjų duomenų naudojimas audito procedūrose ir jų teikiama nauda aktuali ne tik pačioms audito įmonėms, bet ir jų klientams. Tai puikiai atspindi audito kompanijos *KPMG* ir *Forbes Insights* atlikta apklausa, kurioje dalyvavo 200 įmonių atstovai. Apklausos metu nustatyta, kad beveik 80 % apklaustųjų teigia, kad auditoriai turėtų naudoti didesnes atrankas bei kitas duomenų išgavimo ir analizės technikas jų kasdienėje veikloje. Vienas iš respondentų *Arrow Electronics* finansų direktorius C. Stansburis šia tema pasisakė taip: „Klientams nebereikia tik galinio vaizdo veidrodžio atspindžio, mes norime priekinio stiklo vaizdo, norime matyti, kur mes einame ir kaip naviguoti rizikos, galimybių, reglamentų keitimo, konkurencijos ir globalizacijos kraštovaizdyje“ [26]. Taigi

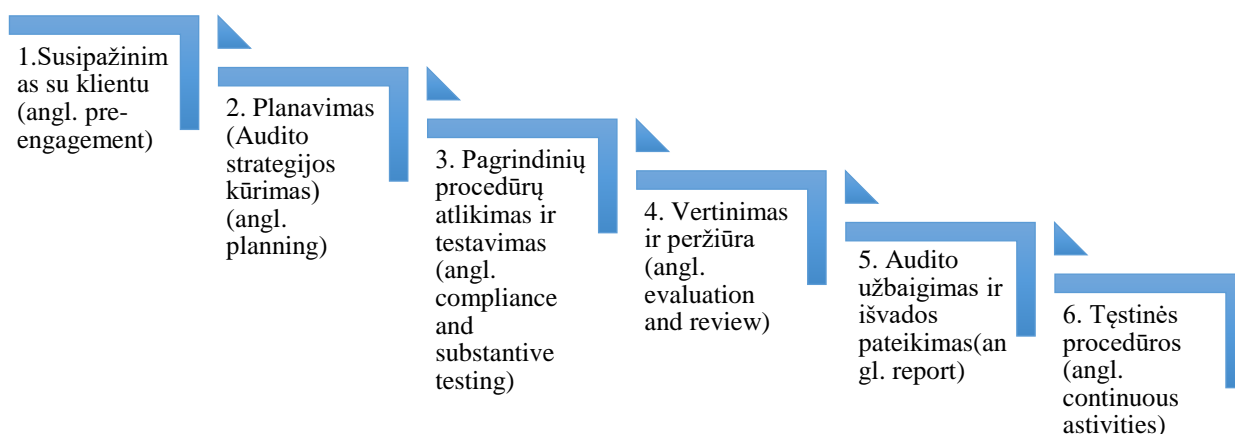
galima teigti, kad duomenų tyrybos naudojimas audito procedūrose prisidėtų prie audito įmonių konkurencinio pranašumo didinimo bei ilgalaikių santykių su klientais palaikymo.

Atsižvelgiant į mokslinę literatūrą bei audito srities profesionalų nuomonę, išplėstinių analitinių procedūrų derinimas su įprastais audito metodais bei ekspertų pagalba galėtų būti išties vertingas audito sričiai [9, 10]. Tačiau pažangių technologijų įtraukimas į audito procesą neabejotinai kelia dvejonių dėl mažiau aiškios audito atlikimo sekos. Nėra aiškių nurodymų, kurias įprastas audito procedūras reikėtų pakeisti duomenų analizės technikomis, ar visuose audituose ir kaip dažnai jas naudoti [24].

### 1.5.2. Audito etapai ir juose taikomos analitinės procedūros

Didžiųjų duomenų banga išties atnešė nemažai galimybių audito veiklai. Tačiau kuriuose audito etapuose ir kokias duomenų tyrybos technikas reikėtų naudoti? Nors TAS reikalauja atlikti analitines procedūras planavimo ir peržiūros fazėse, kokius analizės metodus ir kaip taikyti sprendžia pats auditorius. Be to, taikyti analitinius metodus pagrindinių procedūrų atlikimo stadijoje ar ne – tai pat paliekama spręsti auditoriui [10, 18, 22].

Audito atlikimo eiga skiriama į etapus. Kiekvienas audito atlikimo etapas ruošiamas ankstesnio etapo pagrindu, ir, prieš pradėdant kitą etapą, būtina baigti einamąjį. Be to, audito etapai, gali skirtis atsižvelgiant į audito sudėtingumą ir mastą [4]. Skirtinguose literatūros šaltiniuose galima rasti įvairių audito skirstymo į etapus metodikų, tačiau, iš esmės, atliekamos procedūros išlieka tos pačios. 1986 metais Cushingas ir Loebbecke aštuonių didžiausių audito įmonių faktinę audito užduoties struktūrą perteikė šešių dalių modeliu (žr. 3 pav.). Šis modelis atspindi visą audito eigą, nuo susipažinimo su klientu prieš pasirašant sutartį iki audito užbaigimo ir tęstinių procedūrų [22]. Pabrėžtina, kad audito įrodymai yra renkami visuose audito etapuose [27].



pav. 3 Audito etapai pagal Cushingą ir Loebbecke (1986)

Toliau bus aptariamas kiekvienas finansinių ataskaitų audito etapas. Pristatomi svarbiausi žingsniai ir atliekamos procedūros. Taip pat bus įvardinti kiekvienoje fazėje galimi pritaikyti išplėstinės duomenų analizės metodai. Siūlomi metodai atrinkti remiantis 2018 m. sausio mėnesį paskelbtu Appelbaum, Kogano, ir Vasarhelyi atliktu tyrimu, kuriame pateikiama susisteminta informacija iš 301 skirtingo straipsnio nagrinėjančio išorinio audito užduočių analitinių procedūrų taikymą. Atsižvelgiant į tai, kokie metodai aprašomi literatūroje, autorių jie yra suskirstyti pagal techniką bei audito epatą, kuriame jie galimi naudoti. Išskiriamos 5-ios pagrindinės technikos: audito tyrimas, duomenų tyryba be mokytojo, su mokytoju, regresija ir kitos statistikos. Prie kiekvieno metodo pavadinimo pateikiamas ir skaičius šaltinių, kuriuose buvo aprašytas šio metodo naudojimas atitinkamoje fazėje [22].

**Susipažinimas su klientu.** Prieš pradėdant auditą, yra keli svarbūs žingsniai. Pirmiausia, audito įmonė turi nuspręsti, ar priimti naują klientą, įvertinus ar klientas buvo susijęs su neetišku verslo elgesiu. Tai pat kasmet iš naujo sprendžiama ar toliau dirbti su jau esamais klientais, tikrinant ar šie nepakeitė savo verslo krypties į rizikingesnę ir pan. Kasmetinis audito veiklos atnaujinimas nėra automatinis procesas. Nusprendus pasirašyti sutartį, audito įmonė su klientu aptaria pagrindines užduotis, įskaitant duomenų rinkimo grafiką, audito apimtį, trukmę bei numatomą audito išvados pateikimo datą [28].

Remiantis Appelbaum, Kogano ir Vasarhelyi atlikta literatūros atpžvalga, šiame etape dažniausiai rekomenduojama taikyti rodiklių analizę, regresiją, aprašomąją statistiką bei ekspertų sistemas. Nustatyti tik keli straipsniai, kuriuose minima vizualizacijos, teksto tyrybos, daugiakriterijinių sprendimų priėmimų bei struktūrinių modelių nauda šiame etape (žr. 1 lentelė) [22].

lentelė 1. Galimos duomenų tyrybos technikos susipažinimo su klientu etape

Technikos:	Audito tyrimas	Be mokytojo (angl. unsupervised)	Su mokytoju (angl. supervised)	Regresija	Kitos statistikos
<b>Audito etapas:</b>					
<b>Susipažinimas su klientu:</b>	Rodiklių analizė (Ratio Analysis) (21)	Vizualizacijos (Visualizations) (3)	Ekspertų sistemos /sprendimų pagalba (Expert Systems/ Decision Aids) (7)	Logistinė regresija (Log Regression) (15)	Daugiakriterijiniai metodai (Multi- criteria)(3)
		Teksto tyryba (Text Mining) (4)		Tiesinė regresija (Linear Regression) (7)	Struktūriniai modeliai (Structural Models) (1)
		Procesų tyryba (Process Mining) (1)		Laiko eilutės (Time Series) (2)	Aprašomoji statistika (Descriptive Statistics) (11)
				Vienmatė ir daugiamatė (Univariate and Multivariate) (6)	

Lentelė sudaryta pagal Appelbaum, Kogano, ir Vasarhelyi atliktą tyrimą



**Planavimas.** Planavimo metu auditorius: susipažįsta su audituojamo subjekto veikla, jos ypatumais, taikoma apskaitos sistema; vertina susijusių šalių, apgaulės ir korupcijos riziką; vertina informacinių sistemų audituojamame subjekte naudojimą; vertina ankstesnio audito rezultatus; vertina audituojamo subjekto vidaus kontrolę, susipažindamas su audituojamo subjekto vidaus kontrolės aplinka ir kontrolės procedūromis; nustato kontrolės procedūrų veiksmingumą ir nuoseklumą; nustato reikšmingumą; vertina vidaus auditorių darbą ir jo panaudojimą; sprendžia, ar pasitelkti ekspertus ir išorės auditorius. Planavimo etapas baigiamas audito plano ir audito programų parengimu [4].

Dauguma šaltinių aprašančių šio etapo išplėstinės duomenų analizės galimybes nagrinėja įvairių tipų regresijas, aprašomąją statistiką, rodiklių analizę. Tokie metodai, kaip klasterizavimas, teksto tyryba, procesų optimizavimas ir kiti, yra aptariami kur kas rečiau (žr. 2 lentelė) [22].

lentelė 2. Galimos duomenų tyrybos technikos planavimo etape

Technikos:	Audito tyrimas	Be mokytojo (angl. unsupervised)	Su mokytoju (angl. supervised)	Regresija	Kitos statistikos
<b>Audito etapas:</b>					
<b>Planavimas:</b>	Transakcijų testavimas (Transaction Tests) (20)	Klasterizavimas (Clustering) (6)	Procesų optimizavimas (Process Optimization) (4)	Logistinė regresija (Log Regression) (65)	Daugiakriterijinis sprendimų priėmimas (Multi-criteria Decision Aid) (15)
	Rodiklių analizė (Ratio Analysis) (159)	Teksto tyryba (Text Mining) (6)	Ekspertų sistemos /sprendimų pagalba (Expert Systems/ Decision Aids) (33)	Tiesinė regresija (Linear Regression) (36)	Aprašomoji statistika (Descriptive Statistics) (27)
	CAATS (19)	Vizualizacijos (Visualizations) (7)	BBN (22)	Laiko eilutės (Time Series) (33)	Struktūriniai modeliai (Structural Models) (7)
			Tikimybiniai modeliai (Probability Model) (19)	ARIMA (9)	
				Vienmatė ir daugiamatė (Univariate and Multivariate) (25)	

Lentelė sudaryta pagal Appelbaum, Kogano, ir Vasarhelyi atliktą tyrimą

**Pagrindinių procedūrų atlikimas ir testavimas.** Šis etapas apima išsamias audito procedūras, susijusias su ūkinių operacijų ir buhalterinių sąskaitų likučiais, bei analitines procedūras. Šiame etape renkami būtini ir patikimi duomenys auditui atlikti. Duomenys gali būti iš vidinių ir išorinių šaltinių. Audito įrodymai renkami siekiant pagrįsti finansinės atskaitomybės tvirtinimus: baigtumą, tikrumą, įvertinimą, teisėtumą, atskleidimą. Įrodymams rinkti gali būti taikomos šios audito procedūros: skaičiavimas, patvirtinimas, patikrinimas, stebėjimas, paklausimas, analitinės procedūros [18].

Atlikęs audito procedūras, auditorius turi įvertinti, ar nustatytos klaidos – kiekviena atskirai ir drauge su visomis kitomis neištaisytomis klaidomis – sudaro reikšmingą iškraipymą finansinėse ataskaitose. Auditorius tai pat įvertina ar nustatytos klaidos turi įtakos turto valdymo, naudojimo ir disponavimo juo teisėtumui bei turi atkreipti dėmesį į panašių neaptiktų klaidų galimybę [4].

Kalbant apie duomenų tyrimo metodus šiame etape labai populiarios išlieka audito tyrimo technikos, taip pat visi regresijos metodai, aprašomoji statistika, ekspertų sistemos, neuroniniai tinklai (BBN) bei tikimybiniai modeliai. Šioje fazėje galima atrankas pakeisti į 100 % duomenų tyrimą. Rečiau literatūros šaltiniuose šiame etape minimi duomenų tyrybos metodai be mokytojo, atraminių vektorių metodas, genetiniai algoritmai bei kai kurios kitos statistikos (žr. 3 lentelė) [22].

lentelė 3. Galimos duomenų tyrybos technikos pagrindinių procedūrų ir testavimo etape

Technikos:	Audito tyrimas	Be mokytojo (angl. unsupervised)	Su mokytoju (angl. supervised)	Regresija	Kitos statistikos
<b>Audito etapas:</b>					
<b>Pagrindinės procedūros ir testavimas:</b>	Rodiklių analizė (Ratio Analysis) (139)	Vizualizacijos (Visualizations) (8)	SVM (1)	Tiesinė regresija (Linear Regression) (50)	Benfordo dėsnis (Benford's Law) (7)
	Atrankos (Sampling) (145)	Teksto tyryba (Text Mining) (5)	ANN (8)	Laiko eilutės (Time Series) (36)	Aprašomoji statistika (Descriptive Statistics) (24)
	CAATS (21)	Procesų tyryba (Process Mining) (4)	Genetiniai algoritmai (Genetic Algorithms) (1)	ARIMA (12)	Struktūriniai modeliai (Structural Models) (7)
			Ekspertų sistemos /sprendimų pagalba (Expert Systems/ Decision Aids) (26)	Vienmatė ir daugiamatė (Univariate and Multivariate) (22)	AHP (1)
			Atsitiktinių miškų metodai (Bagging, Boosting) (4)		Monte Carlo modeliavimas (Monte Carlo Study) (3)
			Neuroniniai tinklai (BBN) (29)		
			Tikimybiniai modeliai (Probability Models) (17)		

Lentelė sudaryta pagal Appelbaum, Kogano, ir Vasarhelyi atliktą tyrimą

**Vertinimas ir peržiūra.** Šioje fazėje auditorius peržiūri surinktus audito įrodymus, nustato ar visi suplanuoti tikslai yra pasiekti. Tai pat sprendžiama ar reikia peržiūrėti bendrąją audito strategiją ir

audito planą. Auditorius, prieš pateikiant išvadą, privalo įvertinti ar visi trūkumai, klaidos buvo pašalinti bei ar nekyla abejonių dėl vadovybės sąžiningumo [31].

Peržiūros etape dažniausiai aptariamas santykinės analizės, regresijos, laiko eilučių bei aprašomosios statistikos naudojimas audito procedūrose. Tikimybiniai modeliai, neuroniniai tinklai, procesų tyryba yra minimi kur kas rečiau (žr. 4 lentelė). Šio etapo tyrimai gali būti susiję su kryžminio patikrinimo metodais bei rezultatai, kurie pripažinti išskirtimis analize, naudojant skirtingus metodus [22].

lentelė 4. Galimos duomenų tyrybos technikos vertinimo ir peržiūros etape

Technikos:	Audito tyrimas	Be mokytojo (angl. unsupervised)	Su mokytoju (angl. supervised)	Regresija	Kitos statistikos
<b>Audito etapas:</b>					
<b>Vertinimas ir peržiūra:</b>	Rodiklių analizė (Ratio Analysis) (115)	Vizualizacijos (Visualizations) (8)	Expert Systems/ Decision Aids (24)	Tiesinė regresija (Linear Regression) (36)	Daugiakriterijinis sprendimų priėmimas (Multi-criteria Decision Aid) (16)
	CAATS (14)	Procesų tyryba (Process Mining) (2)	BBN (4)	Laiko eilutės (Time Series) (28)	Aprašomoji statistika (Descriptive Statistics) (23)
			Tikimybiniai modeliai (Probability Models) (16)	ARIMA (10)	Struktūriniai modeliai (Structural Models) (7)
				Vienmatė ir daugiamatė (Univariate and Multivariate) (26)	Hipotezių tikrinimas (Hypothesis Evaluation) (1)

Lentelė sudaryta pagal Appelbaum, Kogano, ir Vasarhelyi atliktą tyrimą

**Audito užbaigimas ir išvados pateikimas.** Atlikus suplanuotas audito procedūras ir gavus pakankamų ir tinkamų įrodymų apie kliento vidaus kontrolės veiksmingumą, auditorius gali pateikti nuomonę apie tai, ar finansinėse ataskaitose nėra reikšmingų iškreipimų. Atsižvelgiant į pasitikėjimą finansinės atskaitomybės kokybe, yra keli audito išvadų tipai, kurias auditorius gali pateikti. Auditoriaus išvadoje auditoriai gali pareikšti besąlyginę, sąlyginę, neigiamą nuomonę arba atsisakyti ją pareikšti [30].

Audito išvados formavimo fazėje pagrindiniai minimi metodai: rodiklių analizė, vizualizacija, ekspertų sistemos, logistinė ir tiesinė regresija, aprašomoji statistika ir daugiakriteriniai sprendimų priėmimo būdai. (žr. 5 lentelė) [21].

lentelė 5. Galimos duomenų tyrybos technikos audito užbaigimo etape

Technikos:	Audito tyrimas	Be mokytojo (angl. unsupervised)	Su mokytoju (angl. supervised)	Regresija	Kitos statistikos
<b>Audito etapas:</b>					
<b>Audito užbaigimas:</b>	Rodiklių analizė (Ratio Analysis) (35)	Vizualizacijos (Visualizations) (3)	Expert Systems/ Decision Aids (7)	Logistinė regresija (Log Regression) (22)	Multi-criteria Decision Aid (3)
		Procesų tyryba (Process Mining) (1)		Tiesinė regresija (Linear Regression) (11)	Aprašomoji statistika (Descriptive Statistics) (10)

Lentelė sudaryta pagal Appelbaum, Kogano, ir Vasarhelyi atliktą tyrimą

**Tęstinės procedūros.** Tęstinės procedūros padeda įvertinti kontrolės veiksmingumą ir nustatyti su rizika susijusius klausimus, kurie gali įtakoti būsimas finansines ataskaitas. TAS šiame etape nereikalauja atlikti analitinių procedūrų. Tačiau didieji duomenys taip pat gali sustiprinti auditoriaus žinias apie klientą šiame etape analogiškai kaip susipažinimo su klientu ir planavimo etapuose [33]. Straipsnių, kuriuose būtų aptariamas analitinių metodų naudojimas, šiame etape nenustatyta [22]. Galima teigti, kad pagal Appelbaum, Kogano ir Vasarhelyi atliktą literatūros apžvalgą dominuojantys metodai visose etapuose priklauso audito tyrimo ir regresijos technikoms. Taip pat gana dažnai aptariami neuroniniai tinklai, tikimybiniai modeliai, aprašomoji statistika [22].

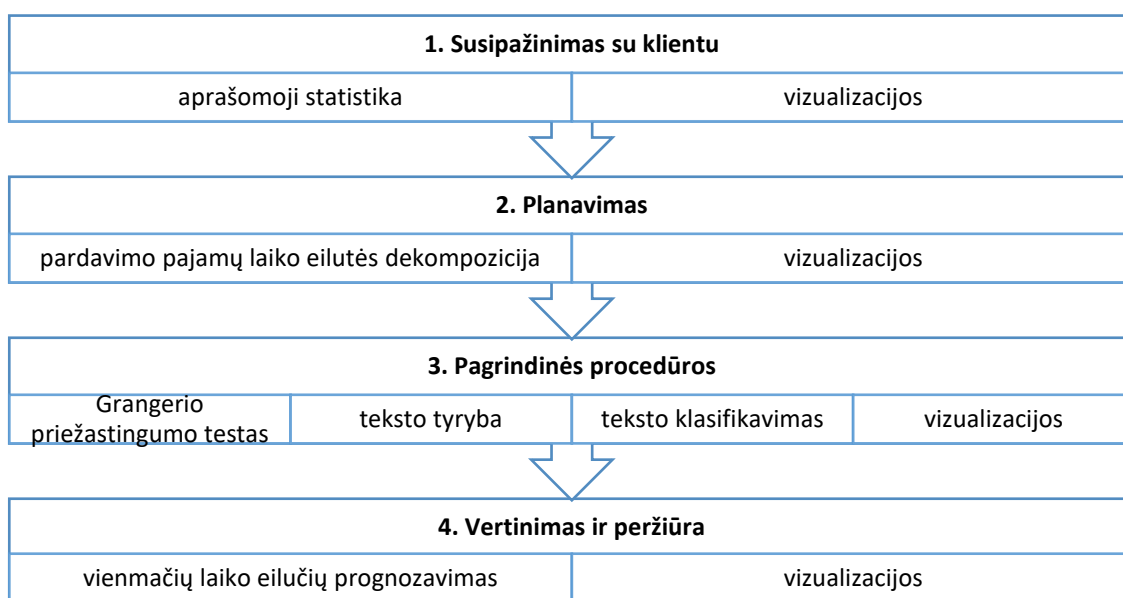
Apibendrinant nagrinėjamą temą galima teigti, kad didžiųjų duomenų analitikos metodai gali reikšmingai prisidėti apie finansinių ataskaitų audito procedūrų stiprinimo visuose audito etapuose, o tai ypač aktualu, kadangi visada siekiama, kad auditas būtų kuo kokybiškesnis. Galima rasti nemažai straipsnių aprašančių, kokius didžiųjų duomenų analitikos metodus būtų galima taikyti audite, tačiau tik keletas jų parodo kaip tai realizuoti praktiškai, be to dauguma šių realizacijų pritaikyti banko arba draudimo sektoriau audito procedūroms [25, 32]. Pastebima, kad literatūros šaltiniuose daugiausia kalbama apie dideles, viešo intereso įmones, o Lietuvos audituojamų įmonių rinka pakankamai siaura, jau nekalbant apie kompanijas, kurių akcijomis būtų prekiaujama biržoje, kurių yra labai mažai. Šio darbo pagrindinė užduotis pritaikyti didžiųjų duomenų analitikos metodus finansinių ataskaitų audito procedūrose, kai audituojama įmonė nėra viešo intereso, kadangi didžioji dalis audito klientų Lietuvoje yra uždarojo tipo akcinės bendrovės.

## 2. Tyrimo metodai

Šiame skyriuje supažindinama su tyrimo teorine schema bei pristatomi tyrimui naudojami metodai.

### 2.1. Tyrimo teorinė schema

Šiame darbe atliekamas tyrimas apims technikas patartinas naudoti pirmuose keturiuose audito etapuose (žr. 4 pav.).



pav. 4 tyrimo teorinė schema

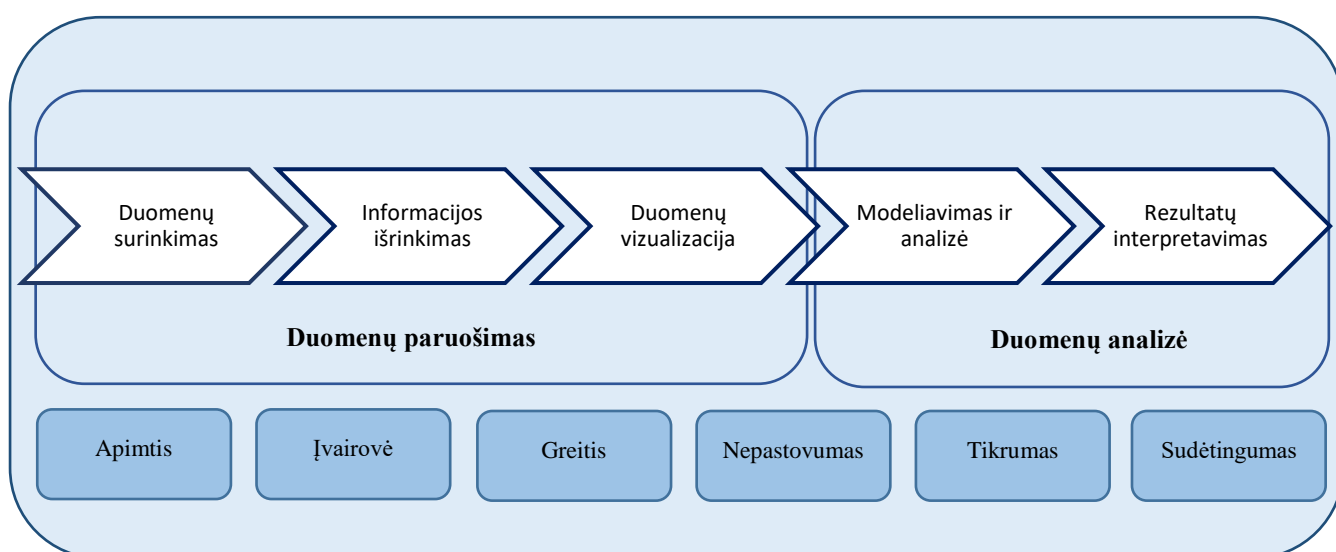
Tyrimo (žr. 3 skyrius) metodų realizacija pateikiama ne tokia tvarka kaip sudarytoje schemoje. Tiriomojoje dalyje metodai išskirti į dvi didesnes grupes t.y. laiko eilučių analizė ir teksto analizė (žr. 6 lentelė).

lentelė 6 tyrimo naudojamų didžiųjų duomenų technikų teorinė schema

Tyrimo technikos	Tyrimo metodai	Naudojami apskaitos duomenys	Tikslai
<b>Kitos statistikos</b>	Aprašomoji statistika	Didžioji knyga	Įvertinti pagrindinių įmonės finansinių duomenų charakteristikas.
<b>Laiko eilučių analizė</b>	Laiko eilutės dekompozicija (žr. 2.5 sk.)	Didžioji knyga	Susipažinti su audituojamo subjekto veikla, jos ypatumais.
	Grangerio priežastingumo testas (žr. 2.5 sk.)	Didžioji knyga	Įvertinti išorės veiksnių poveikį įmonės rodikliams.
	Laiko eilutės prognozavimas vienmačiais metodais (žr. 2.5.1 sk.)	Didžioji knyga	Veiklos tęstinumo prielaidos vertinimas.
<b>Teksto tyryba</b>	Tekstinių duomenų rinkinio analizė (žr. 2.3 sk.)	Bendrasis žurnalas	Bendrojo žurnalo įrašų tekstinės dalies testavimas.
	Teksto klasifikavimas SVM ir Naivaus Bajeso algoritmais (žr. 2.4 sk.)	Bendrasis žurnalas	Siekama patikrinti ar operacijos priskirtos tinkamai DK sąskaitai.

## 2.2. Didžiųjų duomenų apdorojimas ir tyryba

Duomenų analizė įvardijama, kaip pagrindinis didžiųjų duomenų uždavinys [34]. Procesas, kurio metu, naudojant įvairius duomenų analizės įrankius, bandoma nustatyti ir atrasti „užslėptas“ duomenų struktūras ir ryšius, vadinamas duomenų tyryba (angl. data mining). Duomenų tyrybos tikslas yra iš įprastai didelių duomenų aibių išgauti svarbią informaciją ir atmesti nereikšmingą [35]. Norint gauti informaciją iš duomenų, būtina šiuos duomenis tinkamai apdoroti [34]. Pagal žemiau pateiktą schemą (žr. 5 pav.) matome, kad didžiųjų duomenų apdorojimas įprastai skiriamas į dvi dalis: duomenų paruošimą ir po jo sekančią duomenų analizę. Apatinėje grafiko dalyje pateikiamos duomenų savybės, kurios daro duomenų apdorojimo procesą sudėtingu ir kompleksiniu [34].



pav. 5. Didžiųjų duomenų apdorojimo schema [34]

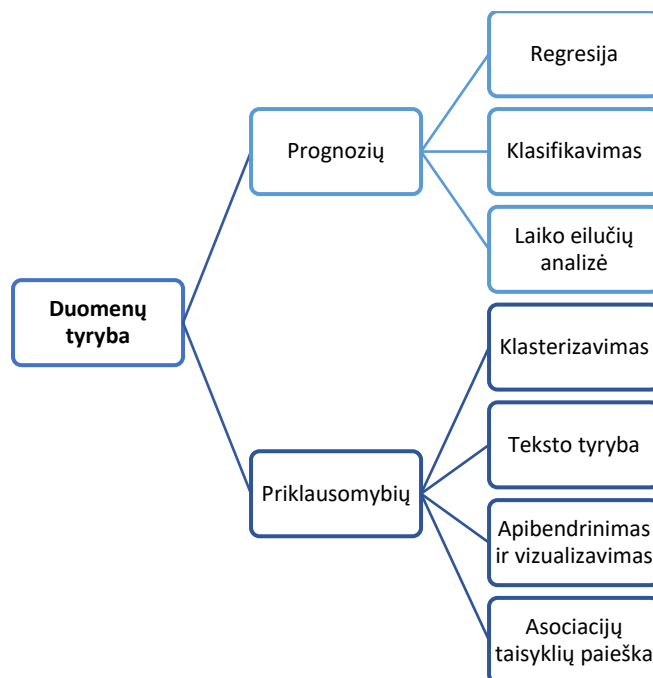
Pirmajame duomenų surinkimo žingsnyje turi būti nuspręsta, iš kokių šaltinių bus renkami duomenys: interneto (naršymo ar paieškos istorija, publikuoti straipsniai, pirkimai elektroninėse parduotuvėse), mobiliųjų telefonų, socialinių tinklų, įmonės vidinių dokumentų. Šiame etape turi būti pasirūpinta duomenų nuasmeninimu ir kitais duomenų tvarkymo teisiniais reikalavimais [36]. Vizualus didžiųjų duomenų pateikimas leidžia išvelgti galimas tendencijas, išskirtis, pasirinkti ir efektyviai naudoti naudingą informaciją [34]. Duomenų analizės etapui turi būti atrinkta tik reikalinga informacija, išgryninta, susieta su galbūt jau turimais duomenimis, surinktais iš kitų šaltinių, ir viskas susisteminta. Paskutinis rezultatų interpretavimo žingsnis yra svarbus tuo, kad skelbiami rezultatai turi būti patikimi, patikrinti bei teisingi [36].

Duomenų tyryba laikoma daugialype sąvoka: ji gali būti apibrėžiama, kaip duomenų bazių struktūros (modelių, jungčių, statistinių modelių ar šablonų) indentifikavimas, taip pat kaip statistikos taikymas duomenų analizėje bei prognozavime, siekiant atrasti tendencijas ir naujus šablonus didžiųjų

duomenų rinkinyje [37]. Mehmedas Kantardzičius taip pat teigia, kad duomenų tyryba yra kilusi iš daugybės disciplinų, tačiau dvi svarbiausios yra statistika ir mašininis mokymas (angl. machine learning) [38]. Duomenų tyrybos procesas apima 6 žingsnius [39]:

- 1) problemos suformulavimas (angl. problem definition);
- 2) duomenų suvokimas (angl. data understanding phase);
- 3) duomenų parengimas (angl. data preparation phase);
- 4) modelių parinkimas ir kūrimas (angl. modeling phase);
- 5) modelių vertinimas (angl. Evaluation phase);
- 6) modelių implementavimas realiam naudojimui (angl. *deployment phase*).

Pasitelkus duomenų tyrybą sprendžiami klasifikavimo, klasterizavimo, prognozavimo, regresijos, susietumo taisyklių paieškos ir kiti uždaviniai [35]. Paprastai duomenų tyrybos uždaviniai skiriami į dvi kategorijas (žr. 6 pav.): priklausomybių (angl. descriptive) ir prognozių (angl. predictive) [40, 41]. Priklausomybių kategorijos modeliai nustato duomenų ryšius ir struktūras. Prognozių modeliai, remdamiesi turimais sukauptais duomenimis, atlieka ateities įvykių prognozę. Duomenų tyrybos metodai gali būti taikomi duomenims, turintiems struktūrą, tačiau taip pat ir duomenims, kurie neturi griežtos struktūros arba yra nestructūrizuoti (angl. unstructured) [40].



pav. 6 Duomenų tyrybos uždaviniai pagal Weiss S.M. ir Indurkha N. [40, (8psl)]

Prognozių analitika (angl. predictive analytics) atsako į klausimą: kas gali įvykti [22]? Weiss'o ir Indurkha-os teigimu prognozė, yra svarbiausias duomenų tyrybos tikslas, ji kuria potencialiai didžiausią pelną ir turi tiksliausią apibūdinimą [41]. Šios grupės duomenų tyryba gali padėti numatyti

stichines nelaimės, tokias kaip potvyniai, uraganai, gali būti taikoma siekiant išvengti finansiniu apgavysčių, prognozuoti nemokumą ar finansines nesėkmes. Priklausomybių radimas puikiai papildo prognozių duomenų tyrybą, tačiau yra tinkamesnis sprendimų parėmimui nei sprendimų priėmimui [41]. Priklausomybių analitika (angl. descriptive analytics) atsako į klausimą: kas įvyko [22]? Priklausomybių duomenų tyrybos uždavinių sričiai priskiriami uždaviniai, kurie nusako duomenų bendrąsias savybes, struktūrą iš anksto jos nežinant. Dažnų šablonų radimas taip pat priskiriamas priklausomybių duomenų tyrybai, jis padeda nustatyti ryšius bei koreliacijas tarp duomenų [40]. Dauguma įmonių naudoja priklausomybių analizę ir dar tik pradeda naudoti prognozių analizę [22].

Duomenų tyryba yra viena iš mašininio mokymosi rūšių [42]. Mašininis mokymasis apibrėžiamas kaip sistemos apmokymas iš duotų duomenų, užuot jį tiksliai užprogramavus. Mašininis mokymasis skiriasi nuo duomenų tyrybos tuo, kad sistemai padavus duomenis ir atsakymus gaunamos taisyklės, kurios vėliau gali būti naudojamos naujiems duomenims apdoroti ir problemoms spręsti, o duomenų tyryba analizuoja jau esamus duomenis [43]. Mašininį mokymąsi, taip pat ir jam priklausančius algoritmus galima skirti į dvi pagrindines kategorijas: sistemų mokymąsi su mokytoju (angl. supervised learning) ir be mokytojo (angl. unsupervised learning) [40]. Jeigu tiriamų duomenų rinkinyje įvesties elementai turi atsako kintamąjį, kuriam priskiriama žymė (angl. label) su teisinga išvestimi, tada gali būti taikomas sistemos mokymasis su mokytoju. Priešingu atveju, jei duomenų rinkinys neturi informacijos apie teisingą išvestį, atliekamas mokymasis be mokytojo [44]. Mokymuisi be mokytojo priskiriami taisyklių išgavimo, klasterizavimo, dimensionalumo mažinimo (pvz.: vizualizavimui) ir teksto tyrybos uždaviniai. Mokymasis su mokytoju apima klasifikavimo bei regresijos uždavinius [45].

### **2.3. Teksto tyryba**

Dauguma verslo įžvalgų šiandien slypi nestruktūrizuotuose ar pusiau struktūrizuotose duomenų rinkiniuose. Aštuoniasdešimt procentų internetinių duomenų yra teksto pavidalo, tai: naujienų straipsniai, moksliniai straipsniai, tinklaraščiai, el. laišakai, klientų atsiliepimai, socialinės žiniasklaidos duomenys (pvz.: „Facebook“ ar „Tweeter“ skelbimai) [46]. Informacijos srautams augant eksponentiškai, informacijos apdorojimo ir vertingų žinių išgavimo procesas darosi vis labiau komplikotas. Dėl didelio duomenų augimo greičio darosi vis sunkiau be papildomų įrankių išanalizuoti visą prieinamą informaciją ir atskirti aktualią informaciją nuo nereikšmingos [46]. Šiai problemai spręsti skirta teksto tyryba (angl. text mining). Teksto tyryba apibrėžiama kaip tekstinių duomenų rinkinio analizė, kuria siekiama nustatyti pagrindines tekste vyraujančias temas, sąvokas, atskleisti paslėptus ryšius bei tendencijas, nereikalaujant iš anksto žinoti, kokius žodžius ar terminus autoriai naudojo šioms idėjoms išreikšti [48].



Apie teksto tyrybos poreikį kalbama ir audito srityje [25, 33, 49]. Atliekami finansinės atskaitomybės tekstinės informacijos tyrimai siekiant nustatyti sukčiavimą, tiriamas apskaitos įrašų (angl. journal entries) pobūdis [49]. Kuriami tekstu paremti įrankiai, kurie apmokomi žinomais įmonių sukčiavimo atvejais, ieško įmonių ataskaitose žodžių galinčių išduoti sukčiavimą [8]. Kiekybinis modeliavimas ir akcijų rinkos prognozavimas, remiantis internetine tekstine informacija, bei sentimentų analizė (angl. sentiment analysis) tai pat yra aktyvi mokslinių tyrimų sritis, kuri tiria didžiųjų duomenų metodų naudą auditui [8]. Standartų nuostatai reikalauja, kad auditoriai įvertintų klientų verslo operacijų atitiktį galiojantiems reguliavimams. Ši procedūra reikalauja labai daug laiko, o dabartiniai kompiuterizuoti audito įrankiai yra nepajėgūs apdoroti tekstinių dokumentų automatinio būdu [50]. Informacinių sistemų ir verslo procesų auditas suteikia vertingų žinių apie verslo valdymą ir reikalavimų laikymąsi. Atsižvelgiant į tai auditoriai turi ištirti verslo operacijas atspindinčią informaciją. Šią informaciją galima išgauti iš atliktų sandorių duomenų, vidaus taisyklių, bendrojo žurnalo įrašų, sutarčių, metinių pranešimų [50]. Tačiau tai reikalauja atlikti teksto tyrybą.

Dauguma duomenų tyrybos įrankių naudoja tik struktūrizuotus, specialiai analizei paruoštus duomenis su tolydziais arba kategoriniais kintamaisiais. Todėl norint analizuoti tekstinius duomenis turime juos paversti skaitine išraiška [47]. Dažniausiai nagrinėjant tekstinio formato failus pirmas ir svarbiausias uždavinys yra juos suvesti į tekstinių dokumentų matricą, o vėliau šiai matricai jau galima taikyti tradicinius duomenų analizės metodus.

Knygoje [47] tekstas yra apibūdinamas, kaip nestruktūrizuotų dokumentų rinkinys, kurių sudarymui nėra jokių specialių reikalavimų. Įsivaizduokime, kad turime dokumentų rinkinį. Tokiu atveju, požymių rinkinį sudarys visų unikalių žodžių esančių dokumentuose rinkinys. Visų dokumentų požymių rinkinys gali būti sudarytas iš žodžių, n-gramų, kolokacijų ir kitų darinių, kurių pagalba galima nagrinėti tekstą. N-gramomis yra laikoma žodžių (būtinai gretimų) seka, sudaryta iš n žodžių, pavyzdžiui, žodžių junginys, sudarytas iš dviejų žodžių yra vadinamas bigrama. Kolokacija – tai žodžių, nebūtinai gretimų seka, kurie pasirodo dažniau nei atsitiktinai.

Tačiau prie pradant nagrinėti tekstą yra keletas svarbių teksto apdorojimo žingsnių, kurie yra būtini siekiant efektyviai atlikti teksto analizę.

Garis Mineris ir kiti išskiria tokius pagrindinius teksto tyrybos žingsnius [51]:

- 1) pasirenkama teksto, kuris turės būti nagrinėjamas apimtis (dokumentas, paragrafas ir pan.);
- 2) tekstas išskaidomas į žodžius. Žodžiai, kiekvienas atskirai, interpretuojami kaip atskiri teksto vienetai – leksemos (angl. tokens);

- 3) iš teksto pašalinami dažnai pasikartojantys bereikšmiai žodžiai (angl. stopwords). Pavyzdžiui: tai, kad, bet ir pan.;
- 4) vykdomas žodžių kamieno išskyrimas (angl. stemming);
- 5) pažymimos sakinių pabaigos;
- 6) visos raidės paverčiamos mažosiomis arba atvirkščiai – didžiosiomis raidėmis.

Pirmiausia svarbu pasirinkti tinkamus šaltinius teksto analizei. Dažniausiai teksto, kuris bus nagrinėjamas apimtį nustatyti nesunku, pavyzdžiui, tiriant el. laiškus, kiekvienas laiškas yra atskiras vektorius. Tiriant ilgesnius dokumentus reikia nuspręsti ar bus nagrinėjamas visas tekstas, ar išskaidytas į dalis, paragrafus ar sakinius. Tuomet, iš pradinių dokumentų pašalinami tarpai, visi skyrybos ženklai, žodžiai ar kiti simboliai, kurie yra nereikšmingi, nieko nepasako apie dokumento turinį, dažniausiai tai įvairūs funkciniai žodžiai: prielinksniai, dalelytės, jungtukai ir pan. Labai svarbus punktas žodžių kamienų išskyrimas ir lemavimas (angl. lemmatization), šie du procesai apjungia žodžius, kurie yra pavartoti skirtingomis formomis ir palieka tik pagrindinės formos žodį, pavyzdžiui „kalbos“, „kalbai“ ir „kalba“ apjungiami į pagrindinę formą „kalba“ [51].

Atlikus visus minėtus punktus atskiros leksemos turi būti transformuotos į vektorinį pavidalą, kuris būtų suprantamas teksto tyrybos algoritmams. Knygoje [51] išskiriami trys vektorinio atvaizdavimo būdai: binarinis atvaizdavimas, požymių dažnumo, svartinio įvertinimo (angl. float-valued weighted). Vienas paprasčiausių ypatybių atspindėjimo metodų yra binaris atvaizdavimas, kuriuo 0 ir 1 vertinamas ypatybės buvimas/nebuvimas tiriamame tekste priskiriant 0, jei ypatybės nėra, ir 1, jei ypatybė yra. Renkantis požymių dažnumo atvaizdavimą vektoriuje saugomas požymių dažnis dokumente, ty. jeigu žodžio nėra saugomas 0, o jeigu žodis pasikartoja dokumente vieną ar daugiau kartų atitinkamai saugomas žodžio pasikartojimų skaičius. Naudojant svartinio įvertinimo atvaizdavimą, pirmiausia reiktų pasirinkti, koks svorinis koeficientas (angl. weighting scheme) bus naudojamas. Kaip populiariausias išskiriamas sąvokos atvirkštinio dažnio požymių atrinkimo (angl. term frequency-inverse document frequency) metodas. Šio metodo esmė yra tokia, kad kuo dažniau žodis pasikartoja, tuo didesnis svoris jam priskiriamas, tačiau, jei šis žodis dažnai kartojasi ir kituose dokumentuose, jo svoris yra mažinamas [51].

Žinant požymių rinkinį, jau galima formuoti dokumentų matricą, kurioje eilutės atitinka dokumentus, o stulpeliai požymius. Sudarius tekstinę dokumentų matricą galima naudoti įprastus statistinės analizės metodus bei taikyti kompleksinius analizės metodus tokius kaip klasterizavimas ar klasifikavimas, taip pat atlikti analizę pasitelkiant įvairių sričių žodynus.

## 2.4. Teksto klasifikavimas

Teksto klasifikavimo problema apibrėžiama, kaip matematinio modelio, galinčio reprezentuoti teksto semantiką ir būti naudojamu pakartotiniams palyginimams, paieška [52]. Tyrimų tikslas yra algoritmo, galinčio priskirti tekstą tam tikrai kategorijai su maksimaliu tikslumu išvengiant priskyrimo daugeliui kategorijų, radimas. Teksto klasifikavimo uždavinio esmė yra tekstą pagal jo turinį priskirti vienai iš kelių iš anksto žinomų kategorijų [52].

Klasifikavimo algoritmo kokybei vertinti naudojama keletas skirtingų būdų, kurių pasirinkimas priklauso nuo paties algoritmo, sprendžiamo uždavinio tikslų, duomenų prigimties ir pan. Praktinius tyrimus aprašančiuose darbuose dažniau vertinami ne algoritmų vidutiniai nuostoliai, o vidutinis algoritmų tikslumas (angl. accuracy) [53]. Klasifikavimo tikslumas skaičiuojamas taip:

$$\text{tikslumas} = \frac{|\{\text{teisingai suklasifikuoti testavimo duomenys}\}|}{|\{\text{visi testavimo duomenys}\}|} \quad (1)$$

Klasifikavimo kokybei vertinti tai pat naudojamas atpažintų arba klasifikuotų objektų kiekis (angl. recall). Atpažintų objektų kiekis apibrėžiamas kaip teisingai klasifikuotų objektų ir visų atpažintų objektų kiekio santykis.

Egzistuoja daug teksto klasifikavimo algoritmų. Toliau plačiau aptariami Naivaus Bajeso (angl. naive Bayes) ir atraminių vektorių klasifikavimo algoritmai (angl. Support Vector Machines).

### 2.4.1. Naivaus Bajeso algoritmas

Bajeso klasifikatorius (angl. Bayes classifier) yra populiarus dėl to, kad yra paprastas tikimybinis klasifikavimo metodas, kuris dažniausiai veikia labai gerai ir greitai [53, 54]. Jis priskiria labiausiai tinkamą klasę dokumentui, kuris yra aprašomas kaip savybių vektorius. Konstruojant Naivaus Bajeso modelį naudojama prielaida, kad visų dokumentų visos savybės yra tarpusavyje nepriklausomos. Nors ši prielaida dažnai neturi pagrindo realiame gyvenime algoritmas įprastai veikia gana gerai. Dėl nepriklausomybės prielaidos kiekvieno atributo parametrus galima apmokti atskirai, o tai labai supaprastina mokymą, ypač kai atributų skaičius yra didelis [54].

Tikimybiniai algoritmai tiesiogiai vertina aposteriorines tikimybes ir šiuos įverčius naudoja klasių nustatymo taisyklėje. Tegul stebimas atsitiktinis dokumentas, kurį reprezentuojantis atsitiktinis požymių vektorius  $X$  įgyja konkrečią reikšmę  $x$ . Tegul šis dokumentas priklauso nestebimai klasei  $\eta \in K$ ,  $d = \dim(X)$ . Naudodamiesi Bajeso teorema, galime užrašyti [53]:

$$P\{\eta = w | X = x\} = \frac{P\{\eta = w\}P\{X = x | \eta = w\}}{P\{X = x\}} \quad (2)$$

Naivaus Bajeso algoritmas remiasi prielaida, kad vektoriaus  $X$  komponentės yra sąlyginai nepriklausomos prie sąlygos  $\eta = w$ , todėl galioja [53]:

$$P\{X = x|\eta = w\} = \prod_{i=1}^Y P\{X_i = x_i |\eta = w\} \quad (3)$$

Čia dešinėje pusėje dauginamos tikimybės, kad atsitiktinai parinkto dokumento, priklausančio klasei  $w$ , atsitiktinai parinktoje pozicijoje  $i$  stebima konkreti reikšmė  $x_i$  [53].

Nors yra nemažai darbų patvirtinančių patobulintų ir sudėtingesnių metodų pranašumą prieš paprasčiausią Naivaus Bajeso klasifikatorių, tačiau neretai šis pranašumas įgyjamas tik tam tikrose specifinėse situacijose, o stebimas tikslumo padidėjimas neatperka skaičiavimų sudėtingumo išaugimo [53].

#### 2.4.2. Atraminų vektorių algoritmas

Atraminų vektorių algoritmas (toliau SVM) kaip ir Naivaus Bajeso priklauso mokymo su mokytoju (angl. supervised learning) klasifikavimo metodams, tačiau jis vietoje empirinės apsirikimo rizikos (klasifikavimo klaidų mokymo imtyse lygis) mažinimo konstruoja klasifikavimo taisyklę taip, kad būtų minimizuojama tikroji klasifikavimo klaida, t. y., tikimybė, kad klasifikatorius suklys, sutikęs nematytą testavimo imties pavyzdį. Paprasčiausiame pavidale metodas yra tiesinis savo parametrų atžvilgiu, be to, skirtas spręsti paprastą vienos klasės uždavinį – teisingai atskirti duomenis, priklausančius klasei nuo nepriklausančių [53, 55].

Šis metodas gali būti naudojamas ne tik dviejų klasių klasifikavimui, jis gali būti išplėstas ir iki  $n$  klasių. Tuomet reikės  $n$  hiperplokštumų norint padalinti vektorinę erdvę dokumentų klasifikavimui.

Sakykime, kad turime apmokymo imtį, kuri apibrėžiama taip [56]:

$$(x_1, y_1), \dots, (x_l, y_l), x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}, i \in [1, \dots, l].$$

Apibrėžiame SVM klasifikatoriaus funkciją:

$$g(x) = \text{sign}(f(x)) \quad (4)$$

$$f(x) = \sum_{i=1}^l y_i \alpha_i K(x_i, x) + b, \quad (5)$$

Čia  $K$  yra branduolio funkcija;  $b \in \mathbb{R}$  ir yra slenkstis;  $\alpha_i$  yra svoriai. Be to svoriai  $\alpha_i$  tenkina tokias nelygybes [56]:

$$\forall i: 0 \leq \alpha_i \leq C \text{ ir } \sum_{i=1}^l \alpha_i y_i = 0,$$

čia  $C$  yra neteisingo klasifikavimo kaina. Vektoriai  $x_i$  su ne nuliniu  $\alpha_i$  yra vadinami atraminiais vektoriais (angl. support vectors). Tiesiniam SVM branduolio funkcija  $K$  apibrėžiama taip [56]:

$$K(x_i, x) = x_i \cdot x \quad (6)$$

Tokiu atveju, galime perrašyti funkciją  $f(x)$  taip:

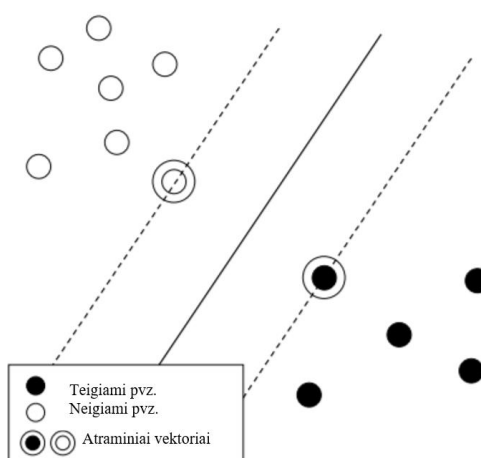
$$f(x) = w \cdot x + b, \quad (7)$$

čia  $w = \sum_{i=1}^l \alpha_i y_i x_i$ .

Apmokyti SVM reiškia rasti  $\alpha_i$  ir  $b$  sprendžiant šią optimizacijos problemą [56]:

$$\begin{cases} \max \left( \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, y_j) \right); \\ \forall i: 0 \leq \alpha_i \leq C; \\ \sum_{i=1}^l \alpha_i y_i = 0. \end{cases} \quad (8)$$

Sprendimas duoda optimalią hiperplokštumą, kuri yra klasifikavimo pasirinkimo kraštas tarp dviejų klasių [56]. 7 pav. vaizduoja optimalią hiperplokštumą ir atraminius vektorius.



pav. 7 hiperplokštuma ir atraminiai vektoriai pagal [56]

## 2.5. Laiko eilučių analizė

Finansuose laiko eilučių analizė bei prognozavimas yra nepaprastai svarbus uždavinys. Tai viena pagrindinių metodikų, leidžiančių numatyti verslo ateitį. Laiko eilute, pagal Stabingienės [57] apibrėžimą yra laikoma „analizuojamo atsitiktinio dydžio  $\zeta$  t stebėjimų, gautų laiko momentais  $t=1, \dots, T$  eilutė  $X_1, X_2, \dots, X_T$ “. Pagal [58]: „visų prognozės modelių pagrindinis tikslas yra tiksliau ar paprasčiau rekonstruoti laiko eilutę generuojančio proceso modelį iš praeities duomenų, ir tuomet jį panaudoti paskutinių laiko eilutės reikšmių ekstrapoliavimui į ateitį“. Tiriant laiko eilutes, laikoma, kad yra žinomos reikšmės  $Z(t_i)$  laiko momentais  $t_1 < t_2 < \dots < t_n$ , o visi stebėjimai atliekami vienodais laiko intervalais, t.y.  $t_{i+1} - t_i = \Delta t$ .

### Laiko eilučių stacionarumas

Laiko eilutės stacionarumo įvertinimas - vienas pirmųjų eilučių analizės uždavinių. Nuo laiko eilutės stacionarumo priklauso vidurkio funkcijos pavidalas. Stacionariame procese laiko eilutės reikšmės kinta atsitiktinai kiekvienu momentu, tačiau vidurkis gana ilgą laiką nekinta [57].

Laiko eilutė  $X_1, X_2, \dots, X_T$  vadinama stacionariąja, jeigu ją atitinkantis procesas  $(X(t), t \in I)$  yra stacionarus plačiąja prasme procesas. Turime skirstinį:

$$P_{t_1, t_2, \dots, t_n} = P_{t_1+h, t_2+h, \dots, t_n+h}, \quad (9)$$

su visais  $t_1 < t_2 < \dots < t_n \in Z$  ir  $h \in Z$ .

Laiko eilutė yra stacionari, jei [57, 59]:

- laiko eilutės vidurkis nepriklauso nuo laiko:

$$E(X_t) = \mu_x = \text{const}_1 \quad (10)$$

- laiko eilutės dispersija nepriklauso nuo laiko:

$$E(X_t - \mu_x)^2 = \sigma_x^2 = \text{const}_2 \quad (11)$$

- laiko eilutės stebėjimų kovariacinė funkcija priklauso nuo laiko skirtumo  $k$ :

$$E[(X_t - \mu_x)(X_{t-k} - \mu_x)] = \text{cov}(X_t, X_{t-k}) = \gamma_k \quad (12)$$

Svarbu iširti kiekvieno kintamojo integravimo tvarką, siekiant nustatyti ar laiko eilutė yra nestacionari ir kiek kartų kintamąjį reiks diferencijuoti, kol gausime stacionarią laiko eilutę.

*Dikio – Fulerio* (angl. *Dickey – Fuller*) testas skirtas patikrinti, ar laiko eilutė turi vienetinę šaknį. Tai vienas populiariausių testų taikomų patikrinti stacionarumo hipotezei. Vienetinė šaknis yra laiko eilutės autoregresijos parametras, lygus 1. Jei laiko eilutė turi vienetinę šaknį, ji nėra stacionari.

Tarsime, kad proceso liekanos yra AR procesas. Kitaip sakant, tiriamasis procesas yra AR(p):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (13)$$

**$H_0$ :** procesas  $X_t$  turi vienetinę šaknį,

**$H_a$ :** procesas  $X_t$  yra stacionarus.

Procesas perrašomas tokiu pavidalu:

$$\Delta X_t = (\alpha - 1)X_{t-1} + \sum_{i=2}^p \beta_i \Delta X_{t-i+1} + a_t; \quad (14)$$

čia  $p$  – vėlavimų skaičius;  $a_t$  – baltasis triukšmas.

Žymima  $1 - \alpha = \gamma$ . Vienetinės šaknies testas vykdomas nulinei hipotezei  $\gamma = 0$  su stacionarumo alternatyva, kad  $\gamma < 0$ . Tada testo statistika:

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}; \quad (15)$$

čia  $SE(\hat{\gamma})$  – dydžio  $\hat{\gamma}$  standartinė paklaida;

Statistika (12) yra lyginama su atitinkama *Dikio – Fulerio* testo, kritine reikšme.

Jei hipotezė  $H_0$  priimama, tai daroma išvada, kad  $X_t$  nėra stacionarus, prieinama prie analogiško  $(I-L)X_t$  nagrinėjimo. Jei gaunama išvada, kad ir  $(I-L)X_t$  nestacionarus, tiriama antros eilės pokyčiai  $(I-L)^2 X_t$ .

## Laiko eilutės dekompozicija

Paprastai pirmasis žingsnis duomenų analizėje yra grafinis pavaizdavimas. Duomenų vizualizacija leidžia įtarti koks modelis turėtų būti taikomas. Populiariausias yra išskaidymas į trendo, sezoniskumo ir triukšmo komponentių sumą, jeigu procesas yra stacionarus ir daugybą nestacionaraus proceso atveju [59]:

$$X_t = m_t + s_t + Z_t \text{ arba } X_t = m_t \cdot s_t \cdot Z_t, \quad (16)$$

Čia  $m_t$  yra trendo komponentė;  $s_t$  yra periodinė funkcija su žinomu periodu  $d$ ;  $Z_t$  yra atsitiktinė triukšmo komponentė, matematiškai aprašoma stacionaraus proceso pagalba.

Laiko eilutės dekompozicija išskaido dinamiką į tokius komponentus [57, 59]:

- tiesinio **trendo** komponentas (angl. level, trend) – tai ilgalaikis augimas/kritimas, kuris pastebimas visoje laiko eilutėje. Trendas gali būti ir netiesinis (pavyzdžiui, logaritminis, eksponentinis ir pan.). Beje, trendo gali ir nebūti (tiksliau, jis gali būti lygus konstantai);
- **sezoninių svyravimų** komponentas (angl. seasonal) – reguliarus stebimo kintamojo reikšmių didėjimas bei mažėjimas griežtai apibrėžtais laiko periodais. Sezoniniai svyravimai yra identifikuojami lengviausiai iš visų sudėtinių laiko eilutės dedamųjų. Jau iš grafinio duomenų vaizdavimo paprastai galima nustatyti, ar tokie svyravimai egzistuoja, o jeigu taip, koks yra jų periodas. Paprastai prieš atliekant svyravimų skaičiavimus, iš laiko eilutės reikšmių yra atmetamas trendas. Tuomet gauti rezultatai iliustruoja sezoninius atsilenkimus nuo trendo;
- **ciklinio svyravimo** (verslo arba ekonomikos) – yra panašūs į sezoninius, tačiau neturi tokio griežto matematinio aprašymo; jų pasikartojimo periodas nėra toks apibrėžtas, o kalbant apie istorinius procesus - paprastai gerokai ilgesnis. Jeigu sezoninis svyravimas dažnai sutapatinamas su stebimo kintamojo reikšmių svyravimu, susijusiu su metų sezonų ar dar trumpesnių laiko intervalų - savaitių, parų, valandų ir t.t. kaita, tai ciklinis svyravimas istorijoje paprastai apima dešimtmečius, šimtmečius, tūkstantmečius ar dar ilgesnius laikotarpius;
- **klaidos** komponentas (liekanos) – tai nesumodeliuoti, nepaaiškinami svyravimai, tai ko nepavyksta sumodeliuoti trendu ir sezoniskumu.

Šie komponentai gali būti visi, gali kai kurių (o gal ir visų, jei laiko eilutė griežtai atsitiktinė) nebūti.

## Grangerio priežastinio ryšio nustatymo testas

Grangerio priežastingumas (angl. Granger causality) yra populiarus testas siekiant nustatyti priežastinius ryšius tarp atsitiktinių kintamųjų. Jis ne tik nustato ryšio egzistavimą ir jo kryptis, bet ir priežastingumą, t. y. atsako į klausimą, kuris iš nagrinėjamų veiksnių yra priežastis, o kuris pasekmė.

Grangerio priežastingumo testas laiko eilutėms remiasi prielaida: jeigu  $x$  įtakoja  $y$ , tai prieš  $y$  pokyčius turėtų įvykti  $x$  pokyčiai, bet ne atvirkščiai [60].

Kitaip tariant, turėtų būti išpildytos dvi sąlygos [60]:

- a)  $x$  turėtų įnešti statistiškai reikšmingą indėlį į  $y$  prognozę;
- b)  $y$  neturėtų įnešti statistiškai reikšmingo indėlio į  $x$  prognozę.

Tikrinant Grangerio priežastingumą, sudaromos regresijos lygtys:

$$y_t = a_0 + \sum_{i=1}^m a_i y_{t-i} + \sum_{i=1}^m \beta_i x_{t-i} + \varepsilon_t, \quad (17)$$

$$x_t = a_0 + \sum_{i=1}^m a_i x_{t-i} + \sum_{i=1}^m \beta_i y_{t-i} + u_t, \quad (18)$$

kur  $\varepsilon_t$  ir  $u_t$  yra nekoreliuotos atsitiktinės paklaidos.

Kiekvienai regresijos lygčiai skaičiuojama F statistika pagal formulę, kuri vėliau lyginama su kritine Fišerio skirstinio reikšme:

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n-k-m-1)} \quad (19)$$

Ir tikrinamos hipotezės:

$$\begin{cases} H_0: \beta = 0 \\ H_A: \beta \neq 0 \end{cases} \quad (20)$$

Jei  $F > F_{k..n-k-m-1}$ , tai  $H_0$  hipotezę atmetame tai reiškia, kad  $x_t$  daro įtaka  $y_t$ . Jei  $F < F_{k..n-k-m-1}$ , tai  $H_0$  hipotezės atmesti negalime ir teigiame, jog  $x_t$  nedaro įtakos  $y_t$  su pasirinktu reikšmingumo lygmeniu  $\alpha$ .

Norint nustatyti vėlinimų (toliau. lagų) ilgį tarp tiriamų kintamųjų naudojamas vektorinės autoregresijos (VAR) modelis. VAR modelis vadinamas todėl, kad autoregresijos terminas vartojamas dėl laginių kintamųjų reikšmių, o vektorinis – nes aprašomas laiko eilučių vektorius [61].

Kaip ir Grangerio modelyje atliekant VAR skaičiavimus, sudaromos regresijos lygtys:

$$Y_t = \alpha_1 + \delta_1 t + \sum_{j=1}^p \theta_{1j} y_{t-j} + \sum_{i=1}^m \beta_{1j} x_{t-j} + \varepsilon_{1t} \quad (21)$$

$$X_t = \alpha_2 + \delta_2 t + \sum_{j=1}^p \theta_{2j} y_{t-j} + \sum_{i=1}^m \beta_{2j} x_{t-j} + \varepsilon_{2t} \quad (22)$$

Šioje dviejų kintamųjų lygčių sistemoje pirma lygtis nurodo, kad  $X$  yra  $Y$  kitimo priežastis, o antra – atvirkščiai  $Y$  yra  $X$  kitimo priežastis. Parametrų  $\theta$  ir  $\beta$  pirmasis indeksas nurodo, kuriai lygčių sistemos lygčiai yra priskiriamas. Analogiškai yra ir parametrams  $\alpha$  ir  $\delta$ . Pabrėžtina, kad lagų skaičius  $p$  yra



vienodas visiems kintamiesiems.  $\varepsilon_{1t}$  ir  $\varepsilon_{2t}$  – tai atsitiktinės paklaidos, kurios vadinamos impulsais [61].

Norint įvertinti šią sistema reikia nusistatyti lagų skaičių  $p$  ir maksimalų ligo ilgį  $p_{\max}$ . Parinkus didelį  $p$ , sumažėja laisvės laipsnių skaičius, o parinkus mažą – padidėja liekamosios paklaidos tikimybė. Jei naudojamas labai didelis  $p_{\max}$ , gali būti reikalinga ilga bandymų seka, kuri įtakos testuojamos sekos bendrą I tipo paklaidą, tai yra,  $p_{\max}$  pasirinkimas įtakos nepakankamą  $p$  atrankos tikimybę. Dažniausiai dydžiui  $p$  nustatyti naudojamosi Akaikės (*AICC*) kriterijumi [61]:

$$AIC(m) = \log \det(\Sigma_{\tilde{u}}(m)) + \frac{2}{T} mK^2, \quad (23)$$

Kur  $\varphi(m) = mK^2$  ir  $c_T = \frac{2}{T}$ . Parinkus  $p$ , VAR modelis žymimas VAR( $p$ ).

### 2.5.1. Vienmačiai laiko eilučių prognozavimo metodai

Šiame darbe finansinių rodiklių prognozavimui bus naudojama 13 vienmačių prognozavimo metodų ar jų modifikacijų. Toliau trumpai apžvelgiamas kiekvienas iš jų.

#### *Tiesinės regresijos modelis (angl. linear regression model)*

Paprastas tiesinės regresijos metodas: dominančio dydžio prognozė gaunama aiškinant jo kitimo priežastį kito nepriklausomo kintamojo kitimu [58]. Regresinėje analizėje tiriama vieno kintamojo priklausomybė nuo vieno ar daugiau kitų kintamųjų, siekiant įvertinti ar prognozuoti vėlesnes populiacijos vidutines reikšmes, žinant imties ankstesnes reikšmes. Duomenų seka išdėstoma grafiškai, paprastai tiesėje ir užrašoma tokiu pavidalu:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (24)$$

čia  $Y$ -priklausomas kintamasis,  $X$ - nepriklausomas kintamasis,  $\beta_0$  ir  $\beta_1$  konstantos,  $\varepsilon$  – atsitiktinė paklaida.

Norint regresijoje atsižvelgti į sezoniskumą ir/arba tendą galima įtraukti *dummy* kintamuosius.

#### *Sezoninis naivus metodas (angl. Seasonal naive method)*

Naivusis arba vadinamas paprastasis prognozavimo metodas, tai pats paprasčiausias metodas. Naivusis prognozavimo metodas panaudoja savo prognozėje pačią paskutinę faktinę vertę, kuri panaudojama kaip kito laikotarpio prognozė [63]. Kai duomenys pasižymi sezoniniais svyravimais, paprastasis naivusis metodas ne visada tinka, kadangi jame ignoruojamas sezoninė komponentė. Tuomet yra taikomas sezoninis-naivusis prognozavimo metodas, kur yra numatoma sezoniskumo galimybė, nors šis metodas išlieka labai paprastas. Šiuo atveju mes nustatome, kad kiekviena

prognozė yra lygi paskutinei stebėtai vertei iš to paties metų sezono (pvz., praėjusių metų tą patį mėnesį) [63].

Sezoninis naivus/*ARIMA*(0,0,0)(0,1,0)[12] modelis atrodo taip:

$$Y_t = Y_{t-12} + e_t, \quad (25)$$

kur  $e$  yra paklaida.

### ***Vidurkis (angl. mean)***

Taikant vidurkį prognozavimui priimama, kad visų ateities verčių prognozės yra lygios istorinių duomenų vidurkiui.

### ***ARIMA metodas***

*ARIMA* (angl. Auto Regressive Integrated Moving-Average) - auto regresinis integruoto slenkančio vidurkio modelis, kartais dar vadinamas *Box&Jenkins* modeliu. *ARIMA* modelis kilo iš autoregresijos (AR) modelio, slenkančio vidurkio (MA) modelio ir AR bei MA kombinuoto (*ARMA*) modelio [58]. *ARIMA* modelis prognozuoja dydį pagal laiko eilutės tiesinę kombinaciją su jos praeities dydžiais, praeities paklaidomis ir dabarties bei praeities kitų laiko eilučių įverčiais. *ARIMA* metodo esmė – sujungti autoregresijos, diferencijavimo ir slenkamųjų vidurkių metodo galimybes.

Pagal [62] procesas  $X_t$ , tenkinantis išraišką:

$$\Phi(L)(1 - L)^d X_t = c + \Theta(L)\varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad (26)$$

$$\text{kur } \Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p,$$

$$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q.$$

vadinamas  $p, q$ -eilių autoregresiniu ir slenkamųjų vidurkių procesu integruotu  $d$  eile bei žymimas *ARIMA*( $p, d, q$ ) čia  $p$  – autoregresijos eilė;  $d$  – diferencijavimo eilė;  $q$  -slenkamųjų vidurkių eilė, o  $c, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2$  parametrai yra pastovūs.

### ***Basic structural modelis***

*Basic structural* modelio sudarymas remiasi laiko eilutės skaidymu į komponentus (dekompozicija). Kiekvienam komponentui skaičiuojama nuosava sklaida (pasitaiko, kad kai kuriems komponentams sklaida gali būti lygi 0). *Basic structural* modelis yra tas pats *ARIMA*(0, 1, 1) modelis tik su apribojimais parametru įverčiams.

### ***Atsitiktinio klaidžiojimo (angl. Random Walk) procesas***

Atsitiktinio klaidžiojimo (angl. Random Walk) procesas  $Y_t$  tenkina išraišką [63]:

$$Y_t = Y_{t+1} + \varepsilon_t \sim WN(0, \sigma_\varepsilon^2). \quad (27)$$

Parametras  $\sigma_\varepsilon^2$  yra pastovus,  $\varepsilon_t$ - baltasis triukšmas (žymima  $\varepsilon_t \sim WN$ , angl. white noise). Baltojo triukšmo savybės, pagal kurias, jis yra stacionarus procesas:

- vidurkis  $\mu_\varepsilon = E(\varepsilon_t) = 0$ ,
- dispersija  $\gamma_0 = E(\varepsilon_t^2) = \sigma_\varepsilon^2$ ,
- kovariacijos  $\gamma_k = 0$ .

Jei laiko eilutės turi kylančią arba besileidžiančią kryptį (trendą), kuri tikėtina, kad tęsis ir ateityje, į šį modelį reikia įtraukti konstantą nelygią nuliui. Toks procesas tuomet vadinamas atsitiktinio klaidžiojimo su dreifu procesu (angl. random walk with drift). Atsitiktinis klaidžiojimas su dreifu yra atskiras atvejis autoregresinio integruoto slenkančio vidurkio modelio *ARIMA*, t.y. *ARIMA*(0,1,0) su konstanta.

### ***Atsitiktinio klaidžiojimo su dreifu procesas***

Atsitiktinio klaidžiojimo su dreifu procesas (angl. *Random Walk with Drift*)  $Y_t$  tenkina išraišką [64]:

$$Y_t = c + Y_{t+1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \quad (28)$$

Abu parametrai  $c$  ir  $\sigma_\varepsilon^2$  yra pastovūs. Jei nuliniu laiko momentu pradinė proceso reikšmė yra  $y_0$ , tai proceso momentai:

- $\mu_\varepsilon = y_0 + ct$ ,
- $\gamma_0 = t\sigma_\varepsilon^2$ ,
- $\gamma_k = (t - k)\sigma_\varepsilon^2$ ,
- autokoreliacijos koeficientas  $\rho_k = \sqrt{1 - k/t}$ .

Atsitiktinio klaidžiojimo su dreifu procesas yra nestacionarus, nes jo vidurkis, dispersija ir kovariacijos priklauso nuo laiko.

### ***Theta metodas***

Laiko eilutės  $X_t$  yra transformuojamos į  $Y_{t,\theta}$  ir modelio lygtis atrodo taip [63]:

$$Y_{t,\theta} = a_\theta + b_\theta(t - 1) + \theta X_\theta, \quad (29)$$

kur  $t=1, \dots, n$ ;  $(t-1)$  yra tiesinis trendas,  $a$  yra konstanta, o  $b$  – trendo koeficientas.

Čia  $a$  ir  $b$  reikšmės yra randamos mažiausių kvadratų metodu.

Mėnesiniai ir ketvirtiniai duomenys pirmiausiai pritaikomi sezoniškumui, išskiriant sezoniškumo komponentą, naudojant paprastą dauginamąjį dekomponavimą. Tuomet sezoniškumo komponentas padauginamas iš *Theta* metodu sugeneruotos prognozės.

### ***BATS - ARMA ir eksponentinio glodinimo hibridas***

*BATS* yra modifikuotas modelis, susidedantis iš kelių metodų. *BATS* yra akronimas [65]:

- B – *Box-Cox* transformacija,
- A – *ARMA* paklaidos,
- T – trendas,
- S – sezoniškumo komponentai.

*BATS* modelis yra tradicinių modelių su sezoniškumu generalizacija, leidžianti daugiau sezoninių periodų.

Jis susideda iš  $w, \Phi, p, q, m_2, \dots, m_T$  argumentų, kurie indikuoja *Box-Cox* parametą, slopinimo (angl. damping) parametą, *ARMA* parametrus  $p$  ir  $q$  ir sezoninius periodus  $m_1 - m_t$ .

Naudojant *BATS* metodą, laiko eilutės transformuojamos *Box-Cox* metodu:

$$Y = Y^{\lambda-1}/\lambda \quad (30)$$

Toliau transformuotai laiko eilutei naudojamas *ARMA* modelis su trendu ir sezoniškumu. Gavus prognozę, ji transformuojama atgal.

### ***ETS (eksponentinio glodinimo generalizacija)***

Taikant eksponentinio glodinimo metodą būsima prognozė remiasi praėjusio laikotarpio faktiniu duomenų sekos dydžiu bei praėjusio laikotarpio prognozės suma, kai šie abu dydžiai yra įvertinti specialių koeficientų pagalba. Tai specialus metodas, kuriame imami visi praeities stebėjimai, tačiau prognozei didesnę reikšmę turi paskutiniai stebėjimai laiko intervale nei įvykę daug anksčiau. Jeigu laiko eilutė turi trendą, taikomas dvigubas eksponentinio išlyginimo metodas. Formaliai ši procedūra užrašoma taip [64]:

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad (31)$$

$$S_{t-1} = \alpha S_t + (1 - \alpha)S_{t-1}. \quad (32)$$

Šiose lygtyse egzistuoja tik vienas modelio suderinimo parametras -  $\alpha$ . Praktikoje šis modelis vadinamas *Brauno* tiesinio eksponentinio išlyginimo metodu.

### ***Holt-Winters eksponentinis glodinimas***

Šis eksponentinio glodinimo algoritmas buvo adaptuotas laiko eilučių su sezoniškumu ir trendu prognozavimui [58]. *Holt-Winters* eksponentinis glodinimas atliekamas naudojantis šiomis formulėmis [64]:

$$\hat{y}_{t+l} = (a_t + b_t)s_{t-n+i}, \quad l=1,2,\dots \quad (33)$$

$$a_t = a(y_t - s_{t-n}) + (1 - \alpha)(a_{t-1} + b_{t-1}), \quad (34)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}, \quad (35)$$

$$s_t = \gamma(y_t - a_t) + (1 - \gamma)s_{t-n}, \quad (36)$$

Čia  $\hat{y}$  – prognozė,  $a$  – lygio komponentas,  $b$  – poslinkio komponentas,  $t$  – laikotarpis,  $l$  – sezoniškumo periodas,  $\alpha, \beta, \gamma$  – parametrai,  $0 \leq \alpha, \beta, \gamma \leq 1$ .

Prognozė keletui periodų į priekį randama pagal šias formules:

$$\hat{y}_{t+k} = a_t + kb_t, \quad (37)$$

$$\hat{y}_{t+l+k} = (a_t + kb_t)s_{t-n+l}. \quad (38)$$

### Prognozavimo paklaidų vertinimas

Prognozuojamoji reikšmė visada yra su paklaida. Paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį, o jų skirstinį nusako dispersija (kuo didesnė rodiklio dispersija, tuo didesnė ir prognozės dispersija) [57].

Šiame darbe, prognozavimo modelių įvertinimui bus naudojami šie paklaidų tipai:

- 1) *RMSE* (angl. root mean square error) šaknis iš vidutinės kvadratinės paklaidos. *RMSE* metrika yra tiesiogiai susijusi su paklaidų variacija ir standartiniu nuokrypiu. Dėl šios priežasties neretai vertinant eilutės prognozės paklaidas yra vertinamas būtent paklaidų išsibarstymas apie jų vidurkį [58]. Ji apskaičiuojama:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^n (x_t - \hat{x}_t)^2} \quad (39)$$

- 2) *MAPE* (angl. mean absolute percentage error) tai vidutinė procentinė absoliutinė paklaida. Ši paklaida, nusako santykinę prognozavimo tikslumą, ja galima palyginti skirtingų rodiklių prognozes – *MAPE* (prognozavimas labai tikslus, kai  $MAPE < 10$  ir tikslus, kai  $10 < MAPE < 20$ ) [57]. Ji apskaičiuojama:

$$MAPE = \frac{1}{N} \sum_{t=1}^n \frac{|x_t - \hat{x}_t|}{x_t} \cdot 100 \% \quad (40)$$

- 3) *MASE* (angl. mean absolute scaled error) =  $[0..1..\infty]$  parodo, kiek kartų blogiau (ar geriau) prognozuojama už 1 žingsnio naivia prognozę.

### 3. Tiriamoji dalis

Šiame skyriuje bus pristatomos nagrinėjamos įmonės, jų finansiniai duomenys, aprašomi atlikti tyrimai – jų rezultatai bei išvados. Kiekviename tyrimo žingsnyje bus aptariama, kaip konkretūs didžiųjų duomenų analitikos metodai gali prisidėti prie finansinių ataskaitų audito procedūrų stiprinimo ir kokiam audito etape juos patartina naudoti. Tyrimui atlikti naudojamas programinis paketas *R Studio*.

#### 3.1. Tiriamieji objektai

Šiame darbe tyrimui atlikti bus naudojami dviejų Lietuvoje veikiančių gamybos įmonių *UAB Duona* ir *UAB Sausainis* duomenys. Dėl konfidencialumo reikalavimų įmonių pavadinimai yra pakeisti. Pagrindinė abiejų įmonių veikla pagal EVRK (Ekonominės veiklos rūšių klasifikatorių) patenka į *106000 Grūdų malimo produktų, krakmolo ir krakmolo produktų gamyba* klasę. Įmonėse gaminama duona, batonai, pyrago gaminiai, duonelės, sausainiai, sausi pusryčiai bei duonos ir bulvių traškučiai.

Tiriamus duomenis sudaro įmonių apskaitos įrašų žurnalai, sąskaitų apyvartos, didžiosios knygos bei finansinių ataskaitų rinkiniai. Tiriamasis laikotarpis apima 2012-2017 metus imtinai, jeigu konkrečiame tyrime nenurodyta kitaip. Finansinių duomenų dydžiai išreiškiami eurais.

Pagal įmonių finansinės atskaitomybės įstatymą uždarosioms akcinėms bendrovėms, auditas privalomas, jeigu jų ne mažiau kaip du rodikliai paskutinę finansinių metų dieną viršija šiuos dydžius [5]:

- pardavimo grynosios pajamos per ataskaitinius finansinius metus – 3 500 000 eurų;
- balanse nurodyto turto vertė – 1 800 000 eurų;
- vidutinis metų sąrašinis darbuotojų skaičius per ataskaitinius finansinius metus – 50;

Pateikiami nagrinėjamų įmonių paskutinės finansinių metų dienos (2017.12.31) finansiniai duomenys (žr. 7 lentelė):

lentelė 7. Nagrinėjamų įmonių finansiniai duomenys 2017.12.31 datai

	Pardavimo grynosios pajamos per ataskaitinius finansinius metus	Balanse nurodyto turto vertė	Vidutinis metų sąrašinis darbuotojų skaičius
<i>UAB Duona</i>	8 mln. Eur.	5 mln. Eur.	190
<i>UAB Sausainis</i>	13 mln. Eur.	7,7 mln. Eur.	160

Pagal duomenis pateiktus lentelėje nr. 6, matome, kad nagrinėjamoms įmonėms finansinių ataskaitų auditas yra privalomas [5].

### 3.2. Laiko eilučių analizė

Pagal Appelbaum, Kogano ir Vasarhelyi atliktą tyrimą laiko eilučių analizė naudinga keturiuose pirmuose audito etapuose: susipažįstant su klientu, planuojant auditą, pagrindinių procedūrų atlikimo procese bei vertinant atliktas procedūras [22]. Tyrime lygiagrečiai bus atliekama dviejų įmonių analizė, kadangi pagal TAS lyginamieji duomenys apie konkurentus yra vienas iš audito įrodymų šaltinių. Be to, pagal audito standartą: analitinės procedūros apima svarstymus dėl įmonės finansinės informacijos palyginimo su kitos panašaus dydžio tos pačios ūkio šakos įmonės informacija [18]. Pagal finansinių rodiklių dydžius abi nagrinėjamos įmonės priskiriamos vidutinio dydžio įmonėms.

Kiekvieno ūkio subjekto sėkmingo egzistavimo prielaida yra pardavimo pajamų gavimas. Auditoriui norint susipažinti su gamybinės įmonės veikla, dydžiu, nustatyti audite naudojamą reikšmingumo lygį bei efektyviai įvertinti įmonės veiklos rezultatus ir jų tęstinumą vieni iš svarbiausių įmonės finansinių rodiklių yra pardavimo pajamos bei bendrasis pelningumas (angl. gross profit ratio), kuris įvertina kiek buvo uždirbta ir kiek tam buvo išleista. Taigi toliau bus nagrinėjamos būtent šių dydžių laiko eilutės.

Bendrasis pelningumas apskaičiuojamas pagal tokią formulę:

$$\text{Bendrasis pelningumas} = \frac{\text{Pardavimo pajamos} - \text{pardavimo savikaina}}{\text{Pardavimo pajamos}} \quad (41)$$

Duomenų paruošimo analizei žingsnyje nustatyta, kad duomenyse nėra trūkstamų įrašų, todėl jų interpoluoti nereikėjo.

Pateikiamos pagrindinės abiejų įmonių pardavimo pajamų ir savikainos 2012-2017 metų laiko eilučių charakteristikos metiniams (žr. 8 lentelė) ir mėnesiniams (žr. 9 lentelė) duomenims:

lentelė 8. Pardavimo pajamų ir savikainos metinių duomenų analizė

Įmonė	Dydis	Minimali reikšmė	Pirmasis kvartilis	Mediana	Vidurkis	Trečiasis kvartilis	Maksimali reikšmė
UAB Duona	Pajamos	6203923	7431299	7630233	7478128	7843973	8124172
	Savikaina	3524997	4116293	4572133	4394020	4663854	5036083
UAB Sausainis	Pajamos	10718058	12506345	12878040	12685993	13372407	13722175
	Savikaina	8021890	8640690	9145987	9062639	9277062	10268960

Pagal aukščiau pateiktą lentelę matome, kad *UAB Duona* pardavimo pajamų augimas lyginant mažiausią ir didžiausią įgytas reikšmes yra 31%, atitinkamas savikainos reikšmes skiria 43 %. Iš to galima spręsti, kad bent dalyje nagrinėjamo laikotarpio savikaina augo greičiau nei pajamos, o tai reiškia, kad bendrasis pelningumas mažėjo. Nagrinėjant *UAB Sausainis* laiko eilutes galime pastebėti šiek tiek kitokias tendencijas. Pardavimo pajamų didžiausia įgyta reikšmė yra 28 % didesnė už

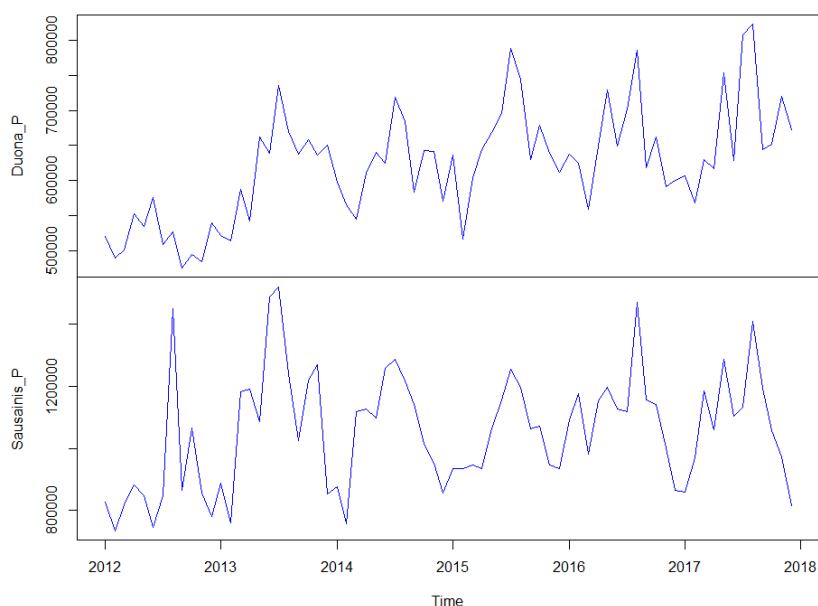
minimalią gautų pajamų ribą, toks taps procentinis pokytis skiria ir mažiausią ir didžiausią savikainos reikšmes. Taigi galima daryti prielaidą, kad *UAB Sausainis* metinių duomenų bendrasis pelningumas nagrinėjamu laikotarpiu buvo stabilus.

lentelė 9. Pardavimo pajamų ir savikainos mėnesinių duomenų analizė

Įmonė	Dydis	Minimali reikšmė	Pirmasis kvartilis	Mediana	Vidurkis	Trečiasis kvartilis	Maksimali reikšmė
UAB Duona	Pajamos	475798	567860	630081	623177	662108	822809
	Savikaina	260398	332072	372062	366168	399317	503449
UAB Sausainis	Pajamos	734620	885946	1065032	1057166	1183343	1520134
	Savikaina	560290	663718	762988	770164	840112	1130046

Mėnesiniuose duomenyse išryškėja kiek kitokie pokyčiai tarp eilučių minimumų ir maksimumų. Pagal mėnesinių duomenų statistiką galime įžvelgti, kad abiejų tiriamų įmonių pajamų bei savikainos kitimo reikšmės yra gana plačiame intervale. Pavyzdžiui, *UAB Sausainis* pardavimo pajamų eilutės minimalią ir maksimalią reikšmes skiria beveik 107 %, atitinkamai savikainos pokytis 102 %. Tokie mėnesinių duomenų šuoliai signalizuoja apie galimą sezonškumo įtaką įmonių veiklai.

Pradiniame laiko eilučių analizės etape didelė dalis informacijos gaunama iš grafiko. Iš jo nesunkiai matosi, ar yra tam tikra ilgalaikė stebimo rodiklio kitimo tendencija, ar yra sezoniniai svyravimai, iš dalies galima pasakyti ar stebimas procesas yra stacionarus. Vizualizacijas, kaip vieną iš galimų audito technikų patariama naudoti visuose audito etapuose [22]. Vizualiai pateikti rezultatai netgi įvardinami, kaip būdas padaryti išpūdį klientams ir gali būti puiki priemonė siekiant dialogą su klientu padaryti informatyvesniu [66]. Pateikiamas *UAB Duona* ir *UAB Sausainis* mėnesinių pardavimo pajamų grafikas (žr. 8 pav.).

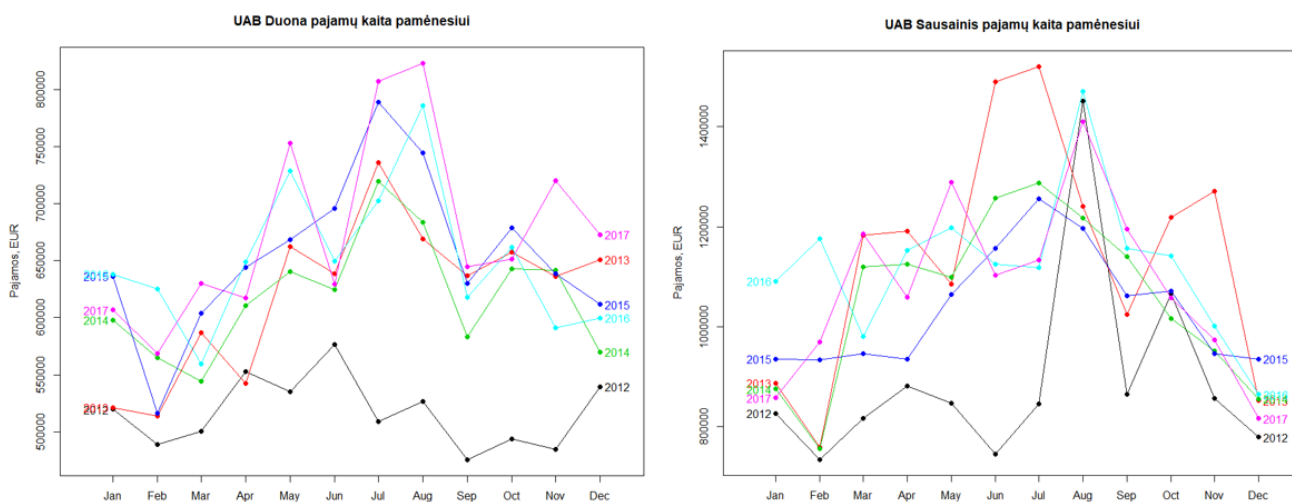


pav. 8 UAB Duona ir UAB Sausainis pajamų kaita pamėnesiui



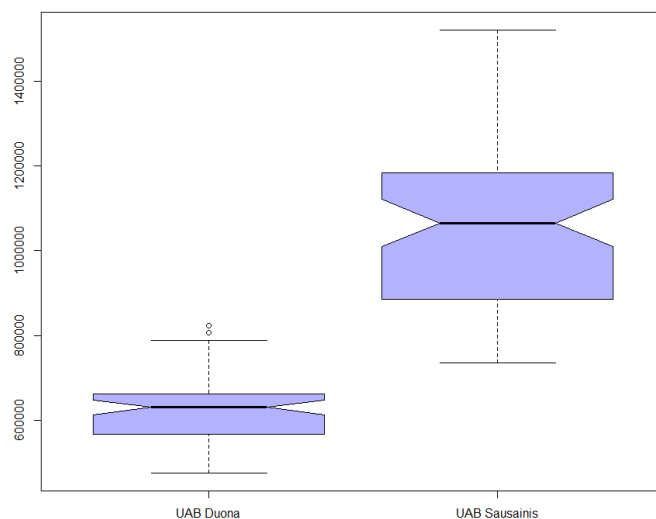
Pastebime, kad abiejų įmonių didžiausių pardavimo pajamų gavimo pikas kasmet sutampa, išimti galima laikyti tik *UAB Duona* 2012 m. pardavimų rezultatus, kurie buvo prasčiausi visu tiriamu laikotarpiu. Iš to galima teigti, kad pastaroji įmonė sunkiau atlaikė tuo metu buvusią pasaulinę ekonominę krizę. Pagal grafiką (žr. 8 pav.) galima numanyti, kad įmonių veikloje vyrauja sezoniskumas. Panašios tendencijos išvelgiamos ir įmonių savikainos laiko eilučių grafikuose (žr. 1 priedas, 21 pav.).

Pagal žemiau pateiktą grafiką (žr. 9 pav.) matome, kad įprastai metų pabaigoje ir metų pradžioje pardavimų apimtys gerokai sumažėja. Pastebima tendencija, kad dažnu atveju, pardavimų pajamos pradeda augti gegužės mėnesį, o geriausius rezultatus abi įmonės fiksuoja vasaros laikotarpiu. Matome, kad neįprastai didelis pajamų šuolis užfiksuotas *UAB Sausainis* duomenyse 2012 m. rugpjūčio mėnesį. Nors šiam mėnesiui pajamų augimas yra būdingas, 2012 metų kontekste jis gerokai išsiskiria iš kitų mėnesių. Šis šuolis reikalauja detalesnių audito procedūrų, tokių kaip pirminių dokumentų tikrinimas.



pav. 9 Mėnesinių pajamų palyginimas 2012 m.- 2017 m. laikotarpiu

Procedūros *Box plot* pagalba nustatyta, kad *UAB Duona* pajamų duomenyse yra 2 duomenų taškai, kurie viršija stačiakampės diagramos “ūso” ilgį (1,5 tarpkvartilinio pločio), todėl gali būti laikomi išskirtimais (žr. 10 pav.). Nustatyta, kad šie du taškai priklauso 2017 m. liepos ir rugpjūčio mėnesiams, o šiuo laikotarpiu kasmet fiksuojami didžiausi pardavimai, todėl šių išskirčių nelaikome neįprastomis ar reikalaujančiomis papildomų procedūrų.

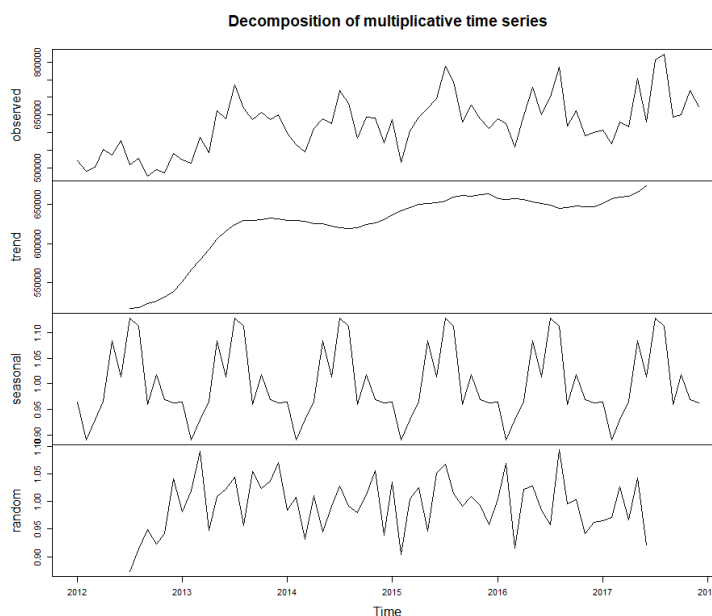


pav. 10 *Box plot* procedūros grafikas

**Laiko eilutės dekompozicija.** Laiko eilutes sudaro trys komponentės: trendas, sezoniškumas ir ciklas. Siekiame nustatyti, ar *UAB Duona* pajamų laiko eilutėje egzistuoja sezoniškumas ir trendas, cikliškumo neieškome, kadangi tai būtų neprasminga žinant, kad tiriami finansiniai duomenys.

Tikriname ar laiko eilutė yra stacionari. Atlikus *Augmented Dickey-Fuller* testą gauta, kad  $p=0.3771 > 0.05$ , taigi  $H_0$  hipotezės atmesti negalime (žr. 2 priedas).

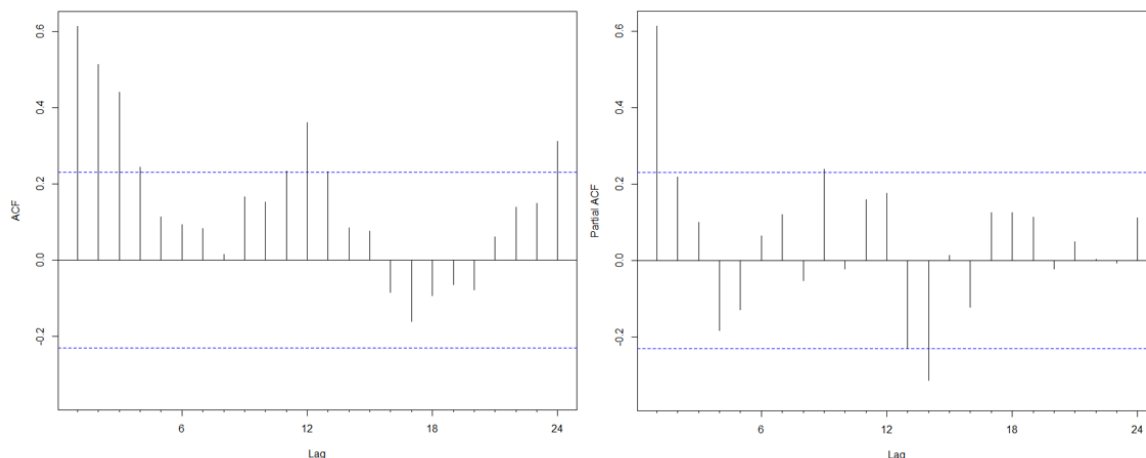
Dekompozicijai naudojome *multiplikatyvinį* modelį, kadangi laiko eilutė yra nestacionari (žr. 11 pav.).



pav. 11 *UAB Duona* pardavimo pajamų laiko eilutės dekompozicija

Nustatyta, kad *UAB Duona* laiko eilutėje nėra tiesinio trendo, tačiau egzistuoja sezoniškumas, kurio periodas metai.

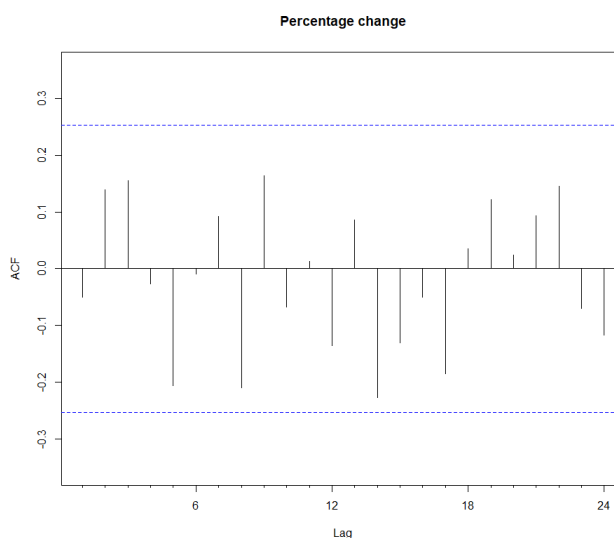
Laiko eilutės autokoreliacinės funkcijos (angl. autocorrelation function) vertinimui naudosime procedūras *acf* ir *pacf*. Pagal autokoreliacijos funkcijos grafiką sprendžiama apie laiko eilutės narių tarpusavio priklausomybę, o tuo pačiu ir apie  $Y_t$  modelio pavidalą. Autokoreliacijos funkcijos reikšmės parodo tiesinės priklausomybės stiprumą tarp laiko eilutės narių  $Y_t$  ir  $Y_{t-k}$ . Pagal tai sprendžiama, kokią įtaką dabartinei stebimo rodiklio reikšmei turi praeityje stebėtos reikšmės. Čia *lag* yra laiko pokytis.



pav. 12 UAB Duona *acf* ir *pacf* procedūrų grafikai

ACF grafike matome išryškėjusią sezoninę 12 ir 24 mėnesio komponentę.

Tiriame ar likusi nepaaiškinama laiko eilutės dalis (toliau, *random* komponentė) yra baltasis triukšmas:



pav. 13 UAB Duona *random* komponentės *acf* grafikas

Tiriama seka yra baltasis triukšmas, kadangi visos autokoreliacinės funkcijos reikšmės, yra melsvos juostos viduje, be to *random* komponentės grafike nėra trendo ar sezoniškumo pėdsakų.

Analogiška UAB Sausainis pajamų laiko eilutės dekompozicija pateikta 2 priede.

Apibendrinant, nustatyta, kad abiejuose tiriamose įmonėse vyrauja sezoniškumas, kurio periodas yra metai. Didžiausios pardavimo pajamos kasmet gaunamos vasaros laikotarpiu.

### 3.3. Veiksnių darančių poveikį bendrajam pelningumui nustatymas

Šiame skyrelyje siekiama nustatyti, kurie išorės veiksniai daro didžiausią poveikį įmonės bendrajam pelningumui (toliau BP). BP pasirinktas todėl, kad jis apima ir pajamas ir savikainą, t.y. parodo kiek buvo uždirbta ir kiek tam buvo išleista. Tirti išorės veiksnių poveikį tik savikainai būtų nekorektiška, kadangi savikainos kitimas nebūtinai susijęs tik su išorės veiksniais, dažniausiai savikaina didžiąja dali priklauso nuo pardavimo pajamų dydžio. Šiam tyrimui bus naudojamos koreliacinė bei priežastingumo analizės. Bus tiriamas BP ryšys su kviečių, cukraus, sviesto kainomis bei vidutiniu šalies darbo užmokesčiu. Šie veiksniai pasirinkti dėl to, kad jie sudaro nemažą dalį gaminio savikainos. Išorės rodiklių duomenų šaltiniai nurodyti 3 priede. Toliau bus nagrinėjamas *UAB Duona* atvejis, analogiškas tyrimas su *UAB Sausainis* duomenimis pateikiamas 4 priede.

Koreliacinės analizės pagalba nustatysime BP atsitiktinio koreliacinio ryšio su išorės rodikliais stiprumą. Šiame darbe ryšio stiprumui įvertinti naudojamas *Spearmano* koreliacijos koeficientas, kuris vertina kintamųjų panašumą monotoniškumo prasme ir gali būti taikomas, kai kintamieji nėra pasiskirstę pagal normalųjį dėsnį. Analizei naudojami mėnesiniai rodikliai. Kadangi vidutinio darbo užmokesčio mėnesinių rodiklių nepavyko rasti, remiantis ketvirčių duomenimis trūkstantis reikšmės interpoliuotos slenkančio vidurkio metodu. Pateikiama koreliacinio ryšio stiprumo įvertinimo diagrama (žr. 14 pav.).



pav. 14 Koreliacinio ryšio stiprumo įvertinimo diagrama, *UAB Duona* atvejis

Pagal pateiktą diagramą matome, kad didžiausias neigiamas koreliacinis ryšys (-0,44) BP sieja su vidutiniu darbo užmokesčiu (DU), tai reiškia, kad didėjant darbo užmokesčiui bendrasis pelningumas mažėja. Didžiausias teigiamas koreliacinis ryšys (0,42) BP sieja su kviečiais, tai gana keista, kadangi augant kviečių kainai bendrasis pelningumas turėtų mažėti, teigiamam ryšiui įtakos galėjo turėti atsitiktinė panaši laiko eilučių kitimo tendencija. Kitų tiriamų produktų ryšiai su BP nėra reikšmingi.

Priežastiniam ryšiui nustatyti naudosime *Grangerio* testą. Jis ne tik nustato ryšio egzistavimą ir jo kryptis, bet ir priežastingumą, t. y. atsako į klausimą, kuris iš nagrinėjamų veiksnių yra priežastis, o kuris pasekmė. Atlikto tyrimo metu gautų rezultatų patikimumą vertinsime taikant šiuos kriterijus: t statistiką ir *F (Fisher)* kriterijų, tačiau pirmumas teikiamas *F* kriterijui, kadangi tyrimuose juo remiamasi dažniausiai. Pasirinktas reikšmingumo lygmuo 0,05.

Pateikiama koreliacinės analizės ir *Grangerio* priežastingumo testo apibendrinamoji lentelė tarp BP ir visų nepriklausomų kintamųjų:

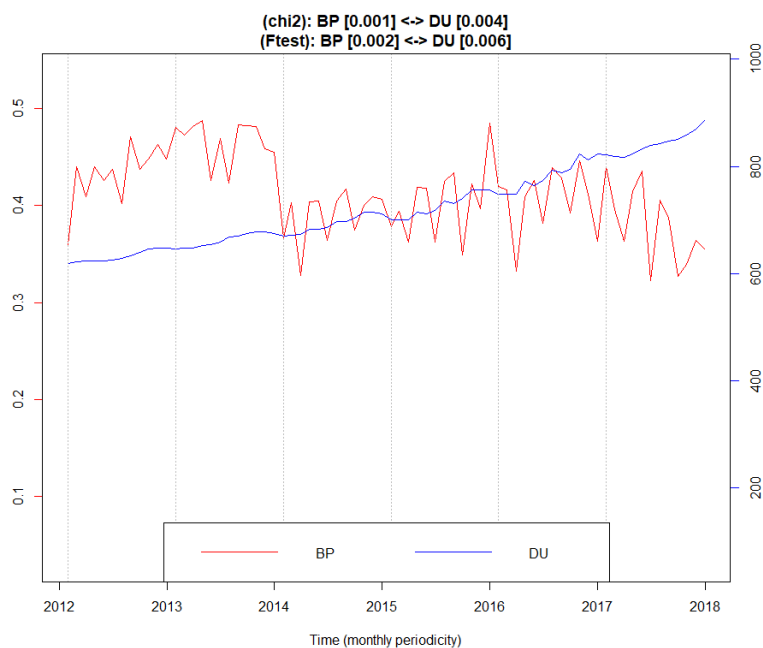
lentelė 10 koreliacinės analizės ir *Grangerio* testo rezultatai, *UAB Duona* atvejis

Testas (su BP)	Sviestas	Cukrus	Kviečiai	DU
Koreliacija	-0,09	0,19	0,42	-0,44
Grangerio testas	—	—	—	<—>

Šioje lentelėje koreliacija matuojama *Spearmano* koreliacijos koeficiento dydžiu. *Grangerio* testo rezultatus reikėtų interpretuoti pagal rodyklių kryptis, t.y. rodyklės kryptis rodo į kurią pusę nukreiptas priežastinis ryšys. Symbolis “-” rodo, kad priežastinio ryšio tarp kintamųjų nėra. Remiantis atlikta priežastingumo analize galime atmesti 3 kintamuosius, kaip nedarančius reikšmingo poveikio BP t.y. sviestą, cukrų ir kviečius. Nustatyta, kad abipusis priežastinis ryšys sieja BP su vidutiniu darbo užmokesčiu, taigi šiuo atveju BP yra ir priežastis ir pasekmė (žr. 15 pav.).

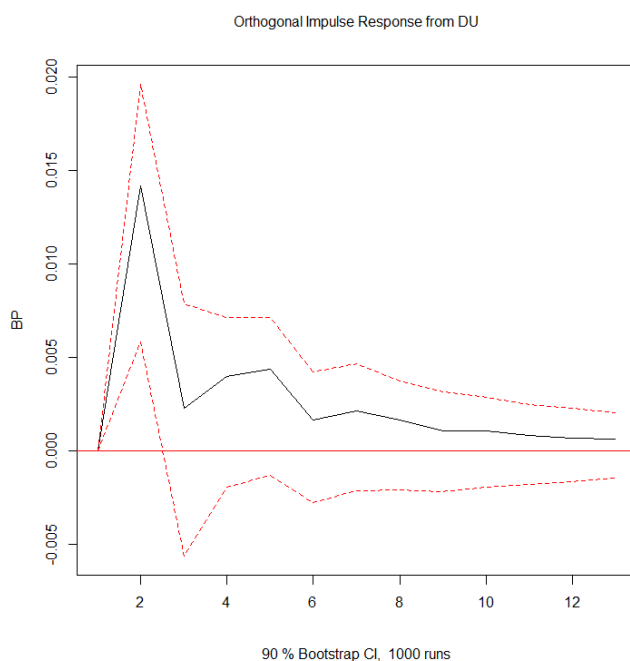
Pagal AR modelio išraišką matome vėlinimų ilgį tarp tiriamų kintamųjų:

Statistika	Grangerio testo rezultatai	AR modelis
<i>chi</i> kvadratas	BP [0.648] — sviestas [0.700]	AR(6+1)
<i>F</i> kriterijus	BP [0.649] — sviestas [0.699]	
<i>chi</i> kvadratas	BP [0.461] — cukrus [0.673]	AR(3+1)
<i>F</i> kriterijus	BP [0.467] — cukrus [0.675]	
<i>chi</i> kvadratas	BP [0.587] - kviečiai [0.065]	AR(3+1)
<i>F</i> kriterijus	BP [0.590] - kviečiai [0.075]	
<i>chi</i> kvadratas	BP [0.001] <-> DU [0.004]	AR(2+1)
<i>F</i> kriterijus	BP [0.002] <-> DU [0.006]	



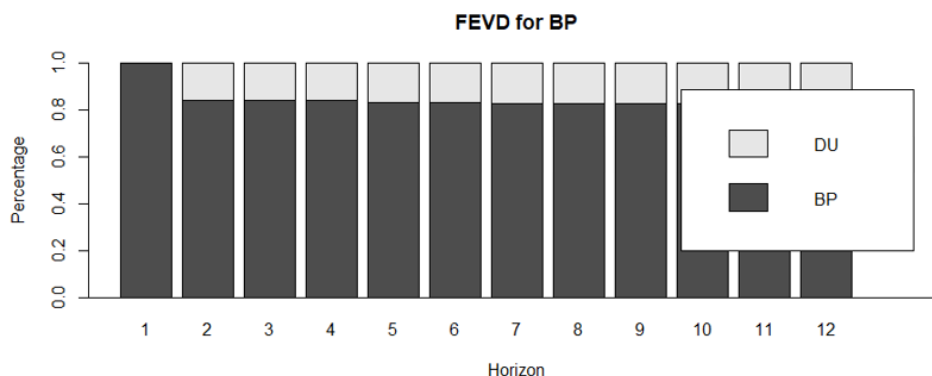
pav. 15 *Grangerio* testo rezultatai tarp bendrojo pelningumo ir vidutinio darbo užmokesčio

Atlikus *Grangerio* priežastingumo analizę patogu nubrėžti impulsų atsako ir FEVD (angl. forecast error variance decompositions) grafikus. Jie yra labai naudingi siekiant įvertinti, kaip atskirų kintamųjų šuoliai (stiprūs pokyčiai) atsispindi nagrinėjamo rodiklio pokyčiuose.



pav. 16 impulsų atsako grafikas

Impulsų atsako funkcija kintamajam DU parodo, kaip pasikeistų BP vidutiniam darbo užmokesčiui šalyje padidėjus vienu standartiniu nuokrypiu (žr. 16 pav.). Kaip matome, pagal aukščiau pateiktą grafiką, BP kitimas būtų intervale nuo -0,005 iki 0,02. Po suteikto impulso vidutiniam darbo užmokesčiui BP kitimas stabilizuojasi maždaug po 6 mėnesių.



pav. 17 FEVD grafikas tarp BP ir DU, UAB Duona atvejis

Aukščiau pateiktame FEVD grafike matome, kokia BP dalis yra veikiamą vidutinio darbo užmokesčio, matome, kad bėgant laikui veikiamą dalis didėja ir sudaro apie 20 % (žr. 17 pav.).

Apibendrinant, nustatyta, kad iš tiriamų išorės rodiklių tik vidutinis darbo užmokestis turi priežastinį ryšį ir tik su UAB Duona bendruoju pelningumu. Tai galima paaiškinti tuo, kad nors UAB Duona pardavimo pajamos bei savikaina yra gerokai mažesnės už UAB Sausainis, UAB Duona dirba daugiau darbuotojų, taigi DU sudaro didesnę dalį savikainos negu UAB Sausainis atveju. Tai pat nustatyta, kad UAB Sausainis BP sieja priežastinis ryšys su sviesto, cukraus bei kviečių kainomis, tačiau visais atvejais BP yra priežastis, o ne pasekmė. Šios krypties priežastinis ryšys galėjo atsirasti dėl to, kad įmonė su tiekėjais sudaro ilgalaikes sutartis ir dažniausiai žino už kokią kainą ji ateityje pirs žaliavas, pagal tai atitinkamai iš anksto koreguoja gaminių sudėtį bei savikainą.

### 3.4. Finansinių rodiklių prognozavimas

Remiantis 570-uoju TAS: auditorius privalo surinkti pakankamą tinkamą audito įrodymų dėl vadovybės taikomos veiklos tęstinumo prielaidos tinkamumo, rengiant ir pateikiant finansines ataskaitas, ir padaryti išvadas dėl to, ar egzistuoja reikšmingas neapibrėžtumas dėl įmonės gebėjimo tęsti veiklą [18]. Auditorius turi įvertinti vadovybės atliktą įmonės gebėjimo tęsti veiklą vertinimą, kurį gali sudaryti planuojami biudžetai, pardavimo pajamų ar pinigų srautų prognozės ir pan. Norint įvertinti ar įmonės planuojamos pardavimų pajamos bei kiti svarbūs veiklos rodikliai yra patikimi, auditorius gali atlikti analizuojamų dydžių prognozę ir taip palyginti savo ir įmonės atstovų gautus rezultatus. Rodiklių prognozę tikslinga atlikti metams į priekį, kadangi kaip reikalauja TAS auditorius turi prašyti įmonės vadovybės pateikti gebėjimo tęsti veiklą vertinimą bent dvylikos mėnesių laikotarpiui nuo finansinių ataskaitų datos [18].

Kadangi siekiama prognozuoti 12 mėnesių į priekį modelių įvertinimui atidedama 12 paskutinių mėnesių, t.y. pirmiausia, remiantis 2012-2016 metų mėnesiniais duomenimis atliksime prognozę 2017 metams, tuomet įvertinsime gautus rezultatus naudojami RMSE, MAPE ir MASE paklaidų tipus.

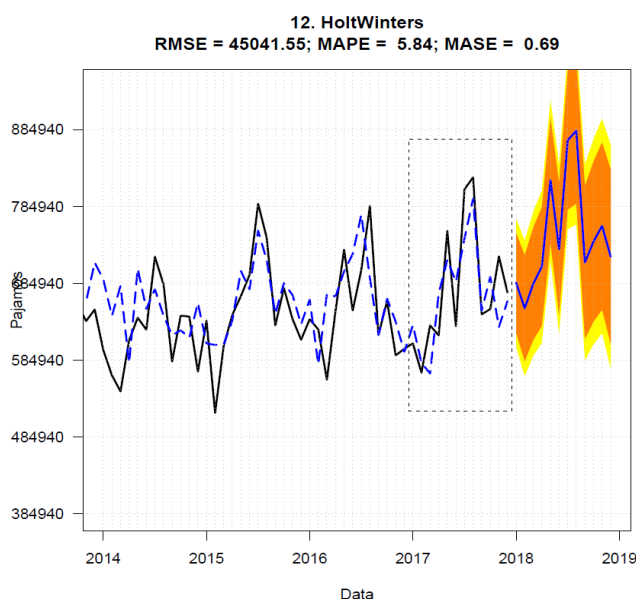
Modeliai pagal gerumą rikiuojami remiantis *RMSE* įverčiu. Nustačius mažiausią paklaidą padariusį prognozavimo modelį bus atliekama prognozė 2018 metams.

Siekiant prognozuoti pardavimo pajamų ir bendrojo pelningumo įverčius 12 mėnesių į ateitį buvo sudaryti įvairūs vienmačiai prognozės modeliai. Žemiau pateikiama lentelė su *UAB Duona* pardavimo pajamoms prognozuoti išbandytais prognozavimo modeliais bei klaidų tipų *RMSE*, *MAPE*, *MASE* įverčiais (žr. 11 lentelė).

lentelė 11 *UAB Duona* pajamų prognozės klaidų įverčiai

Nr.	Metodo pavadinimas	RMSE	MAPE	MASE
1	HoltWinters	45041.55	5.84	0.69
2	Linear regression model	53100.03	7.13	0.81
3	Basic structural model	55538.42	6.85	0.83
4	ARIMA(0,1,2)(0,1,1)[12]	58243.58	6.84	0.84
5	Seasonal naive method	62127.17	7.49	0.9
6	ETS(M,Ad,M)	65332.56	6.89	0.88
7	BATS(0.005, {0,0}, -, {12})	66825.73	7.17	0.91
8	Linear regression model	79377.69	10.99	1.28
9	Theta	95861.64	8.97	1.17
10	Mean	100968.63	9.85	1.27
11	Random walk with drift	102070.6	10.02	1.29
12	ARIMA(0,1,1)	103496.17	10.21	1.32
13	Random walk	109491.75	11.21	1.44

Kaip matyti iš lentelės geriausius paklaidų įverčius gavome taikydami *Holt-Winters* modelį. Šis eksponentinio glodinimo algoritmas buvo adaptuotas laiko eilučių su sezoniškumu ir trendu prognozavimui. Jis yra ypač tinkamas trumpalaikėms pardavimų ar paklausos laiko eilučių prognozėms [67].



pav. 18 *UAB Duona* pardavimo pajamų prognozė



Pateiktame grafike (žr. 18 pav.) galima matyti kaip modelis atitinka turimus duomenis ir preliminarias prognozes 12 mėnesių į priekį. Geltona spalva grafike yra žymimas pasikliautinis intervalas, kai reikšmingumo lygmuo yra  $\alpha=0.1$  ir oranžine spalva žymimi pasikliautinieji intervalai, kai reikšmingumo lygmuo  $\alpha=0.2$ .

Remiantis modelio gautomis prognozėmis galime teigti, kad per 2018 metus bus gauta ~10,32 % daugiau pardavimų pajamų negu per 2017 metus. Įmonės vadovas metiniame pranešime nurodė, kad 2018 metais planuojamas 5 % pardavimo pajamų augimas, taigi pagal atliktą tyrimą galime vertinti jo prognozes kaip patikimas. Gautos prognozės kiekvienam mėnesiui bei pasikliautiniųjų intervalų įverčiai pateikti 5 priede.

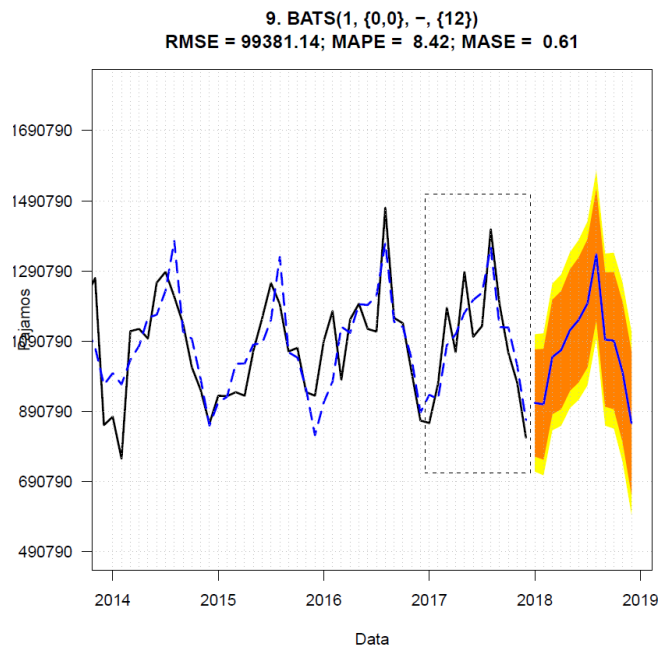
Toliau nagrinėjamas *UAB Sausainis* atvejis.

Žemiau pateikiama lentelė su *UAB Sausainis* pardavimo pajamoms prognozuoti išbandytais prognozavimo modeliais bei klaidų tipų *RMSE*, *MAPE*, *MASE* įverčiais (žr. 12 lentelė). Kaip matyti iš lentelės geriausius paklaidų įverčius pagal *RMSE* gavome taikydami *BATS* modelį. Pagal *MAPE* ir *MASE* įverčius geriausiai pasirodė *Holt-Winters* metodas.

lentelė 12 *UAB Sausainis* pajamų prognozės klaidų įverčiai

Nr.	Metodo pavadinimas	RMSE	MAPE	MASE
1	BATS(1, {0,0}, -, {12})	99381.14	8.42	0.61
2	ETS(M,N,A)	101481.29	8.09	0.59
3	HoltWinters	113119.66	7.98	0.56
4	Seasonal naive method	119844.17	9.11	0.62
5	Linear regression model	123323.75	10.58	0.72
6	ARIMA(1,0,0)(1,0,0)[12] with non-zero mean	130127.51	9.92	0.72
7	ARIMA(1,0,0) with non-zero mean	164564.18	11.57	0.85
8	Mean	167995.07	12.37	0.89
9	Linear regression model	180744.31	14.78	0.97
10	Basic structural model	196658.95	13.67	1.04
11	Theta	201875.66	14.1	1.08
12	Random walk with drift	273320.78	19.52	1.52
13	Random walk	276628.16	19.74	1.54

Pateiktame grafike (žr. 19 pav.) galima matyti kaip *BATS* modelis atitinka turimus duomenis ir preliminarias prognozes 12 mėnesių į priekį. Geltona spalva grafike yra žymimas pasikliautinis intervalas, kai reikšmingumo lygmuo yra  $\alpha=0.1$  ir oranžine spalva žymimi pasikliautinieji intervalai, kai reikšmingumo lygmuo  $\alpha=0.2$ .



pav. 19 UAB Sausainis pardavimo pajamų prognozė

Remiantis modelio gautomis prognozėmis galime teigti, kad per 2018 metus bus gauta ~2 % mažiau pardavimų pajamų negu per 2017 metus. Įmonės vadovas metiniame pranešime nurodė, kad 2018 metais planuojamas 1 % pardavimo pajamų augimas, taigi pagal atliktą tyrimą galime vertinti jo prognozes kaip gana patikimas, kadangi *BATS* metodu gautos prognozės pasikliautinumo intervalai numato ir optimistiškesnes augimo galimybes. Gautos prognozės kiekvienam mėnesiui bei pasikliautinių intervalų įverčiai pateikti 5 priede.

Apibendrinant, pagal gautus prognozių rezultatus, įmonių vadovybė pateikė patikimus pardavimo apimčių planus 2018 metams. Prognozuojama, kad abiejų įmonių BP turėtų išlikti panašiam lygyje. Abiejų įmonių bendrojo pelningumo prognozavimo rezultatai pateikti 6 priede. Rizikos susijusios su veiklos tęstinumo prielaida nenustatyta.

### 3.5. Bendrojo žurnalo įrašų teksto tyryba

Bendrajį žurnalą sudaro visos įmonės ūkinės operacijos t.y. visų per ataskaitinį laikotarpį įvestų pirminių dokumentų dvejetainiai įrašai. Iš bendrojo žurnalo duomenų formuojama didžioji knyga, balansas ir daug kitų ataskaitų. Bendrajame žurnale pateikiama išsami informacija apie įrašą: operacijos ir jos įvedimo datos, sąskaitų korespondencija, suma, valiuta, iš kokio modulio operacija perkelta, kas operaciją įvedė ir kita svarbi informacija. Šiandien dažnai bendrajame žurnale pateikiamas ir įrašo apibūdinimas, kurį sudaro tekstinė informacija. Auditoriui perskaityti visų įrašų apibūdinimus atimtų labai daug lauko, o didelių įmonių atveju tai būtų praktiškai neįmanoma. Puikus būdas auditoriui susidoroti su tekstine informacija yra pasitelkti teksto tyrybos metodus.

Toliau bus atliekamas *UAB Duona* bendrojo žurnalo išlaidų sąskaitų tekstinės informacijos testavimas. Į tyrimą nebus įtraukiamos didžiosios knygos (toliau DK) sąskaitos susijusios su darbo užmokesčiu, kadangi šių sąskaitų aprašymus sudaro konfidenciali įmonės darbuotojų informacija. Tyrime taip pat nebus nagrinėjamos savikainos DK sąskaitos, kadangi jų aprašymai nesuteikia papildomos informacijos apie konkretų įrašą t.y. visų sąskaitų apibūdinimas yra *tiesioginė savikaina*.

Pateikiamas pradinių duomenų fragmentas:

DK Sąskaita	Aprašymas
61113 Techninės apžiūros sąnaudos	Už techninę apžiūra MAN12.224
61113 Techninės apžiūros sąnaudos	Už techninę apžiūra MAN14.225
61105 Transporto remontas ir eksploatacija	Padangų montavimas
61112 Vinjetės	Kelių vinjetė
61112 Vinjetės	Kelių mokestis
61116 Samdomiems vežėjams sąnaudos	Pavyzdžių siuntimas Kinija, Lenkija
61116 Samdomiems vežėjams sąnaudos	Pavyzdžių siuntimas i Kinija

Tiriamąją imtį sudaro 10304 įrašai. Kiekvienas bendrojo žurnalo įrašas yra atskiras vektorius. Prieš pradėdant tyrimą iš teksto pašaliname lietuviškas raides. Tuomet iš pradinių dokumentų pašalinami tarpai, visi skyrybos ženklai, žodžiai ar kiti simboliai, kurie yra nereikšmingi, nieko nepasako apie dokumento turinį. *UAB Duona* atveju iš teksto pašalinami šie dažnai pasikartojantys bereikšmiai žodžiai ar žodžių trumpiniai: „kt“, „am“, „pagal“, „ad“, „sf“, „np“, „uz“, „ir“, „men“, „per“, „islaidos“, „sanaudos“, „paslaugos“.

Sujungiamo žodžius, kurie tekste turėtų būti traktuojami, kaip vienas žodis, t.y. sudarome žodžių n-gramas. *UAB Duona* atveju naudosime šias n-gramas: „nekilnojamo turto mokestis“, „Neig val pasikeitimo itaka“, „palukanos uz kredita“, „neleidziami atskaitymai“, „duonos broko nurasymas“, „el energija“, „mob rysio“, „logistikos paslaugos“, „garantinio fondo mokestis“.

Atlikus visus aptartus punktus atskiros leksemos transformuojamos į vektorinį pavidalą, kuris bus suprantamas teksto tyrybos algoritmams. Požymiams atvaizduoti vektoriuje bus naudojamas požymių dažnumo atvaizdavimas. Žinant požymių rinkinį, formuojame dokumentų matricą, kurioje eilutės atitinka dokumentus, o stulpeliai požymius.

Siekiant nustatyti neįprastus bei labai retus įrašus ieškome žodžių, kurie turi mažiausią dažnį. Žemiau pateikiami 20 rečiausiai bendrojo sąnaudų žurnalo įrašuose pasitaikančių žodžių bei jų dažniai:

Žodis	Dažnis
baltimore	1
bwi	1
usa	1
uzregistruota	1
keitimas	1
registracijos	1
imbiss	1
arabija	1
saudu	1
geguzes	1

Žodis	Dažnis
beiruta	1
dokumentu	1
santjiego	1
boxer	2
partner	2
bmw	2
plovimui	2
aparatu	2
guoliu	2
parsiuntimas	2

Nustatyta, kad žodžiai reiškiantys šalių ar miestų pavadinimus susiję su produkcijos pavydžių siuntimu į užsienio rinkas, dalyvavimą parodose. Informacija apie sieki didinti eksportą tai pat atskleista įmonės metiniame pranešime. Dalis retų žodžių susiję su automobilių remontu, priežiūra bei technine apžiūra. Kadangi tarp retų žodžių yra automobilių markių, modelių auditoriui reikėtų patikrinti ar buvo įsigyta naujų automobilių, sudaryta nuomos sutarčių. Apibendrinant, neišprastų įrašų nenustatyta.

Ieškome dažniausiai pasikartojančių žodžių, kurie padės atskleisti pagrindinius sąnaudų šaltinius. Pateikiami 20 daugiausiai kartų tiriamoje imtyje pasikartojusių žodžių bei jų dažniai:

Žodis	Dažnis
banko	1656
automobilio	915
remontas	897
komandiruotes	724
elektros	664
reklama	636
dujas	503
nusidevejimas	430
siuntimas	368
draudimas	317

Žodis	Dažnis
logistikos	293
traskuciu	268
nuotekas	244
vandeni	244
apziura	230
nurasymas	224
mokymams	212
sveikatai	212
bauda	205
prekes	194

Nustatyta, kad dažniausiai pasikartojantis žodis yra *banko*. Iš to galima spręsti, kad įmonė galimai turi finansinių įsipareigojimų bankui. Šiuo atveju auditui reikia patikrinti kreditavimo sutartis, jų paskirtį, jose numatytas sąlygas, terminus, atsiskaitymų su banku savalaikiškumą, įvertinti įmonės galimybes grąžinti kreditą laiku. Taip pat reikėtų paprašyti įmonės atstovų užsakyti banko paklausimą, kuris būtų atsiųstas tiesiai auditui, kuriame būtų atskleistos kreditavimo sąlygos, atlikti ir laukiantys mokėjimai, nurodytos paskolos užtikrinimo priemonės. Kiti dažnai pasikartojantys žodžiai tokie kaip: *automobilio*, *remontas*, *draudimas*, *apziura* susiję su įmonės automobilių parku. Galime daryti prielaidą, kad įmonė savo veikloje naudoja nuosavą transportą. Taigi auditui reikia nustatyti automobilių finansavimo šaltinius, patikrinti naujus įsigijimus, įvedimo į eksploataciją



Gautus tekstų analizės rezultatus atvaizduojame dažnių diagrama (žr. 6 priedas), tačiau yra ir modernesnių diagramų, tokių kaip žodžių debesis (angl. *word cloud*), kuris leidžia moderniai ir suprantamai vizualizuoti žodžių dažnius - kuo didesnis žodis, tuo jis dažnesnis dokumente (žr. 20 pav.).

Atliekame analogišką *UAB Sausainis* bendrojo žurnalo išlaidų sąskaitų tekstinės informacijos testavimą.

Pateikiamas pradinių duomenų fragmentas:

DK Sąskaita	Aprašymas
60205 Elektros sąnaudos	Elektra
61011 Muitinės dok. formavimo sąnaudos	Muit. Dok.
61013 Dizaino darbai, klišės, maketai	Pakuočių dizainas
61014 Degalų sąnaudos	Kuras
61015 Nuoma, sandėliavimas, krova	Mašinos nuoma 03 mėn.
61017 Parduotuvių priežiūra	Prekių priežiūra

Siekiant nustatyti neįprastus bei labai retus įrašus ieškome žodžių, kurie turi mažiausią dažnį. Žemiau pateikiami 20 rečiausiai bendrojo sąnaudų žurnalo įrašuose pasitaikančių žodžių bei jų dažniai:

Žodis	Dažnis	Žodis	Dažnis
administravimas	1	naujas	4
dokumentu	1	taikymas	4
pastu	1	kurimas	4
siuntima	1	strategijos	4
tvarkymo	2	dubajus	4
pazyma	2	saskaitos	8
auditoriams	3	vadovams	8
berlynas	4	grudu	12
marketingas	4	korteles	12
kodeksas	4	klausimais	16

Nustatyta, kad žodžių kurie būtų pavartoti žurnale vieną kartą yra tik keturi. Žodis *administravimas* susijęs su reklamos socialiniame tinkle *Facebook* administravimu. Net keli reti žodžiai yra iš vieno įrašo: mokestis už dokumentų auditoriams siuntimą paštu. Matome, kad be pastarojo atvejo žodis *auditoriams* panaudotas dar du kartus, jis susijęs su banko pažymos auditoriams užsakymu. Kaip ir *UAB Duona* atveju prie retų žodžių matome priskirtus keleto miestų pavadinimus, jie susiję su dalyvavimu parodose. Informacija apie dalyvavimą parodose atskleista įmonės metiniame pranešime. Didžioji dalis retų įrašų susiję su mokymais pvz.: naujas darbo kodeksas taikymas, strategijos kūrimas, mokymai vadovams ir pan. Peržiūrėjus visus retus įrašus, neįprastų įrašų nenustatyta.

Ieškome dažniausiai pasikartojančių žodžių, kurie padės atskleisti pagrindinius sąnaudų šaltinius. Pateikiami 20 daugiausiai kartų tiriamoje imtyje pasikartojusių žodžių bei jų dažniai:

Žodis	Dažnis	Žodis	Dažnis
kuras	1026	prekiu	472
elektra	962	priežiura	472
nusidevejimas	962	administravima	472
nuoma	806	lesu	472
dok	726	remontas	392
muit	726	plovimas	362
mokestis	628	masinos	332
komisinis	604	valiutos	244
masinu	542	siuntos	226
tyrimai	530	transportas	224

Matome, kad dažniausiai pasikartojantys žodžiai susiję su įprastos veiklos sąnaudų elektra, ilgalaikio turto nusidėvėjimu, degalais.

Net 806 kartus tiriamame žurnale kartojasi žodis *nuoma*. Siekiant nustatyti ką įmonė nuomojasi tiriamo su kokiais žodžiais žodis *nuoma* pasižymi didžiausia koreliacija:

Žodis	<i>masinos</i>	<i>gpm</i>	<i>sandelio</i>	<i>krautuvo</i>	<i>padekli</i>
<b>Koreliacijos dydis</b>	0,63	0,5	0,33	0,3	0,27

Nustatyta, kad įmonė nuomojasi mašinas, sandėlius, krautuvą bei padėklus. Tai, kad žodis *nuoma* turi vidutinio stiprumo 0.5 koreliaciją su *GPM* reiškia, kad dalį nuomojamų objektų įmonė nuomojasi iš fizinių asmenų.

Matome, kad tarp dažniausių žodžių yra žodis *plovimas*. Pagal koreliaciją nustatyta, kad kalbama būtent apie mašinų plovimą.

Žodis	<i>masinu</i>
<b>Koreliacijos dydis</b>	1

Žodžiai *muit* ir *dok* yra muitinės ir dokumentų trumpiniai. Koreliacija tarp šių dviejų žodžių yra lygi 1, tai reiškia, kad šie žodžiai visada vartojami kartu.

Žodis	<i>muit</i>
<b>Koreliacijos dydis</b>	1

Žodžiai *komisinis, administravimą, bei lėšų* turi labai stiprią tarpusavio koreliaciją:

Žodis	<i>mokestis</i>	<i>administravima</i>	<i>lesu</i>	<i>iseinancius</i>	<i>mokejimus</i>	<i>pavedima</i>	<i>tarptautini</i>
<b>Koreliacijos dydis</b>	0,98	0,88	0,88	0,33	0,33	0,31	0,31

Nustatyta, kad šie žodžiai dažniausiai naudojami kartu tokio tipo įrašuose: komisinis mokestis už išeinančius mokėjimus, komisinis mokestis už lėšų administravimą. Visi šie įrašai priskirti DK 61107 *banko sąnaudos*.

Žodį *valiutos* lydi šie žodžiai:

Žodis	<i>koregavimas</i>	<i>kurso</i>	<i>keitimo</i>
<b>Koreliacijos dydis</b>	0,96	0,96	0,28

Šie žodžiai yra iš įrašų susijusių su valiutos kurso koregavimu bei valiutos keitimo sąnaudomis. Tai sufleruoja, kad dalis įmonės produkcijos parduodama į ne euro zonos šalis. Kaip teigiama įmonės metiniame pranešime: įmonė vis didesnę dėmesį skiria eksporto plėtrai, todėl dalyvauja ir reprezentuoja įmonę bei jos produkciją įvairiose parodose.

Gautus tekstinės informacijos analizės rezultatus atvaizduojame dažnių diagrama (žr. 6 priedas), bei žodžių debesimi (žr. 21 pav.)



pav. 21 UAB Sausainis žodžių debesis, 50 dažniausių žodžių



### 3.5.1. Bendrojo žurnalo įrašų klasifikavimas pagal operacijos aprašymą

Pagal verslo apskaitos standartus finansinės ataskaitos turi būti rengiamos taip, kad jų informacijos vartotojai galėtų palyginti jose pateiktą informaciją su kitų ataskaitinių laikotarpių informacija [68]. Tai reiškia, kad konkrečios rūšies ūkinės operacijos kasmet turi būti priskiriamos tai pačiai DK sąskaitai. Puikus būdas neskaitant visų bendrojo žurnalo įrašų patikrinti ar operacijos priskirtos tinkamai DK sąskaitai yra klasifikavimas.

Klasifikavimo uždaviniui spręsti naudosime dvidešimties DK sąnaudų sąskaitų bendrojo žurnalo įrašus. Įrašai bus klasifikuojami pagal jų aprašymus. Klasifikavimo modelio apmokymo imtį sudarys 2015 ir 2016 metų duomenys. Testavimui naudosime 2017 metų analogiškų DK sąskaitų įrašus. Darome prielaidą, kad 2015 ir 2016 metai jau yra audituoti ir šių metų bendrojo žurnalo įrašai yra priskirti tinkamai DK sąskaitai. Norint patikrinti ar 2017 metų įrašai priskirti tinkamoms sąskaitoms auditoriui užteks patikrinti tik tuos įrašus, kurie bus klasifikuojami klaidingai. Klasifikavimui naudosime *SVM* bei *Naivaus Bajeso* klasifikatorius.

Duomenų paruošimo klasifikavimui žingsnyje nustatyta, kad naudojama duomenų imtis nėra subalansuota t.y. atskiros DK sąskaitos turi skirtingus įrašų kiekius. Prieš pradėdant klasifikavimą kaip ir 3.5 skyriuje duomenis transformuojame į dokumentų matricą.

Toliau nagrinėjamas *UAB Duona* atvejis. Pagal bendrą klasifikavimo tikslumą (angl. accuracy) daug geriau pasirodė *SVM* algoritmas su tiesiniu branduoliu negu *Naivaus Bajeso* (žr. 13 lentelė). Algoritmo klasifikavimo tikslumas apibrėžiamas, kaip dalis testavimo duomenų aibės elementų, kuriems tas algoritmas klasę priskyre teisingai.

lentelė 13 Klasifikavimo tikslumas, *UAB Duona* atvejis

Algoritmas	Tikslumas, %
SVM	99.33
Naivaus Bajeso	51.43

Matome, kad *SVM* algoritmas teisingai suklasifikavo net 99,33 % įrašų. Žemiau pateikiama sumaišymų matrica (angl. confusion matrix) čia eilutės simbolizuoja klasifikatoriaus priskirtą klasę, o stulpeliai klasę kuriai iš tiesų priklauso stebėjimas (faktinę).

lentelė 14 UAB Duona sumaišymų matrica

DK sąsk.	6085	60913	61105	61106	61107	61112	61113	61116	61117	61119	61128	61129	61131	61133	61135	61140	6301	6302	6320	636
6085	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60913	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61105	0	0	313	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61106	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61107	0	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61112	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61113	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0
61116	0	0	0	0	0	0	0	138	0	0	0	0	0	0	0	0	0	0	0	0
61117	0	8	0	0	0	0	0	4	160	0	0	0	0	0	0	0	0	0	0	0
61119	0	0	0	0	0	0	0	0	0	145	0	0	0	0	0	0	0	0	0	0
61128	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	0	0	0	0	0
61129	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
61131	0	0	0	0	0	0	0	0	0	0	0	0	51	0	0	0	0	0	0	0
61133	0	0	0	0	0	0	0	0	0	0	0	0	0	53	0	0	0	0	0	0
61135	0	0	0	0	0	0	0	0	0	0	0	0	0	0	133	0	0	0	0	0
61140	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	119	0	0	0	0
6301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	62	0	0	0
6302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0
6320	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	147	0
636	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52
Recall, %	100	85.19	100	100	100	100	100	97.18	100	100	100	100	100	100	100	100	100	100	100	100

Pagal sumaišymų matricą matome, kad tik 12 įrašų buvo priskirti ne tai DK sąskaitai. Taigi šias 12 operacijų reikia patikrinti auditui, jos pateikiamos žemiau:

Tikroji DK sąsk.	Įrašas	Klasifikatoriaus priskirta DK sąsk.
61116	<i>Siunčiami dokumentai į SEB - Vilnius</i>	61117
61116	<i>Siunta BLL Billund, Denmark</i>	61117
61116	<i>Laikiklių pristatymas tr.c.</i>	61117
61116	<i>Laikiklių pristatymas tr.c.</i>	61117
60913	<i>Mokymai elektrokrautuvo vairuotojo</i>	61117
60913	<i>Mokymai elektrokrautuvo vairuotojo</i>	61117
60913	<i>Mokymai elektrokrautuvo vairuotojo</i>	61117
60913	<i>Seminaras pakuotės</i>	61117
60913	<i>Seminaras "Maisto sudėtis"</i>	61117
60913	<i>Seminaras pakuotės</i>	61117
60913	<i>Seminaras pakuotės</i>	61117
60913	<i>Už seminarą finansinė atskaitomybė</i>	61117

Toliau nagrinėjamas UAB Sausainis atvejis. Kaip ir UAB Duona atveju pagal bendrą klasifikavimo tikslumą (angl. accuracy) daug geriau pasirodė SVM algoritmas su tiesiniu branduoliu negu Naivaus Bajeso (žr. 15 lentelė). Matome, kad SVM algoritmas teisingai suklasifikavo net 99,88 % įrašų.

lentelė 15 Klasifikavimo tikslumas, *UAB Sausainis* atvejais

Algoritmas	Tikslumas, %
SVM	99.88
Naivaus Bajeso	64.26

lentelė 16 *UAB Sausainis* sumaišymų matrica

	60205	61000	61003	61004	61005	61006	61007	61008	61010	61011	61013	61014	61015	61017	611040	61106	61107	61111	61114	63003
60205	188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61000	0	215	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
61003	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61004	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61005	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61006	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61007	0	0	0	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0	0	0
61008	0	0	0	0	0	0	0	57	0	0	0	0	0	0	0	0	0	0	0	0
61010	0	0	0	0	0	0	0	0	135	0	0	0	0	0	0	0	0	0	0	0
61011	0	0	0	0	0	0	0	0	0	126	0	0	0	0	0	0	0	0	0	0
61013	0	0	0	0	0	0	0	0	0	0	67	0	0	0	0	0	0	0	0	0
61014	0	0	0	0	0	0	0	0	0	0	0	173	0	0	0	0	0	0	0	0
61015	0	0	0	0	0	0	0	0	0	0	0	0	124	0	0	0	0	0	0	0
61017	0	0	0	0	0	0	0	0	0	0	0	0	0	66	0	0	0	0	0	0
611040	0	0	0	0	0	0	0	0	0	0	0	0	0	0	131	0	0	0	0	0
61106	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	0	0	0	0
61107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	73	0	0	0
61111	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0
61114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83	0
63003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40
Recall, %	100	100	100	100	100	100	100	96,61	100	100	100	100	100	100	100	100	100	100	100	100

Aukščiau pateikiama sumaišymų matrica (angl. confusion matrix) čia eilutės simbolizuoja klasifikatoriaus priskirtą klasę, o stulpeliai klasę kuriai iš tiesų priklauso stebėjimas (faktinę) (žr. 16 lentelę).

Pagal sumaišymų matricą matome, kad tik 2 įrašai buvo priskirti ne tai DK sąskaitai. Taigi šias dvi operacijas reikia patikrinti auditui, jos pateikiamos žemiau:

Tikroji DK sąsk.	Įrašas	Klasifikatoriaus priskirta DK sąsk.
61008	<i>marketingas</i>	61000
61008	<i>marketingas</i>	61000

Apibendrinant, pagal atliktą tyrimą matome, kad teksto analizė puikiai tinka bendrojo žurnalo įrašų testavimui. Pasitelkiant teksto analizę būtų galima analizuoti, ne dalį, bet 100 % įrašų. Nustatyta, kad teksto klasifikavimas gali sumažinti tikrinamą imtį net 99 %. Tai gerokai sumažindamas auditoriaus darbo laiko sąnaudas skirtas ūkinių operacijų priskyrimo tinkamai DK sąskaitai tikrinimui.

## Išvados

1. Atlikus mokslinės ir metodinės literatūros šaltinių analizę nustatyta, kad dažniausiai finansinių ataskaitų audito procedūrose rekomenduojami taikyti didžiųjų duomenų analitikos metodai yra: įvairūs regresijos tipai, teksto tyryba, laiko eilučių analizė, neuroniniai tinklai, atraminių vektorių metodas, Naivaus Bajeso algoritmas, autoregresyvus integruotas slenkantis vidurkis, vizualizacijos. Nustatyta, kad penkiuose audito etapuose iš šešių galima taikyti didžiųjų duomenų analitikos metodus, metodai yra parenkami pagal konkretaus etapo tikslus ir uždavinius. Straipsnių, kuriuose būtų aptariamasi analitinių metodų naudojimas, tęstinių procedūrų etape nenustatyta;
2. sudaryta tyrime naudojamų didžiųjų duomenų analitikos metodų schema, kuri apima technikas patartinas naudoti pirmuose keturiuose audito etapuose. Susipažinimo su klientu etape naudojama aprašomoji statistika. Planavimo etape naudojama pardavimo pajamų laiko eilutės dekompozicija. Pagrindinių procedūrų etape naudojamas Grangerio priežastingumo testas, teksto tyryba bei teksto klasifikavimas. Vertinimo ir peržiūros etape naudojamas vienmačių laiko eilučių prognozavimas. Vizualizacijos pasitelkiamos visuose audito etapuose;
3. teorinės prielaidos ir pasirinkti tyrimo metodai buvo verifikuojami dviejų įmonių pagrindu. Tyrimo rezultatai parodė, kad pasirinkti metodai prisidėjo prie abiejų įmonių audito procedūrų efektyvumo, našumo bei kokybės gerinimo. Galima daryti prielaidą, kad naudoti tyrimo metodai pagerintų audito procedūras ir kitų uždarojo akcinio kapitalo bendrovių audituose. Atlikus *Grangerio* priežastingumo testą nustatyta, kad:
  - iš tiriamų išorės rodiklių tik vidutinis darbo užmokestis turi priežastinį ryšį su *UAB Duona* bendroju pelningumu, bet ir šiuo atveju bendrasis pelningumas yra ir priežastis ir pasekmė;
  - *UAB Sausainis* bendrąjį pelningumą sieja priežastinis ryšys su sviesto, cukraus bei kviečių kainomis, tačiau visais atvejais bendrasis pelningumas yra priežastis, o ne pasekmė;
4. atlikus pardavimo pajamų laiko eilutės analizę nustatyta, kad:
  - prognozuojant *UAB Duona* pardavimo pajamas geriausius paklaidų įverčius gavome taikydami *Holt-Winters* modelį. Rizikos susijusios su veiklos tęstinumo prielaida nenustatyta;
  - prognozuojant *UAB Sausainis* pardavimo pajamas geriausius paklaidų įverčius gavome taikydami *BATS* modelį. Rizikos susijusios su veiklos tęstinumo prielaida nenustatyta;

- atlikus pardavimo pajamų laiko eilutės dekompoziciją nustatyta, kad abiejuose tiriamose įmonėse nėra tiesinio trendo, tačiau vyrauja sezoniškumas, kurio periodas yra metai. Didžiausios pardavimo pajamos kasmet gaunamos vasaros laikotarpiu;
5. nustatyta, kad teksto analizė gali būti taikoma bendrojo žurnalo įrašų testavimui. Pasitelkiant teksto analizę būtų galima analizuoti, ne dalį įrašų, kas atsitiktinės atrankos pagalba audito praktikoje vyksta šiandien, bet 100 % įrašų. Nustatyta, kad teksto klasifikavimas gali sumažinti tikrinamą imtį net 99 %. Taip gerindamas auditorių darbo našumą ir efektyvumą.

### **Rekomendacijos**

1. Siekiant pagerinti audito procedūrų efektyvumą bei kokybę bendrojo žurnalo įrašų testavimui taikyti teksto tyrybos metodus.
2. Bendrojo žurnalo įrašų klasifikavimui taikyti atraminių vektorių metodą bei išbandyti neuroninius tinklus.
3. Pritaikyti laiko eilučių analizę balanso straipsniams, rekomenduotina atsargoms bei skoloms.
4. Atnaujinti tarptautinius audito standartus įtraukiant nurodymus: kaip galima išmatuoti audito įrodymų kiekį, kuriuos analitikai pateikia didžiųjų duomenų kontekste? Kaip metodologiškai pagrįstu būdu šiuos įrodymus galima sujungti su kitų tipų audito įrodymais? Kaip ir kokios kiekybinės priemonės gali būti naudojamos siekiant padėti auditoriui priimti sprendimą dėl audito įrodymų pakankamumo?

## Literatūros sąrašas

1. Ramlukan, R. (2015): How big data and analytics are transforming the audit | EY Reporting - EY - Global. [Žiūrėta 2018-02-23]. Prieiga per:  
<http://www.ey.com/gl/en/services/assurance/ey-reporting-issue-9-how-big-data-and-analytics-are-transforming-the-audit#item4>
2. Tang, F., Norman, C. S., & Vendirzyk, V. P. (2017): Exploring perceptions of data analytics in the internal audit function. *Behaviour & Information Technology*, 36(11), 1125–1136. [Žiūrėta 2018-02-23]. Prieiga per: <https://doi.org/10.1080/0144929X.2017.1355014>
3. Segall, S. R., & Cook, S. J. (2018): Handbook of Research on Big Data Storage and Visualization Techniques - Google knygos. [Žiūrėta 2018-02-23]. Prieiga per:  
[https://books.google.lt/books?id=vQ9MDwAAQBAJ&pg=PR27&lpg=PR27&dq=Handbook+of+Research+on+Big+Data+Storage+and+Visualization+Techniques&source=bl&ots=PFzn91\\_ZQ\\_&sig=jqNqDfUoyFA8AD0SI4aRgCNTbWA&hl=lt&sa=X&ved=0ahUKEwic8-y01LzZAhVRTt8KHVUxDfYQ6AEIUTAG#v=one](https://books.google.lt/books?id=vQ9MDwAAQBAJ&pg=PR27&lpg=PR27&dq=Handbook+of+Research+on+Big+Data+Storage+and+Visualization+Techniques&source=bl&ots=PFzn91_ZQ_&sig=jqNqDfUoyFA8AD0SI4aRgCNTbWA&hl=lt&sa=X&ved=0ahUKEwic8-y01LzZAhVRTt8KHVUxDfYQ6AEIUTAG#v=one)
4. Rusteikienė, J. (2016): Finansinis (teisėtumo) auditas: savivaldybės atvejis. Magistro baigiamasis darbas. Mykolo Riomerio universitetas. [Žiūrėta 2018-02-27]. Prieiga per:  
<https://vb.mruni.eu/object/elaba:15885005/>
5. Lietuvos Respublikos finansų ministerija. *Finansinių ataskaitų auditas*. [Žiūrėta 2018-02-27]. Prieiga per: <https://finmin.lrv.lt/lt/veiklos-sritys/apskaita-ir-atskaitomybe/finansiniu-ataskaitu-auditas>
6. EY, (2015): Big data and analytics in the audit process. [Žiūrėta 2018-02-27]. Prieiga per:  
<http://www.alumni.events.ey.com/pdfs/ey-big-data-and-analytics-in-the-audit-process.pdf>
7. Whithouse, T. (2014). Auditing in the Era of Big Data. Compliance Week, Balandis 2014. [Žiūrėta 2018-02-27]. Prieiga per:  
[http://www.acl.com/pdfs/CW\\_AuditingintheEraofBigData.pdf/](http://www.acl.com/pdfs/CW_AuditingintheEraofBigData.pdf/)
8. Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018): Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40, 102–115. [Žiūrėta 2018-02-27]. Prieiga per:  
<https://doi.org/10.1016/J.ACCLIT.2017.05.003>
9. Alles, M.G. and Gray, G.L. (2016), “Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors”, *International Journal of Accounting Information Systems*, Vol. 22, pp. 44-59 [Žiūrėta 2018-02-27]. Prieiga per:  
[https://www.researchgate.net/publication/305922074\\_Incorporating\\_big\\_data\\_in\\_audits\\_Identifying\\_inhibitors\\_and\\_a\\_research\\_agenda\\_to\\_address\\_those\\_inhibitors](https://www.researchgate.net/publication/305922074_Incorporating_big_data_in_audits_Identifying_inhibitors_and_a_research_agenda_to_address_those_inhibitors)
10. Wang, T. and Cuthbertson, R. (2015), “Eight Issues on Audit Data Analytics We Would Like Researched”, *Journal of Information Systems*, Vol. 29 Iss.1, pp. 155-162. [Žiūrėta 2018-02-27]. Prieiga per: <https://fardapaper.ir/mohavaha/uploads/2017/12/Fardapaper-Incorporating-big-data-in-audits-Identifying-inhibitors-and-a-research-agenda-to-address-those-inhibitors.pdf>  
[https://www.researchgate.net/publication/276076947\\_Eight\\_Issues\\_on\\_Audit\\_Data\\_Analytics\\_We\\_Would\\_Like\\_Researched](https://www.researchgate.net/publication/276076947_Eight_Issues_on_Audit_Data_Analytics_We_Would_Like_Researched)
11. Gartner, (2016): Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest. STAMFORD, Conn., October 4, 2016. [Žiūrėta 2018-02-27]. Prieiga per: <https://www.gartner.com/newsroom/id/3466117>
12. Griffin, P.A. and Wright, A.M. (2015), “Commentaries on Big Data's Importance for Accounting and Auditing”, *Accounting Horizons*, Vol.29 No.2, pp. 377-379 [Žiūrėta 2018-02-27]. Prieiga per:  
[https://www.researchgate.net/publication/276392572\\_Commentaries\\_on\\_Big\\_Data%27s\\_Importance\\_for\\_Accounting\\_and\\_Auditing](https://www.researchgate.net/publication/276392572_Commentaries_on_Big_Data%27s_Importance_for_Accounting_and_Auditing)

13. IAASB (2016): Exploring the Growing Use of Technology in the Audit, with a Focus on Data Analytics, edited by D. A. W. Group. [Žiūrėta 2018-02-27]. Prieiga per: <https://www.ifac.org/publicationsresources/exploring-growing-use-technology-audit-focus-data-analytics>: International
14. Vaicekauskas, D. (2015). Finansinio audito paslaugų kokybės analizė ir vertinimas: daktaro disertacija. Vilniaus universitetas. [Žiūrėta 2018-02-27]. Prieiga per: <https://elaba.lvb.lt/>
15. Kostic, N. (2017): The future of audit: Examining the opportunities and challenges stemming from the use of Big Data Analytics and Blockchain technology in audit practice: magistro baigiamasis projektas. Lundo universitetas. [žiūrėta 2018-02-28]. Prieiga per: [https://www.researchgate.net/publication/318394961\\_The\\_future\\_of\\_audit\\_Examining\\_the\\_opportunities\\_and\\_challenges\\_stemming\\_from\\_the\\_use\\_of\\_Big\\_Data\\_Analytics\\_and\\_Blockchain\\_technology\\_in\\_audit\\_practice\\_Title\\_The\\_future\\_of\\_audit\\_Examining\\_the\\_opportunities\\_and\\_challenges\\_stemming\\_from\\_the\\_use\\_of\\_Big\\_Data\\_Analytics\\_and\\_Blockchain\\_technology\\_in\\_audit\\_practice](https://www.researchgate.net/publication/318394961_The_future_of_audit_Examining_the_opportunities_and_challenges_stemming_from_the_use_of_Big_Data_Analytics_and_Blockchain_technology_in_audit_practice_Title_The_future_of_audit_Examining_the_opportunities_and_challenges_stemming_from_the_use_of_Big_Data_Analytics_and_Blockchain_technology_in_audit_practice_Title_The_future_of_audit_Examining_the_opportunities_and_challenges_stemming_from_the_use_of_Big_Data_Analytics_and_Blockchain_technology_in_audit_practice)
16. Zhang, Juan; Yang, Xiongsheng; Appelbaum, Deniz. (2015): *Toward Effective Big Data Analysis in Continuous Auditing*; Accounting Horizons Jun2015, Vol. 29 Issue 2, p469-476. 8p. [žiūrėta 2018-03-02]. Prieiga per: [https://www.researchgate.net/publication/276391068\\_Toward\\_Effective\\_Big\\_Data\\_Analysis\\_in\\_Continuous\\_Auditing](https://www.researchgate.net/publication/276391068_Toward_Effective_Big_Data_Analysis_in_Continuous_Auditing)
17. Yoon, Kyunghye; Hoogduin, Lucas; Zhang, Li. (2015): Big Data as Complementary Audit Evidence; Accounting Horizons June 2015, Vol. 29, No. 2, pp. 431-438. [žiūrėta 2018-03-02]. Prieiga per: [https://www.researchgate.net/publication/276391168\\_Big\\_Data\\_as\\_Complementary\\_Audit\\_Evidence](https://www.researchgate.net/publication/276391168_Big_Data_as_Complementary_Audit_Evidence)
18. Lietuvos auditorių rūmai. Tarptautiniai audito standartai. [žiūrėta 2018-03-02]. Prieiga per: <http://lar.lt/www/new/page.php?326>
19. Cangemi, M.P. Addressing the C-Level Question: *How Effectively are Assurance Functions Contributing and Using Automated Analytics?*; The EDP Audit, Control, and Security Newsletter 25 May 2017. VOL. 55, NO. 5. ISSN: 0736-6981. [žiūrėta 2018-03-02]. Prieiga per: <https://www.tandfonline.com/toc/uedp20/current>
20. Appelbaum, D., Kogan, A., & Vasarhelyi A. M. (2017): Big Data and Analytics in the Modern Audit Engagement: Research Needs. Article in Auditing A Journal of Practice & Theory February 2017. [žiūrėta 2018-03-10]. Prieiga per: [https://www.researchgate.net/publication/313286738\\_Big\\_Data\\_and\\_Analytics\\_in\\_the\\_Modern\\_Audit\\_Engagement\\_Research\\_Needs](https://www.researchgate.net/publication/313286738_Big_Data_and_Analytics_in_the_Modern_Audit_Engagement_Research_Needs)
21. Cangemi, M.P. & Sinnett, W.M. (2016): Data Analytics and Financial Compliance: How Technology is Changing Audit and Business Systems. Financial Executives Research Foundation, Inc. ISBN: 978-1-61509-211-6. [žiūrėta 2018-03-10]. Prieiga per: <https://www.financialexecutives.org/Research/Publications/2016/Data-Analytics-and-Financial-Compliance-How-Techn.aspx>
22. Appelbaum, D., Kogan, A., & Vasarhelyi A. M. (2018): Analytical Procedures in External Auditing: A Comprehensive Literature Survey and Framework for External Audit Analytics. Journal of Accounting Literature 40, January 2018. [žiūrėta 2018-03-11]. Prieiga per: [https://www.researchgate.net/publication/322841767\\_Analytical\\_Procedures\\_in\\_External\\_Auditing\\_A\\_Comprehensive\\_Literature\\_Survey\\_and\\_Framework\\_for\\_External\\_Audit\\_Analytics](https://www.researchgate.net/publication/322841767_Analytical_Procedures_in_External_Auditing_A_Comprehensive_Literature_Survey_and_Framework_for_External_Audit_Analytics)
23. Aykut, B., Klaus, S., Jan, K. Wiebke, G. (2016): Perspectives on Big Data and Business Intelligence Technologies in the Context of Audit Tasks. The Bucharest University of Economic Studies. [žiūrėta 2018-03-11]. Prieiga per: [https://www.researchgate.net/publication/305386022\\_Perspectives\\_on\\_Big\\_Data\\_and\\_Business\\_Intelligence\\_Technologies\\_in\\_the\\_Context\\_of\\_Audit\\_Tasks](https://www.researchgate.net/publication/305386022_Perspectives_on_Big_Data_and_Business_Intelligence_Technologies_in_the_Context_of_Audit_Tasks)

24. Brown-Liburd, H.; Vasarhelyi, M. A., (2015): Big Data and audit evidence. *Journal of Emerging Technologies in Accounting* 12 (1): 1– 16. [žiūrėta 2018-03-11]. Prieiga per: [https://www.researchgate.net/publication/290221256\\_Big\\_Data\\_and\\_Audit\\_Evidence](https://www.researchgate.net/publication/290221256_Big_Data_and_Audit_Evidence)
25. Kuenkaikaew, S, (2013): Predictive Audit Analytics: Evolving to a New Era; Daktaro disertacija, Naujojo Džersio valstybinis universitetas. [žiūrėta 2018-03-11]. Prieiga per: <https://rucore.libraries.rutgers.edu/rutgers-lib/41494/>
26. KPMG, (2017): *Audit 2025, the future is now*; Forbes insights (March). [žiūrėta 2018-03-11]. Prieiga per: <https://assets.kpmg.com/content/dam/kpmg/us/pdf/2017/03/us-audit-2025-final-report.pdf>
27. Gray, I., Manson, S. (2008): *The Audit Process– Principles, Practice and Cases*. ISBN: 978-1-84480-678-2. [žiūrėta 2018-03-12]. Prieiga per: <https://books.google.lt/books?id=NwPRL6QVr9EC&pg=PA237&lpg=PA237&dq=audit+stages&source=bl&ots=G9bgTDt8hi&sig=BMU8HmblAtGoV-IUWXnx4rNeUp8&hl=lt&sa=X&ved=0ahUKEwjO8er8tOfZAhWMDcAKHRtoAFM4ChDoAQhtMAw#v=onepage&q=audit%20stages&f=false>
28. Clements, J. (2017): Steps of an Audit Engagement. [žiūrėta 2018-03-12]. Prieiga per: <https://bizfluent.com/info-8786746-steps-audit-engagement.html>
29. Stanisloviėnė, Eglė (2006): *Audito procesas ir praktinis jo taikymas* (UAB „Baldų alėja“ pavyzdžiu). Magistro baigiamasis projektas. Šiaulių universitetas. [žiūrėta 2018-03-12]. Prieiga per: <http://gs.elaba.lt/object/elaba:1965512/>
30. Lietuvos Respublikos valstybės kontrolė. Valstybinio audito reikalavimai. [žiūrėta 2018-03-19]. Prieiga per: <https://www.vkontrolė.lt/page.aspx?id=32>
31. Crumbley, L. D; Rezaee, Z; Ziegenfuss, D.E. (2004): *U.s. Master Auditing Guide*. CHH Incorporated, Chicago. ISBN 0-8080-1191-X. [žiūrėta 2018-03-21]. Prieiga per: <https://books.google.lt/books?id=MclAITZZBVQC&pg=PA148&lpg=PA148&dq=audit+engagement+stages&source=bl&ots=Djo2Z8UqXT&sig=AYVvk95VAbi5BQCy0pia9G18Abj0&hl=lt&sa=X&ved=0ahUKEwjw6uOQgf7ZAhVDDiwKHTO6CwM4ChDoAQhrMAw#v=onepage&q=audit%20engagement%20stages&f=false>
32. Caldeira, M. A., Gassenferth, W., Machado, M. A., Santos, D. J. (2015): Auditing Vehicles Claims using Neural Networks. [žiūrėta 2018-03-21]. Prieiga per: <https://www.sciencedirect.com/science/article/pii/S1877050915014830>
33. Appelbaum, D. (2016): Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box as Reliable Evidence. Rutgers, The State University of New Jersey, Newark. *JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING* Vol. 13, No. 1 pp. 17–36. [žiūrėta 2018-03-24]. Prieiga per: [https://www.researchgate.net/publication/300562117\\_SECURING\\_BIG\\_DATA\\_PROVENANCE\\_FOR\\_AUDITORS\\_THE\\_BIG\\_DATA\\_PROVENANCE\\_BLACK\\_BOX](https://www.researchgate.net/publication/300562117_SECURING_BIG_DATA_PROVENANCE_FOR_AUDITORS_THE_BIG_DATA_PROVENANCE_BLACK_BOX)
34. Ruzgas, T., Jakubėlienė, K & Buivyte, A. (2016): Big Data Mining and Knowledge Discovery. Kauno technologijos universitetas. *Journal of Communications Technology, Electronics and Computer Science*, Issue 9, 2016. ISSN 2457-905X. [žiūrėta 2018-04-12]. Prieiga per: <http://www.jctecs.com/index.php/com/article/view/134>
35. Kurasova, O. (2014): Programinės sistemos duomenų tyrybos mokymui. Lietuvos matematikų draugijos darbai, ser. B 55 t., 2014, 60-65, ISSN 0132-2818. [žiūrėta 2018-04-12]. Prieiga per: <https://www.mii.lt/LMR/B/2014/55B12.pdf>
36. Zubova, J. (2015): DIDELĖS APIMTIES DUOMENŲ VIZUALI ANALIZĖ. Mokslinė ataskaita MII-DS-09P-15-9, Vilniaus universitetas. [žiūrėta 2018-04-12]. Prieiga per: [http://old.mii.lt/files/09p\\_zubova\\_ataskaita2015.pdf](http://old.mii.lt/files/09p_zubova_ataskaita2015.pdf)

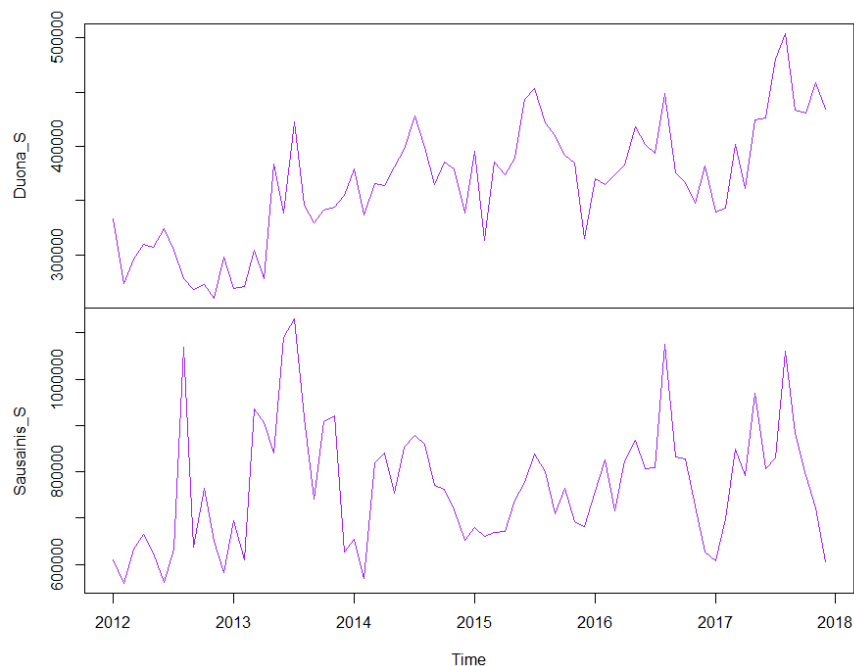


37. Ruzgas, T., Dabulytė-Bagdonavičienė, J. (2017): Business Intelligence for Big Data Analytics. Kaunas University of Technology International Journal of Computer Applications Technology and Research Volume 6–Issue 01, 01-08, 2017, ISSN:-2319–8656. [žiūrėta 2018-04-12]. Prieiga per:  
[https://www.researchgate.net/profile/Tomas\\_Ruzgas/publication/312269363\\_Business\\_Intelligence\\_for\\_Big\\_Data\\_Analytics/links/58bd5c3492851c471d5a01f5/Business-Intelligence-for-Big-Data-Analytics.pdf](https://www.researchgate.net/profile/Tomas_Ruzgas/publication/312269363_Business_Intelligence_for_Big_Data_Analytics/links/58bd5c3492851c471d5a01f5/Business-Intelligence-for-Big-Data-Analytics.pdf)
38. KANTARDZIC M. (2011): Data Mining – Concepts, Models, Methods and Algorithms. [žiūrėta 2018-04-13] ISBN 978-0-470-89045-5 Prieiga per:  
<https://books.google.lt/books?id=ZZ716v0CvRMC&lpg=PA1&dq=data%20mining%20importance&lr=&hl=lt&pg=PR4#v=onepage&q=data%20mining%20importance&f=false>
39. LAROSE D. T. (2005): Discovering Knowledge in Data: An Introduction to Data Mining. Wiley-Interscience. ISBN 0-471-66657-2 (cloth). [žiūrėta 2018-04-12]. Prieiga per:  
[http://secs.ac.in/wp-content/CSE\\_PORTAL/DataMining\\_Daniel.pdf](http://secs.ac.in/wp-content/CSE_PORTAL/DataMining_Daniel.pdf)
40. HAN J., M. KAMBER and J. PEI. (2012) Data Mining: Concepts and Techniques. Amsterdam: Elsevier. ISBN 978-0-12-381479-1 [žiūrėta 2018-04-13] Prieiga per:  
<https://books.google.lt/books?id=pQws07tdpjoC&printsec=frontcover&dq=data+mining&hl=lt&sa=X&ved=0ahUKEwiCy9bm8LfaAhUDWiwKHUtbCIMQ6AEIUzAG#v=onepage&q=data%20mining&f=false>
41. Weiss S.M., Indurkha N. (1998) Predictive Data Mining. Morgan Kaufmann, San Francisco. ISBN 1-55860-403-0 [žiūrėta 2018-04-13] Prieiga per:  
[https://books.google.lt/books?hl=lt&lr=&id=xzVD8C2YpnQC&oi=fnd&pg=PR11&dq=data+mining+predictive+descriptive&ots=IO47qBI8NN&sig=XOYRYbHrRMwv\\_e3RKJNcxQXhDFE&redir\\_esc=y#v=onepage&q=data%20mining%20predictive%20descriptive&f=false](https://books.google.lt/books?hl=lt&lr=&id=xzVD8C2YpnQC&oi=fnd&pg=PR11&dq=data+mining+predictive+descriptive&ots=IO47qBI8NN&sig=XOYRYbHrRMwv_e3RKJNcxQXhDFE&redir_esc=y#v=onepage&q=data%20mining%20predictive%20descriptive&f=false)
42. Alapaydin, E. (2016): Machine Learning: The New AI. MIT Press, Cambridge, Massachusetts. ISBN 0262529513. [žiūrėta 2018-04-14] Prieiga per:  
[https://books.google.lt/books?id=AGQ4DQAAQBAJ&hl=lt&source=gbs\\_navlinks\\_s](https://books.google.lt/books?id=AGQ4DQAAQBAJ&hl=lt&source=gbs_navlinks_s)
43. Ratner, B. (2017): Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, Third Edition. CRC Press, ISBN 1351652389. [žiūrėta 2018-04-14]. Prieiga per:  
[https://books.google.lt/books?id=ulAsDwAAQBAJ&printsec=frontcover&hl=lt&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.lt/books?id=ulAsDwAAQBAJ&printsec=frontcover&hl=lt&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
44. ZHOU, Lina, Shimei PAN, Jianwu WANG and Athanasios VASILAKOS (2017): Machine learning on big data: Opportunities and challenges. Neurocomputing (237), 350-361 [žiūrėta 2018-04-14]. Prieiga per:  
[https://www.researchgate.net/publication/312398923\\_Machine\\_Learning\\_on\\_Big\\_Data\\_Opportunities\\_and\\_Challenges](https://www.researchgate.net/publication/312398923_Machine_Learning_on_Big_Data_Opportunities_and_Challenges)
45. Alapaydin, E. (2014): Introduction to Machine Learning. MIT Press, ISBN 0262028182. [žiūrėta 2018-04-14]. Prieiga per:  
[https://books.google.lt/books?id=NP5bBAAAQBAJ&dq=machine+learning&hl=lt&source=gbs\\_navlinks\\_s](https://books.google.lt/books?id=NP5bBAAAQBAJ&dq=machine+learning&hl=lt&source=gbs_navlinks_s)
46. Guha, R. (2017): Exploring the Field of Text Mining. International Journal of Computer Applications (0975 – 8887) Volume 177 – No.4. [žiūrėta 2018-04-14]. Prieiga per:  
[https://www.researchgate.net/publication/321102244\\_Exploring\\_the\\_Field\\_of\\_Text\\_Mining](https://www.researchgate.net/publication/321102244_Exploring_the_Field_of_Text_Mining)
47. Weiss, M. S., Indurkha, N., Zhang, T., Damerau, F. J. (2010): Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer Science & Business Media. ISBN 0-387-95433-3. [žiūrėta 2018-04-14]. Prieiga per:  
[https://books.google.lt/books?id=NZteXd4qf9sC&printsec=frontcover&hl=lt&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.lt/books?id=NZteXd4qf9sC&printsec=frontcover&hl=lt&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)

48. IBM: About text mining. [žiūrėta 2018-04-14]. Prieiga per:  
[https://www.ibm.com/support/knowledgecenter/en/SS3RA7\\_18.1.0/ta\\_guide\\_ddita/textmining/shared\\_entities/tm\\_intro\\_tm\\_defined.html](https://www.ibm.com/support/knowledgecenter/en/SS3RA7_18.1.0/ta_guide_ddita/textmining/shared_entities/tm_intro_tm_defined.html)
49. Gray, G. L., R. S. Debreceeny. (2014): A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems* 15 (4):357-380. [žiūrėta 2018-04-14]. Prieiga per: doi: 10.1016/j.accinf.2014.05.006.
50. Szabó, I., Ternai K. (2016) Semantic Audit Application for Analyzing Business Processes. In: Tjoa A., Xu L., Raffai M., Novak N. (eds) Research and Practical Issues of Enterprise Information Systems. CONFENIS 2016. Lecture Notes in Business Information Processing, vol 268. Springer, Cham. ISBN 978-3-319-49944-4. [žiūrėta 2018-04-14]. Prieiga per: [https://link.springer.com/content/pdf/10.1007%2F978-3-319-49944-4\\_1.pdf](https://link.springer.com/content/pdf/10.1007%2F978-3-319-49944-4_1.pdf)
51. Miner, G., Elder, J., Nisbet, B. (2012): Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Academic Press, ISBN:978-0-12-386979-1. [žiūrėta 2018-04-15]. Prieiga per:  
[https://books.google.lt/books?id=SM94BMsy50gC&printsec=frontcover&hl=lt&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.lt/books?id=SM94BMsy50gC&printsec=frontcover&hl=lt&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
52. Jaillet, S., Teisseire M., Chauche, J., Prince, V. (2003): Classification of Documents by Content. Conference: Cognitive Informatics, 2003. Proceedings. The Second IEEE International Conference. DOI 10.1109/COGINF.2003.1225983. [žiūrėta 2018-04-15]. Prieiga per:  
[https://www.researchgate.net/publication/4031861\\_Classification\\_of\\_Documents\\_by\\_Content](https://www.researchgate.net/publication/4031861_Classification_of_Documents_by_Content)
53. Balys, V. (2009): MOKSLINĖS TERMINIJOS MATEMATINIAI MODELIAI IR JŲ TAIKYMAS LEIDINIŲ KLASIFIKAVIME. Daktaro disertacija. Vilniaus Gedimino technikos universitetas. VGTU leidykla TECHNIKA. ISBN 978-9955-28-467-3. [žiūrėta 2018-04-15]. Prieiga per: [http://old.mii.lt/files/mii\\_dis\\_2009\\_balys.pdf](http://old.mii.lt/files/mii_dis_2009_balys.pdf)
54. McCallum, A., Nigam, K. (2001): A Comparison of Event Models for Naive Bayes Text Classification. [žiūrėta 2018-04-15]. Prieiga per:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.9324&rep=rep1&type=pdf>
55. Thorsten, J. (2005): Text categorization with Support Vector Machines: Learning with many relevant features. Dortmundo Universitetas. European Conference on Machine Learning ECML 1998: Machine Learning: ECML-98 pp 137-142. [žiūrėta 2018-04-15]. Prieiga per: <https://link.springer.com/content/pdf/10.1007%2FBFb0026683.pdf>
56. Sassano, M. (2003): Virtual Examples for Text Classification with Support Vector Machines. DOI 10.3115/1119355.1119382. [žiūrėta 2018-04-15]. Prieiga per:  
[https://www.researchgate.net/publication/2921663\\_Virtual\\_Examples\\_for\\_Text\\_Classification\\_with\\_Support\\_Vector\\_Machines](https://www.researchgate.net/publication/2921663_Virtual_Examples_for_Text_Classification_with_Support_Vector_Machines)
57. Stabingienė, L. (2014): Ekonometrika. Klaipėdos universitetas, matematikos ir statistikos katedra. Elektroninė knyga. [žiūrėta 2018-04-15]. Prieiga per:  
[http://www.ilab.lt/stabingiene/sk8\\_1.html](http://www.ilab.lt/stabingiene/sk8_1.html)
58. Lukoševičiūtė, K. (2012): Chaotinių procesų rekonstravimo bei algebrinių sekų modeliai laiko eilučių prognozavime. Daktaro disertacija: Fiziniai Mokslai, Informatika (09P). Kauno technologijos universitetas. [žiūrėta 2018-04-15]. Prieiga per:  
[https://www.personalas.ktu.lt/~kriluko/Disertacija/Disertacija\\_Kristinos\\_Lukoseviciutes.pdf](https://www.personalas.ktu.lt/~kriluko/Disertacija/Disertacija_Kristinos_Lukoseviciutes.pdf)
59. Leupus, R. (2016): Ekonometrija II. [žiūrėta 2018-04-20]. Prieiga per:  
[http://uosis.mif.vu.lt/~remis/fin\\_laik\\_eil\\_3.pdf](http://uosis.mif.vu.lt/~remis/fin_laik_eil_3.pdf)
60. Danilenko, S. (2009): Makroekonominių procesų poveikio akcijų rinkai tyrimas. EKONOMIKA IR VADYBA: 2009. 14. ISSN 1822-6515. [žiūrėta 2018-04-20]. Prieiga per: <https://www.researchgate.net/publication/237696192>
61. Virbukaitė, S. (2011): Vektorinės autoregresijos (VAR) modelių taikymas ekonomikos procesams analizuoti. 14-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas –

- Lietuvos ateitis“, 2011 metų teminės konferencijos straipsnių rinkinys: matematika. ISBN 978-9955-28-836-7 [žiūrėta 2018-04-20]. Prieiga per: <http://dspace.vgtu.lt/bitstream/1/731/1/Virbukaite.pdf>
62. Kvedaras, V. (2005): Taikomoji laiko eilučių ekonometrija. [žiūrėta 2018-04-20]. Prieiga per: [http://web.vu.lt/mif/v.kvedaras/files/2013/09/Konspektas\\_2005.pdf](http://web.vu.lt/mif/v.kvedaras/files/2013/09/Konspektas_2005.pdf)
63. Athanasopoulos, G., Hyndman, R., Song, H., Wu, D. (2010), The tourism forecasting competition. *International Journal of Forecasting*, Vol. 27, Iss: 3, pp. 822-844. [žiūrėta 2018-04-27]. Prieiga per: [https://ac.els-cdn.com/S016920701000107X/1-s2.0-S016920701000107X-main.pdf?\\_tid=fab1176f-dbfd-4cbc-bce4-b00aa036abae&acdnat=1526574624\\_f789ab6d7101f8ab6ece36c06ceff47d](https://ac.els-cdn.com/S016920701000107X/1-s2.0-S016920701000107X-main.pdf?_tid=fab1176f-dbfd-4cbc-bce4-b00aa036abae&acdnat=1526574624_f789ab6d7101f8ab6ece36c06ceff47d)
64. Boguslauskas, V., Bliedienė, R. (2012): Ekonometrija. Laiko eilučių modeliai: laboratoriniai darbai : mokomoji knyga. Kaunas, Technologija. ISBN: 9786090205952. [žiūrėta 2018-04-27]. Prieiga per: [https://www.ebooks.ktu.lt/eb/621/ekonometrija\\_laiko\\_eiluciu\\_modeliai\\_laboratoriniai\\_darbai/](https://www.ebooks.ktu.lt/eb/621/ekonometrija_laiko_eiluciu_modeliai_laboratoriniai_darbai/)
65. De Livera, A., Hyndman, J. R., Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106:496, 1513-1527. [žiūrėta 2018-04-27]. Prieiga per: <https://www.tandfonline.com/doi/pdf/10.1198/jasa.2011.tm09771?needAccess=true>
66. George Salijeni, Anna Samsonova-Taddei & Stuart Turley (2018): Big Data and changes in audit technology: contemplating a research agenda, *Accounting and Business Research*, DOI: 10.1080/00014788.2018.1459458. ISSN: 0001-4788 (Print) 2159-4260 (Online). [žiūrėta 2018-05-05]. Prieiga per: [https://www.tandfonline.com/doi/full/10.1080/00014788.2018.1459458?utm\\_medium=email&utm\\_source=EmailStudio\\_JB&utm\\_campaign=JME00769+VIB\\_2731146](https://www.tandfonline.com/doi/full/10.1080/00014788.2018.1459458?utm_medium=email&utm_source=EmailStudio_JB&utm_campaign=JME00769+VIB_2731146)
67. Chatfield, C., Yar, M. (1988): Holt-Winters Forecasting: Some Practical Issues. *Journal of the Royal Statistical Society Series D (The Statistician)* 37(2). DOI 10.2307/2348687. [žiūrėta 2018-05-10]. Prieiga per: [https://www.researchgate.net/publication/246382642\\_HoltWinters\\_Forecasting\\_Some\\_Practical\\_Issues](https://www.researchgate.net/publication/246382642_HoltWinters_Forecasting_Some_Practical_Issues)
68. AUDITO, APSKAITOS, TURTO VERTINIMO IR NEMOKUMO VALDYMO TARNYBA. Verslo apskaitos standartai. [žiūrėta 2018-05-10]. Prieiga per: <http://www.avnt.lt/veiklos-sritys/apskaita/verslo-apskaitos-standartai/priimti-standartai/>

### Įmonių savikainos kaita pamėnesiui



pav. 22 savikainos kaita pamėnesiui

### Pardavimo pajamų laiko eilutės dekompozicija

#### UAB Duona atvejis

Stacionarumo hipotezės tikrinimo rezultatai:

```
> adf.test(Duona_P, k=12)
```

Augmented Dickey-Fuller Test

```
data: Duona_P  
Dickey-Fuller = -2.4868, Lag order = 12, p-value = 0.3771  
alternative hypothesis: stationary
```

#### UAB Sausainis pajamų laiko eilutės dekompozicija

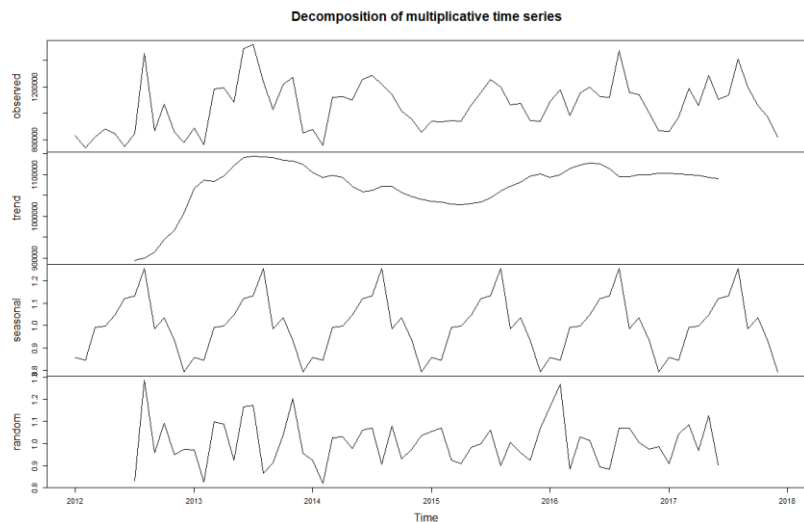
Tikriname laiko eilutės stacionarumą:

Augmented Dickey-Fuller Test

```
data: Sausainis_P  
Dickey-Fuller = -3.121, Lag order = 12, p-value = 0.1189  
alternative hypothesis: stationary;
```

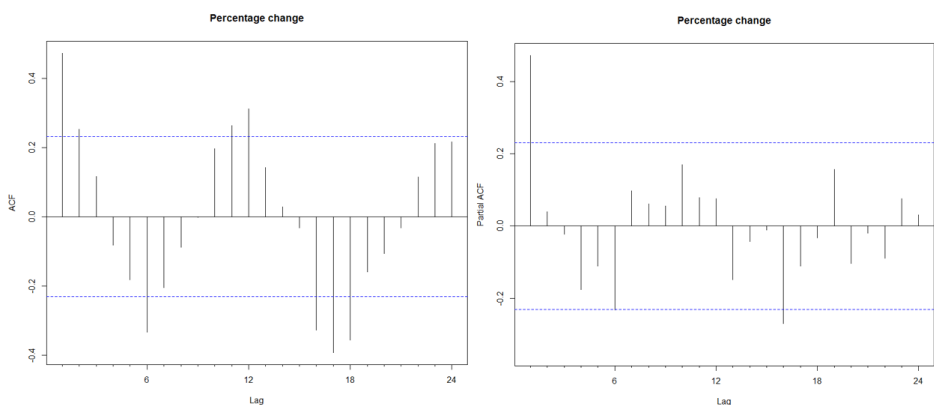
Pagal *Augmented Dickey-Fuller* testo rezultatus matome, kad  $p=0.1189 > 0.05$ , taigi  $H_0$  hipotezės atmesti negalime.

Dekompozicijai naudojome *multiplikatyvinį* modelį, kadangi laiko eilutė yra nestacionari (žr. 21 pav.). Nustatyta, kad *UAB Sausainis* pardavimo pajamų laiko eilutėje nėra tiesinio trendo, tačiau egzistuoja sezoniškumas, kurio periodas metai.



pav. 23 *UAB Sausainis* pajamų laiko eilutės dekompozicija

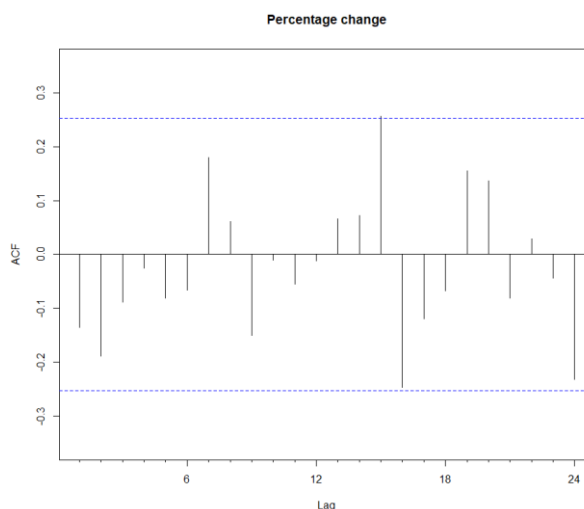
Laiko eilutės autokoreliacinės funkcijos (angl. autocorrelation function) vertinimui naudosime procedūras *acf* ir *pacf* (žr. 22 pav.).



pav. 24 *UAB Sausainis* *acf* ir *pacf* procedūrų grafikai

ACF grafike matome išryškėjusią sezoniškumo, įtaką.

Tiriame ar likusi nepaaiškinama laiko eilutės dalis (toliau, *random* komponentė) yra baltasis triukšmas (žr. 23 pav.). Nustatyta, kad tiriama seka yra baltasis triukšmas, kadangi visos autokoreliacinės funkcijos reikšmės, yra melsvos juostos viduje, be to *random* komponentės grafike nėra trendo ar sezoniškumo pėdsakų.



pav. 25 UAB Sausainis random komponentės ACF grafikas

3 priedas

### Tiriamų rodiklių informacijos šaltiniai

Rodiklis	Matavimo vienetas	Informacijos šaltinis
Sviesto kaina	EUR/100kg	<a href="https://ec.europa.eu/agriculture/sites/agriculture/.../eu-historical-price-series_en.xls">https://ec.europa.eu/agriculture/sites/agriculture/.../eu-historical-price-series_en.xls</a>
Kviečių kaina	EUR/t	<a href="http://www.produktukainos.lt/?mid=86&amp;id=4587">http://www.produktukainos.lt/?mid=86&amp;id=4587</a>
Cukraus kaina	EUR/t	<a href="https://ec.europa.eu/agriculture/sites/agriculture/files/market-observatory/sugar/doc/price-industrial_en.pdf">https://ec.europa.eu/agriculture/sites/agriculture/files/market-observatory/sugar/doc/price-industrial_en.pdf</a>
Vidutinis darbo užmokestis Lietuvoje	EUR	<a href="https://osp.stat.gov.lt/statistiniu-rodikliu-analize#/">https://osp.stat.gov.lt/statistiniu-rodikliu-analize#/</a>

4 priedas

### Veiksnių darančių poveikį UAB Sausainis bendrajam pelningumui nustatymas



pav. 26 Koreliacinio ryšio stiprumo įvertinimo diagrama, UAB Sausainis atvejis

Pateikiama koreliacinio ryšio stiprumo įvertinimo diagrama (žr. 24 pav.). Pagal diagramą matome, kad didžiausias neigiamas ryšys BP sieja su cukrumi. Taip pat nustatytas, neigiamas -0,37 ryšys tarp BP ir kviečių kainos. Teigiamas koreliacinis ryšys sieja BP su DU, tai gana keista, tačiau tai paaiškinama tuo, kad didžioji dalis darbininkų atlyginimų yra artima minimaliam darbo užmokesčiui, taigi vidutinio darbo užmokesčio augimas turi mažai įtakos savikainos augimui.

lentelė 17 koreliacinės analizės ir *Grangerio* testo rezultatai, *UAB Sausainis* atvejis

Testas (su BP)	Sviestas	Cukrus	Kviečiai	DU
Koreliacija	-0.17	-0.48	-0.37	0.32
Grangerio testas	—>	—>	—>	—

Remiantis atlikta priežastingumo analize galime atmesti kintamąjį DU, kaip nedarantį reikšmingo poveikio BP. Nustatyta, kad BP sieja priežastinis ryšys su sviesto, cukraus bei kviečių kainomis, tačiau visais atvejais BP yra priežastis, o ne pasekmė (žr. 25pav, 26pav, 27pav.).

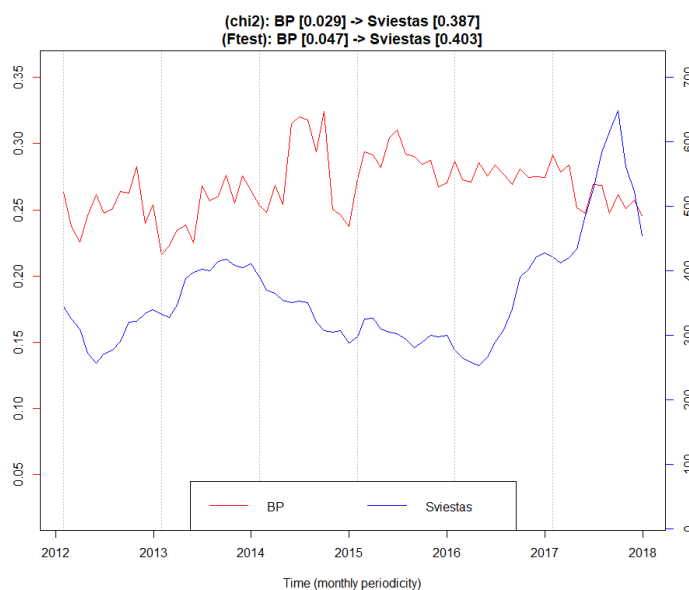
Pagal AR modelio išraišką matome vėlinimų ilgį tarp tiriamų kintamųjų:

AR(7+1)  
 (chi2): BP [0.029] -> sviestas [0.387]  
 (Ftest): BP [0.047] -> sviestas [0.403]

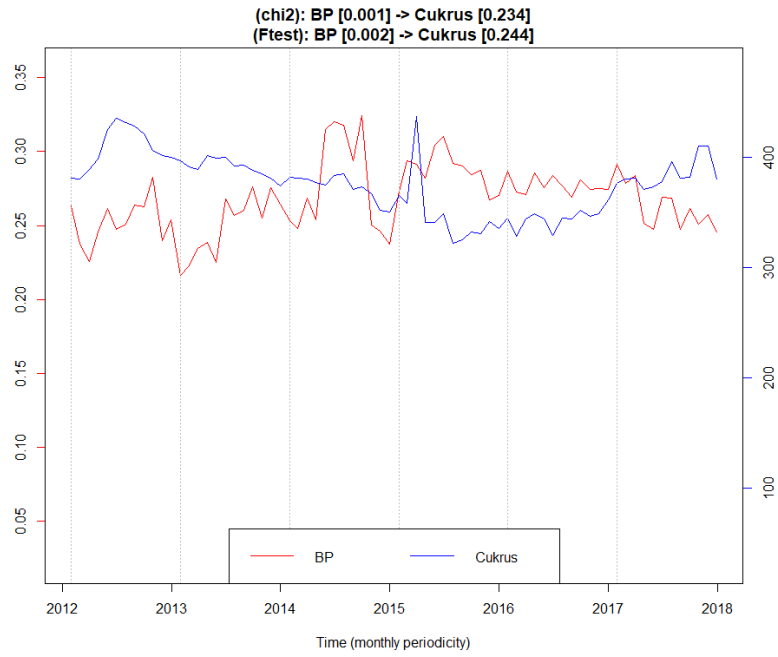
AR(3+1)  
 (chi2): BP [0.001] -> cukrus [0.234]  
 (Ftest): BP [0.002] -> cukrus [0.244]

AR(7+1)  
 (chi2): BP [0.005] -> kviečiai [0.164]  
 (Ftest): BP [0.014] -> kviečiai [0.191]

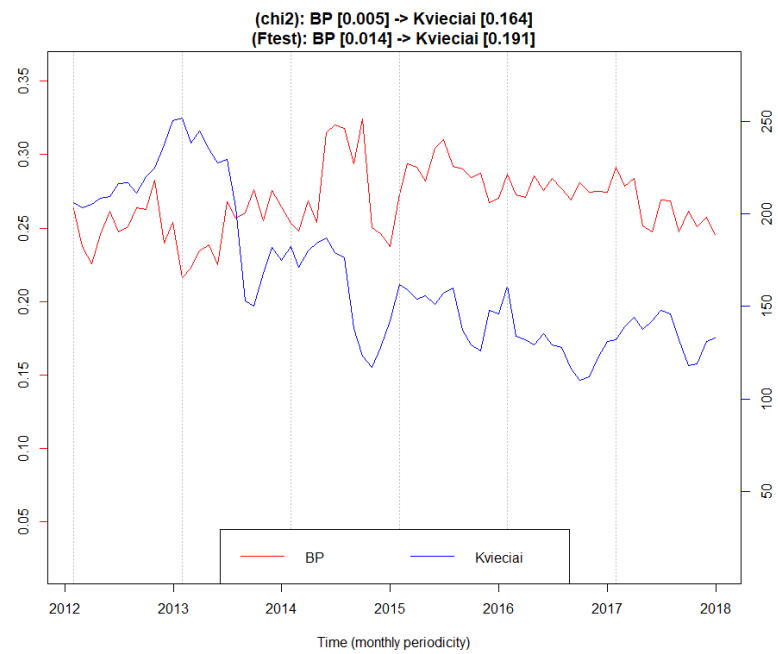
AR(1+1)  
 (chi2): BP [0.229] - DU [0.149]  
 (Ftest): BP [0.233] - DU [0.153]



pav. 27 *Grangerio* testo rezultatai tarp bendrojo pelningumo ir sviesto kainos



pav. 28 *Grangerio* testo rezultatai tarp bendrojo pelningumo ir cukraus kainos



pav. 29 *Grangerio* testo rezultatai tarp bendrojo pelningumo ir kviečių kainos



## Pardavimo pajamų prognozės 2018 metams

lentelė 18 *UAB Duona* pardavimo pajamų prognozė 2018 metams

Method name	saus. 2018	vas. 2018	kov. 2018	bal. 2018	geg. 2018	birž. 2018	liep. 2018	rugp. 2018	rugs. 2018
HoltWinters	685852	652450	684365	706860	819175	729324	871051	883243	711919
Linear regression model	679291	638908	663410	695134	757118	728203	802816	797839	690531
Basic structural model	691003	650647	700047	683608	812305	681250	848543	866825	695610
ARIMA(0,1,2)(0,1,1)[12]	676556	655043	682277	706549	804547	723667	847059	866560	710819
Seasonal naive method	606770	568718	630231	617278	753058	629240	807176	822809	644479
ETS(M,Ad,M)	674842	631619	656916	689484	761955	722498	805720	797060	684201
BATS(0.005, {0,0}, -, {12})	664557	624030	648039	680361	743569	708104	786797	778836	665628
Linear regression model	706527	708810	711094	713377	715661	717944	720228	722512	724795
Theta	676536	625360	653677	680920	765016	715938	798350	788061	680942
Mean	623177	623177	623177	623177	623177	623177	623177	623177	623177
Random walk with drift	675055	677208	679362	681516	683670	685823	687977	690131	692285
ARIMA(0,1,1)	685994	685994	685994	685994	685994	685994	685994	685994	685994
Random walk	672901	672901	672901	672901	672901	672901	672901	672901	672901

(18 lentelės tęsinys)

Method name	spal. 2018	lapkr. 2018	gruod. 2018
HoltWinters	739153	759562	719661
Linear regression model	723658	711162	699961
Basic structural model	705319	765714	717580
ARIMA(0,1,2)(0,1,1)[12]	736150	747698	719665
Seasonal naive method	651546	719965	672901
ETS(M,Ad,M)	720806	698844	689612
BATS(0.005, {0,0}, -, {12})	698222	685588	670956
Linear regression model	727079	729362	731646
Theta	723779	689970	686123
Mean	623177	623177	623177
Random walk with drift	694438	696592	698746
ARIMA(0,1,1)	685994	685994	685994
Random walk	672901	672901	672901

Pateikiama *UAB Duona* pardavimo pajamų prognozės 2018 metams gautos *Holt-Winters* metodu pasikliautinių intervalų įverčiai (žr. 18 lentelė).

lentelė 19 UAB Duona pardavimo pajamų prognozės pasikliautinieji intervalai

	Point Forecast	Lo 80	Hi 80	Lo 90	Hi 90
Jan 2018	685852.4	620917.8	750787.0	602509.8	769195.0
Feb 2018	652449.7	583490.7	721408.8	563941.8	740957.7
Mar 2018	684365.4	611297.9	757432.9	590584.3	778146.6
Apr 2018	706860.3	629602.2	784118.4	607700.6	806020.0
May 2018	819174.8	737645.8	900703.8	714533.4	923816.1
Jun 2018	729324.4	643445.8	815203.0	619100.4	839548.4
Jul 2018	871050.9	780745.5	961356.3	755145.2	986956.7
Aug 2018	883242.8	788434.9	978050.8	761558.1	1004927.6
Sep 2018	711918.6	612533.7	811303.5	584359.5	839477.8
Oct 2018	739152.6	635117.7	843187.6	605625.2	872680.1
Nov 2018	759562.4	650805.4	868319.3	619974.4	899150.4
Dec 2018	719661.4	606111.7	833211.1	573921.9	865400.8

lentelė 20 UAB Sausainis pardavimo pajamų prognozė 2018 metams

Method name	saus. 2018	vas. 2018	kov. 2018	bal. 2018	geg. 2018	birž. 2018	liep. 2018	rugp. 2018
BATS(1, {0,0}, -, {12})	915110	911198	1045952	1065432	1122222	1151800	1198864	1338082
ETS(M,N,A)	905633	896692	1056249	1063624	1116632	1194996	1212772	1320161
HoltWinters	923095	986603	1066685	1026918	1164218	1076233	1127451	1369488
Seasonal naive method	858073	969699	1186106	1059748	1288546	1103384	1133895	1409487
Linear regression model	1001461	977477	1127599	1146637	1185768	1235249	1282876	1420133
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean	901351	982349	1080626	1041642	1130352	1063109	1075190	1178024
ARIMA(1,0,0) with non-zero mean	936734	995412	1023986	1037900	1044676	1047976	1049582	1050365
Mean	1057166	1057166	1057166	1057166	1057166	1057166	1057166	1057166
Linear regression model	1137814	1140024	1142233	1144443	1146652	1148862	1151071	1153281
Basic structural model	910996	927238	959566	939445	965068	949584	945687	1011306
Theta	918534	907087	1066651	1074072	1128719	1206318	1223103	1355812
Random walk with drift	816084	815932	815779	815627	815475	815323	815170	815018
Random walk	816236	816236	816236	816236	816236	816236	816236	816236

(19 lentelės tęsinys)

Method name	rugs. 2018	spal. 2018	lapkr. 2018	gruod. 2018
BATS(1, {0,0}, -, {12})	1095595	1092326	1001568	857782
ETS(M,N,A)	1049629	1098186	995245	840833
HoltWinters	1126097	1045007	953272	823312
Seasonal naive method	1195246	1057836	973226	816236
Linear regression model	1163029	1184553	1088970	939683
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean	1098487	1047446	1016006	957629
ARIMA(1,0,0) with non-zero mean	1050746	1050931	1051022	1051066
Mean	1057166	1057166	1057166	1057166
Linear regression model	1155490	1157700	1159910	1162119
Basic structural model	944534	981504	984817	951626
Theta	1066142	1118897	1011264	860706

Random walk with drift	814866	814714	814561	814409
Random walk	816236	816236	816236	816236

Pateikiami *UAB Sausainis* pardavimo pajamų prognozės 2018 metams gautos *BATS* metodu pasikliautinių intervalų įverčiai (žr. 20 lentelė).

lentelė 21 *UAB Sausainis* pardavimo pajamų prognozės pasikliautiniai intervalai

	Point Forecast	Lo 80	Hi 80	Lo 90	Hi 90
Jan 2018	915110.3	762758.5	1067462	719568.9	1110652
Feb 2018	911198.2	753511.4	1068885	708809.4	1113587
Mar 2018	1045952.5	883105.3	1208800	836940.4	1254965
Apr 2018	1065432.4	897583.4	1233281	850000.5	1280864
May 2018	1122221.5	949515.5	1294928	900555.8	1343887
Jun 2018	1151800.0	974369.8	1329230	924070.9	1379529
Juļ 2018	1198864.1	1016832.5	1380896	965229.0	1432499
Aug 2018	1338081.9	1151562.2	1524602	1098686.4	1577477
Sep 2018	1095595.1	904692.8	1286497	850574.7	1340616
Oct 2018	1092325.7	897139.2	1287512	841806.6	1342845
Nov 2018	1001567.9	802189.2	1200947	745668.2	1257468
Dec 2018	857781.9	654297.4	1061266	596612.4	1118951

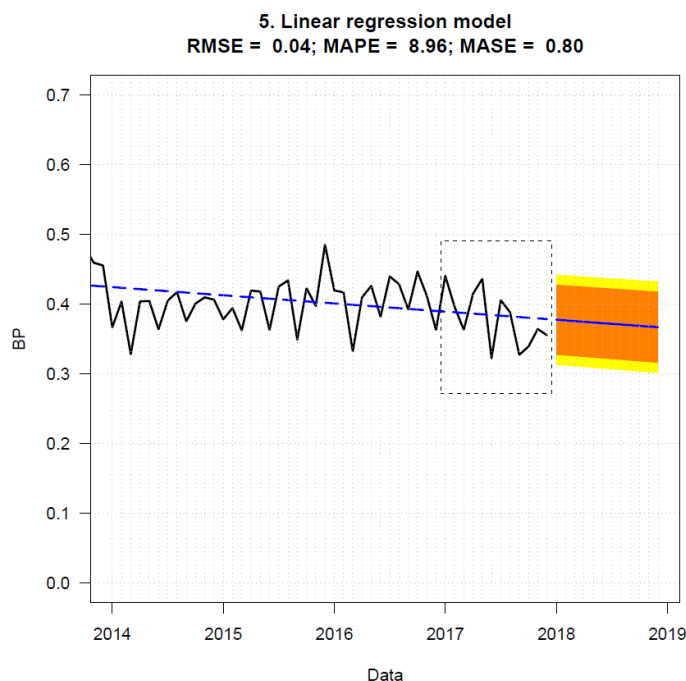
6 priedas

### Bendrojo pelningumo prognozės 2018 metams

Toliau pateikiama lentelė su *UAB Duona* bendrojo pelningumo prognozėms išbandytais modeliais bei klaidų tipų *RMSE*, *MAPE*, *MASE* įverčiais (žr. 19 lentelė). Kaip matyti iš lentelės geriausius paklaidų įverčius gavome taikydami *tiesinės regresijos su trendu* modelį.

lentelė 22 *UAB Duona* bendrojo pelningumo prognozės klaidų įverčiai

Nr.	Metodo pavadinimas	RMSE	MAPE	MASE
1	Linear regression model	0.04	8.96	0.8
2	HoltWinters	0.04	8.81	0.8
3	Basic structural model	0.04	8.66	0.77
4	Linear regression model	0.04	9.74	0.89
5	Random walk with drift	0.04	8.78	0.84
6	Random walk	0.04	8.78	0.84
7	Theta	0.04	9.92	0.87
8	BATS(1, {0,0}, -, {12})	0.04	10.08	0.91
9	ARIMA(0,1,1)	0.04	10.43	0.91
10	ARIMA(0,1,1)(2,0,0)[12]	0.04	10.43	0.91
11	ETS(M,N,N)	0.05	10.59	0.92
12	Seasonal naive method	0.05	10.48	0.91
13	Mean	0.06	13.2	1.14



pav. 30 UAB Duona bendrojo pelningumo prognozė

28 pav. pateikiamos bendrojo pelningumo prognozės 2018 metams taikant tiesinės regresijos su trendu modelį. Remiantis modelio gautomis prognozėmis galime teigti, kad 2018 metais bendrasis pelningumas turėtų svyruoti tarp  $\sim 0,36$  ir  $\sim 0,37$ . Žemiau pateiktos BP prognozės naudojant tiesinės regresijos modelį ir pasikliautinieji intervalai, kai reikšmingumo lygmuo 0,1 ir 0,2 atitinkamai (žr. 22 lentelė).

lentelė 23 UAB Duona bendrojo pelningumo prognozė

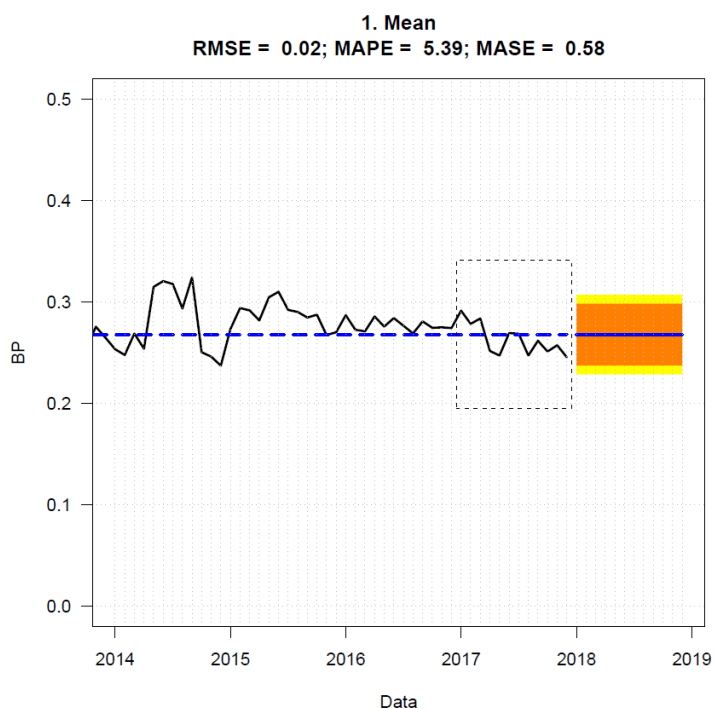
	Point Forecast	Lo 80	Hi 80	Lo 90	Hi 90
Jan 2018	0.3774472	0.3272076	0.4276868	0.3127174	0.4421770
Feb 2018	0.3764730	0.3261769	0.4267691	0.3116703	0.4412757
Mar 2018	0.3754988	0.3251447	0.4258529	0.3106214	0.4403762
Apr 2018	0.3745246	0.3241111	0.4249382	0.3095706	0.4394786
May 2018	0.3735505	0.3230760	0.4240249	0.3085180	0.4385829
Jun 2018	0.3725763	0.3220395	0.4231131	0.3074635	0.4376891
Juī 2018	0.3716021	0.3210015	0.4222027	0.3064071	0.4367971
Aug 2018	0.3706279	0.3199621	0.4212937	0.3053489	0.4359069
Sep 2018	0.3696537	0.3189212	0.4203862	0.3042888	0.4350186
Oct 2018	0.3686795	0.3178790	0.4194801	0.3032269	0.4341322
Nov 2018	0.3677054	0.3168353	0.4185754	0.3021632	0.4332475
Dec 2018	0.3667312	0.3157902	0.4176721	0.3010976	0.4323647

Toliau pateikiama lentelė su UAB Sausainis bendrojo pelningumo prognozėms išbandytais modeliais bei klaidų tipų *RMSE*, *MAPE*, *MASE* įverčiais (žr. 23 lentelė). Kaip matyti iš lentelės geriausius paklaidų įverčius gavome taikydami paprastą vidurkio metodą.

lentelė 24 UAB Sausainis bendrojo pelningumo prognozės klaidų įverčiai

Nr.	Metodo pavadinimas	RMSE	MAPE	MASE
1	Mean	0.02	5.39	0.58
2	Random walk	0.02	6.49	0.69
3	ARIMA(1,1,1)	0.02	6.52	0.69
4	Basic structural model	0.02	6.53	0.69
5	ARIMA(1,1,1)	0.02	6.55	0.69
6	BATS(1, {0,0}, -, -)	0.02	6.55	0.69
7	ETS(M,N,N)	0.02	6.56	0.69
8	Random walk with drift	0.02	6.89	0.73
9	Seasonal naive method	0.02	7.13	0.75
10	Theta	0.02	7.38	0.78
11	HoltWinters	0.02	7.79	0.83
12	Linear regression model	0.04	12.68	1.33
13	Linear regression model	0.04	12.96	1.36

29 pav. pateikiamos bendrojo pelningumo prognozės 2018 metams prognozuojant vidurkio metodu. Remiantis modelio gautomis prognozėmis galime teigti, kad 2018 metais bendrasis pelningumas turėtų būti apie ~0,27. Žemiau pateiktos BP prognozės ir pasikliautinieji intervalai, kai reikšmingumo lygmuo 0,1 ir 0,2 atitinkamai (žr. 24 lentelė).



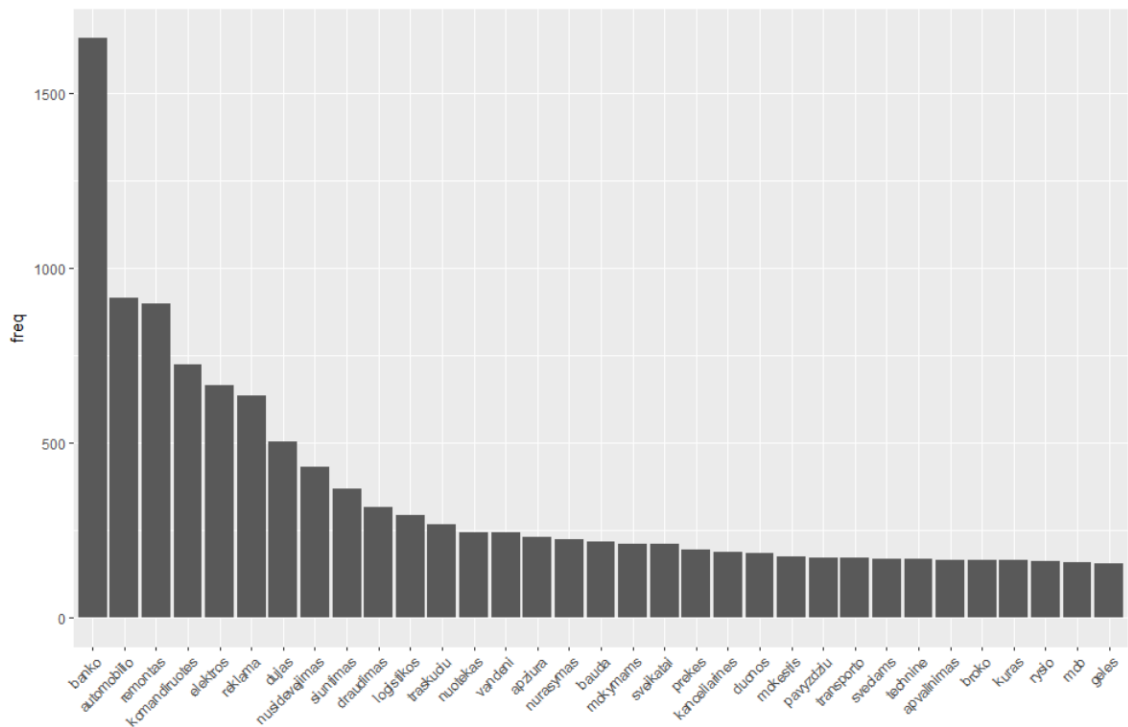
pav. 31 UAB Sausainis bendrojo pelningumo prognozė

lentelė 25 UAB Sausainis bendrojo pelningumo prognozė

	Point Forecast	Lo 80	Hi 80	Lo 90	Hi 90
Jan 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Feb 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Mar 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Apr 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
May 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Jun 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Juī 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Aug 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Sep 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Oct 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Nov 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956
Dec 2018	0.2677784	0.2373694	0.2981874	0.2286008	0.306956

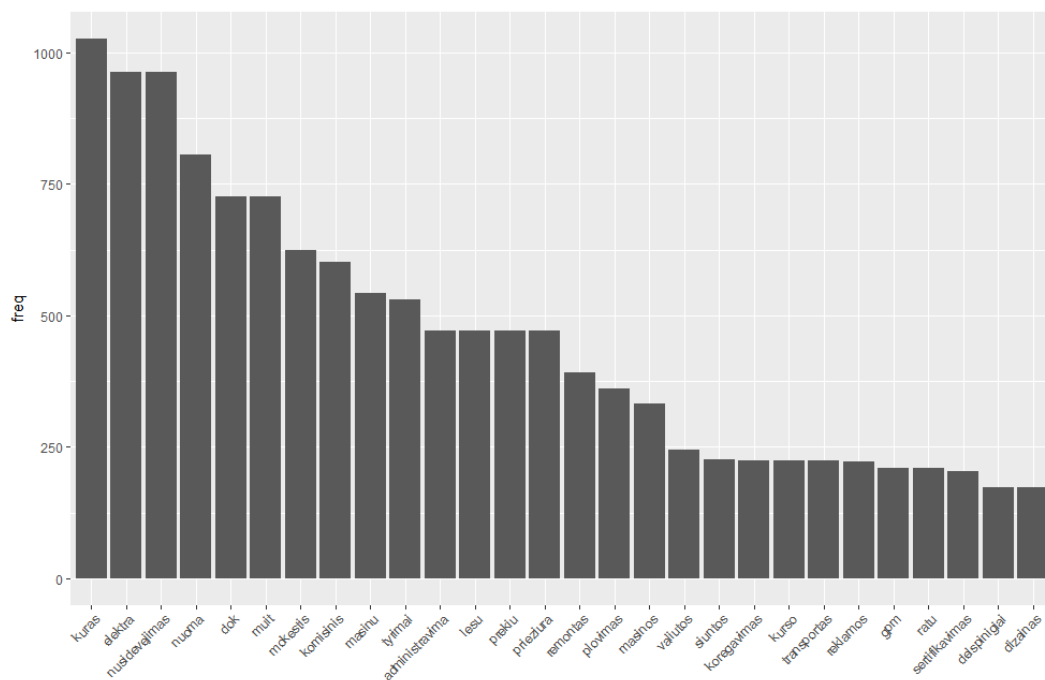
6 priedas

UAB Duona teksto analizės rezultatų vaizdavimas histograma, pateikiami 50 dažniausiai išlaidų bendrajame žurnale pasitaikančių žodžių.



pav. 32 UAB Duona 50 dažniausių išlaidų bendrojo žurnalo žodžių

*UAB Sausainis* teksto analizės rezultatų vaizdavimas histograma, pateikiami 50 dažniausiai išlaidų bendrajame žurnale pasitaikančių žodžių.



pav. 33 *UAB Sausainis* 50 dažniausių išlaidų bendrojo žurnalo žodžių