



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

**Internetinės reklamos technologijų kompanijos pelno  
prognozės neapibrėžtumo tyrimas**

Baigiamasis magistro studijų projektas

---

**Vilija Tigrūdytė**

Projekto autorė

**Prof. Robertas ALZBUTAS**

Vadovas

**Doc. Alfreda ŠAPKAUSKIENĖ**

Vadovė

---

**Vilnius, 2019**



**Kauno technologijos universitetas**  
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

# **Internetinės reklamos technologijų kompanijos pelno prognozės neapibrėžtumo tyrimas**

Baigiamasis magistro studijų projektas  
Didžiųjų verslo duomenų analitika (6213AX001)

---

**Vilija Tigrūdytė**

Projekto autorė

**Prof. Robertas Alzbutas**

Vadovas

**Doc. Alfreda Šapkauskienė**

Vadovė

**Dr. Doc. Mantas Landauskas**

Recenzentas

**Dr. Aura Drakšaitė**

Recenzentė

**Vilnius, 2019**



**Kauno technologijos universitetas**  
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS  
Vilija Tigrūdytė

**Internetinės reklamos technologijų kompanijos pelno  
prognozės neapibrėžtumo tyrimas**  
Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Vilijos Tigrūdytės, baigiamasis projektas tema „Internetinės reklamos technologijų kompanijos pelno prognozės neapibrėžtumo tyrimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjusi.

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

---

(vardą ir pavardę įrašyti ranka)

---

(parašas)

Tigrūdytė, Vilija. Internetinės reklamos technologijų kompanijos pelno prognozės neapibrėžtumo tyrimas. Magistro studijų baigiamasis projektas, vadovas prof. Robertas Alzbutas; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų kryptių grupė): Didžiųjų verslo duomenų analitika (6213AX001).

Reikšminiai žodžiai: internetinės reklamos technologijos, pelno prognozė, duomenų neapibrėžtumas.

Vilnius, 2019. 65 p.

### **Santrauka**

Šiame tyrime nagrinėjami neapibrėžtumo šaltiniai, prognozuojantys internetinės reklamos technologijų kompanijos pelną trumpuoju laikotarpiu, t.y. iki einamojo mėnesio pabaigos. Pirmajame etape aprašomas įmonės verslo modelis, turimi produktai, bei pelno skaičiavimai, įvardijami galimi veiksniai, įtakojantys pelno pokyčius. Toliau apžvelgiami esami tyrimai siekiant įvertinti jau sukurtus finansų prognozių metodus, taip pat duomenų neapibrėžtumo panaudojimo atvejai. Tiriamoje dalyje aplikama duomenų statistinė aprašomoji analizė naudojantis nagrinėjamos įmonės realiais duomenimis. Šiems duomenų rinkiniams pritaikomi matematiniai laiko eilučių tyrimo metodai siekiant atrasti tinkamiausią pelno prognozės įverčiui paskaičiuoti metodą bei įvertinti duomenų neapibrėžtumo įtaką rezultatui.

Duomenų rinkiniams buvo pritaikyti šie metodai – laiko eilutės, tiesinė regresija. Pritaikius duomenų neapibrėžtumui tikrinti skirtus matematinius metodus, įvertinti jų įtaką galutinės prognozės rezultatui ir tikslumui.

Tigrūdytė, Vilija. Survey of profit forecast uncertainty for online advertising technology company. Master's Final Degree Project supervisor prof. Robertas Alzbutas; Mathematics and applied science faculty, Kaunas University of Technology.

Study field and area (study field group): Big Data Analytics (6213AX001).

Keywords: Online marketing technologies, revenue forecasting, data uncertainty

Vilnius, 2019. 65 pages.

### **Summary**

This study explores sources of uncertainty when calculating the prediction of profit for an online advertising technology company in the short term, i.e. until the end of the current month. The first stage describes the company's business model, available products, and profit calculations, and identifies potential factors that affect profit changes. The following is an overview of existing studies to assess existing financial forecasting methods, as well as the use of data uncertainty. The descriptive analysis of the data is based on the actual data of the company concerned. Digital datasets are applied to these datasets to discover the most appropriate forecasting method and to assess the impact of data uncertainty on the result.

The following methods were used for data sets: time series, linear regression. After applying mathematical methods for checking the uncertainty of data, evaluate their influence on the result and accuracy of the final forecast.

## Turinys

<b>Lentelių sąrašas .....</b>	<b>7</b>
<b>Paveikslų sąrašas .....</b>	<b>8</b>
<b>Įvadas .....</b>	<b>9</b>
<b>1. Literatūros apžvalga .....</b>	<b>11</b>
1.1. Internetinės reklamos technologijų kompanijos verslo modelis.....	11
1.2. Realaus laiko aukcionas .....	12
1.2.1. Realaus laiko aukciono tipai.....	12
1.2.2. Realaus laiko aukciono kainodaros modeliai .....	12
1.3. Internetinės reklamos technologijų įmonės pelno modeliai .....	13
1.4. Pelno prognozių tyrimai.....	14
1.5. Matematinų modelių parametrų neapibrėžtumo ir jautrumo rezultatui tyrimai .....	15
1.5.1. Kiekybinė neapibrėžtumo analizė.....	16
1.5.2. Kokybinė neapibrėžtumo analizė.....	17
<b>2. Metodologija, metodų aprašymas.....</b>	<b>20</b>
2.1. Pelno prognozės lygtis .....	20
2.2. Slenkantis vidurkis.....	21
2.3. Eksponentinis glodinimas.....	21
2.4. Croston` s metodas.....	22
2.5. ARIMA laiko eilučių prognozavimas .....	22
2.6. Pelno prognozės sėkmingumo (tikslumo) vertinimas .....	23
2.7. Neapibrėžtumo tyrimai.....	25
<b>3. Tiriamoji dalis.....</b>	<b>26</b>
3.1. Aprašomoji statistinė duomenų analizė .....	26
3.2. Realaus laiko aukciono pelno prognozė.....	28
3.2.1. Išlaidų prognozė.....	28
3.2.2. Pelningumo prognozė.....	34
3.3. Pelno prognozės neapibrėžtumo tyrimas.....	40
3.3.1. Pradinių duomenų neapibrėžtumas .....	40
3.3.2. Neapibrėžtumo šaltiniai.....	40
3.3.3. Modelio parametrų jautrumo vertinimas .....	43
<b>Tyrimo apžvalga ir išvados.....</b>	<b>47</b>
<b>Literatūros sąrašas .....</b>	<b>48</b>
<b>Priedai .....</b>	<b>50</b>

## Lentelių sąrašas

1 lentelė. Pasitikėjimo laipsnių kategorijų vertinimo kriterijai.....	17
2 lentelė. Neapibrėžtumo ir jautrumo tyrimų panaudojimo atvejai.....	19
3 lentelė. Koreliacijos koeficientų reikšmių skalė.....	25
4 lentelė. Klientų išlaidų dieninių vidutinių prognozių koeficientų modelių lygtys.....	34
5 lentelė. Pelningumo prognozių metodų rezultatai.....	37
6 lentelė. Klientų išlaidų koeficientų modelių skirstinių rezultatai.....	42
7 lentelė. Koreliacijų koeficientų reikšmės.....	45

## Paveikslų sąrašas

1 pav. Įmonės klientų išlaidos realaus laiko aukcione dalyvaujančiose kampanijose.....	26
2 pav. Registruotų parodymų kiekis klientų, dalyvaujančių išankstinio susitarimo kampanijose ....	27
3 pav. Pelnas, gautas už realaus laiko aukciono kampanijas .....	27
4 pav. Pelnas, gautas už išankstinio susitarimo kampanijas .....	28
5 pav. Dieninis klientų išlaidų pokytis mėnesio eigoje .....	29
6 pav. Klientų vidutinės dienos išlaidos .....	30
7 pav. Klientų vidutinių dienos išlaidų koeficientų mėnesiniai vidurkiai .....	31
8 pav. Sausio mėnesio koeficientų kitimo lygtis.....	32
9 pav. Vasario mėnesio koeficientų kitimo lygtis .....	32
10 pav. Gruodžio mėnesio koeficientų kitimo lygtis.....	32
11 pav. 30 dienų mėnesių koeficientų kitimo lygtis.....	33
12 pav. 31 dienos mėnesių koeficientų kitimo lygtis .....	33
13 pav. Pelningumo kaita.....	35
14 pav. Pelningumo pokytis su pažymėtais pirmu ir trečiu kvartiliais.....	35
15 pav. Pelningumo išskirtys .....	36
16 pav. Pelningumo rodiklių tankio histograma .....	36
17 pav. Laiko eilutės išskaidymas pagrindinėmis komponentėmis.....	37
18 pav. Laiko eilutės likučių autokoreliacijos testo rezultatas.....	37
19 pav. Pelningumo prognozė vidurkio metodu .....	38
20 pav. Slenkančio vidurkio vidutinių kvadratinių paklaidų įverčiai.....	39
21 pav. Pelningumo prognozė naudojant eksponentinio glodinimo metodą .....	39
22 pav. 31 dienos mėnesių kelintų išlaidų koeficientų tankio diagrama.....	41
23 pav. 30 dienų mėnesių klientų išlaidų koeficientų tankio diagrama.....	41
24 pav. Sausio mėnesio klientų išlaidų koeficientų tankio diagrama.....	41
25 pav. Vasario mėnesio klientų išlaidų koeficientų tankio diagrama.....	41
26 pav. Gruodžio mėnesio klientų išlaidų koeficientų tankio diagrama.....	42
27 pav. Mėnesio dieninių vidutinių išlaidų koeficientų neapibrėžtumo srities amplitudė .....	43
28 pav. Pelno prognozės ir išlaidų sklaidos diagrama.....	44
29 pav. Pelno prognozės ir pelningumo rodiklio sklaidos diagrama.....	44
30 pav. Pelno prognozės neapibrėžtumo intervalo kaita mėnesio eigoje .....	46
31 pav. Sausio mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis.....	56
32 pav. Vasario mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis.....	56
33 pav. Gruodžio mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis.....	56
34 pav. 31 dienos mėnesių klientų dieninių išlaidų koeficientų neapibrėžtumo sritis .....	57
35 pav. 30 dienų mėnesių klientų dieninių išlaidų koeficientų neapibrėžtumo sritis.....	57



## **Įvadas**

Finansinių rodiklių modeliavimas yra skirtas argumentuotam sprendimų priėmimui. Visada versle yra problema, kurią reikia spręsti, dėl ko yra kuriami įvairūs skaičiavimai, išvedamos formulės, jog būtų galima pagrįstai priimti verslo sprendimus.

Kadangi pelnas nėra vienintelis svarbiausias įmonės finansinis rodiklis, svarbu turėti patikimą ir tikslų pelno planą tam, kad įmonė sėkmingai veiktų. Ilgalaikių biudžetų planavimas yra vienas didžiausių iššūkiu bet kurioje įmonėje, o tam atlikti, įmonėje yra remiamasi finansinių prognozių modeliais [1]. Skaitmeninės reklamos technologijų įmonė suinteresuota prognozuoti pelną, gaunamą iš klientų skaitmeninių reklamos kampanijų ir išsiaiškinti nuo ko jis labiausiai priklauso. Duomenų pagrindu sudarytą pelno prognozę šiuo metu teikia keli departamentai, kurie remiasi tais pačiais duomenimis, tačiau naudoja skirtingas priemones ir metodus.

Tobulėjant skaitmeninio mokymosi algoritmams ir populiarėjant įžvalgų ir prognozavimo analitikai, įmonių finansų analitikai gali tiksliau prognozuoti įmonės pelną. Tai leistų sudaryti tikslesnius ilgalaikius biudžetus, planuoti išlaidas, tiksliau įvertinti būsimą darbuotojų poreikį ir pan. Trumpalaikės (einamojo mėnesio) prognozės įprastai yra naudojamos pardavimų vadybininkų, kurių tikslas yra stebėti pardavimų tendencijas mėnesio eigoje ir žinoti, ar pavyks einamąjį mėnesį pasiekti nustatytą biudžetą. Dažniausiai didelėse kompanijose vadovams gana sudėtinga paprastai ir greitai sužinoti bendrą kompanijos pardavimų tendenciją einamuoju mėnesiu. Nagrinėjamos įmonės atveju, tokie pardavimai yra gana nenuspėjami, nėra fiksuoti mėnesio pradžioje, sudaromi išankstiniai sandoriai, kurie yra preliminarūs.

Pelno planavimas atliekamas remiantis istoriniais duomenimis, todėl ne visada yra tikslus. Pritaikius matematinius prognozių metodus, skirtus prognozuoti būsimas reikšmes (koks bus ateityje pelno dydis, atsižvelgiant į esamas tendencijas), galima gana tiksliai tai paskaičiuoti. Duomenys, kurie yra kaupiami analizuojamos įmonės duomenų saugyklose yra labai didelės apimties, kuriuose daugybė įvairios informacijos. Radus tinkamus veiksnus, kurie veikia pelno pokytį, būtų galima gana tiksliai prognozuoti pelną ir išvengti klaidų nustatant biudžetus, bei kuo anksčiau reaguoti, kai jie nėra pasiekiami.

Skaičiavimui ir modeliavimui naudojama MS Excel programa, skirta greitam nedidelių duomenų apdorojimui, rezultatų grafiniam vaizdavimui ir skaičiavimui. Papildomam grafiniam vizualizavimui naudojamas *Tableau* įrankis. Taip pat integruotas programinės įrangos rinkinys R, skirtas duomenų valdymui, skaičiavimui ir grafiniam vaizdavimui.

### **Darbo problematika**

Kaip įvertinti pelno prognozės neapibrėžtumo sritį.

### **Darbo tikslas**

Įvertinti internetinės reklamos technologijų kompanijos pelno prognozės neapibrėžtumą, jo šaltinius ir prognozės įverčio jautrumą atsižvelgiant į parametrų neapibrėžtumą.

## **Darbo uždaviniai**

- Remiantis literatūros šaltiniais atlikti internetinės reklamos technologijų tyrimus, apžvelgti pajamų šaltinius ir verslo modelį, įvardinti pagrindinius kintamuosius, įtakojančius pelną. Nustatyti dažniausiai taikomus metodus finansų prognozei modeliuoti bei globaliai sprendžiamas finansų prognozių problemas. Remiantis publikuotais moksliniais tyrimais, apžvelgti pagrindinius duomenų neapibrėžtumo analizių metodus, jų taikymo sritis.
- Atlikti aprašomąją duomenų statistinę analizę internetinės reklamos technologijų kompanijos atveju ir pritaikyti pelno prognozės metodą, tinkamiausią einamojo mėnesio pelno prognozei skaičiuoti.
- Įvertinti pelno prognozės metodo neapibrėžtumo šaltinius, neapibrėžtumo intervalą. Įvertinti pelno prognozės parametrų jautrumą atsižvelgiant į parametrų indeksus, taip nustatant reikšmingumo lygį pelno prognozės įverčiui.

## 1. Literatūros apžvalga

### 1.1. Internetinės reklamos technologijų kompanijos verslo modelis

Reklamos pateikimas internete (angl. *Ad serving*) apibūdina technologiją ir paslaugą, kuri skelbimus pateikia interneto svetainėse. Skelbimų pateikimo technologijų bendrovė teikia programinę įrangą interneto svetainėms ir reklamuotojams, kuri suteikia galimybę lengvai pateikti skelbimus, juos suskaičiuoti, sekti ir parinkti tuos, kurie suteiks galimybę svetainės savininkui ar reklamuotojui pasiekti didžiausią pelną, taip pat stebėti skirtingų reklamos kampanijų pažangą. Skelbimų serveriai skirstomi į du tipus - leidėjo skelbimų serverius ir reklamuotojo (arba trečiosios šalies) skelbimų serverius.

Internetinė reklama yra pagrindinis socialinių tinklų ir paieškos sistemų pajamų šaltinis pvz., „Google“ [10]. Skelbimų mainų platformos leidžia reklamuotojams pasirinkti tikslią auditoriją naudojant demografinius rodiklius, interesus ir naršymo istorijas. Pritaikius pažangius matematinius metodus vartotojo elgsenos tyrimams, tikslinė auditorija, atsižvelgiant į reklamuotojo kampanijų tikslus, gali būti pasiekama dar tiksliau [5]. Kiekvieną kartą vartotojui patekus į tinklalapį arba naudojantis paieškos sistema, reklamos pirkimo statymai surenkami iš atitinkamų reklamuotojų [6] įvyksta aukcionas ir reklamuotojas sumoka (patiria išlaidas) už laimėtą parodymą [7]. Kadangi reklamuotojai neprivalo kiekvienam vartotojui pateikti atskirų pasiūlymų, reklamuotojas vietoj to pateikia abstrakčią informaciją apie savo biudžetą ir tikslią auditoriją, o platforma siūlo pasiūlymus jų vardu [8].

Internetinė rinkodara yra verslo įmonės komunikavimas su esamais arba būsimais klientais internete. Pačius klientus, pirkėjus ir pardavėjus galima suprasti įvairiai, dėl to pirmiausia apibrėžiamos pagrindinės sąvokos:

- Paklausa ir internetinės reklamos pirkimas [11, 20] (angl. *Demand Side*) – internete yra perkama vieta internetiniame puslapyje, norint pateikti savo reklamą – reklaminį skydelį – (angl. *banner*). Pirkėjai – reklamos agentūros (pvz.: Lietuvoje esančios „Iprospect“, „Omd“, „SPARK LT“ ir t. t.) arba konkretus prekinis ženklas (angl. *Advertiser, Brand*), pavyzdžiui, „Coca-Cola“. Reklamos agentūros atstovauja savo klientų prekinis ženklus, tačiau jie gali ir patys save atstovauti, įprastai tai telekomunikacijų kompanijos, oro linijos ir pan. Internetinės reklamos pirkimo platforma (angl. *Demand Side Platform*) trumpinama DSP, leidžia pirkėjams prisijungti ir pirkti reklamą iš kelių tinklapių grupių, paprasčiau organizuoti ir tvarkyti savo reklamos pirkimo kampanijas ir gauti duomenis bei ataskaitas apie jų progresą. DSP pavyzdžiai: „Adform“, „DoubleClick Bid Manager“ („Google“ DSP) ir kt.
- Pasiūlos pusė ir internetinės reklamos pardavimas (angl. *Supply side* arba *Publisher*) [12, 20]. Pardavėjai parduoda vietą savo interneto puslapiuose arba savo mobiliosiose programėlėse. Pardavėjai – internetinio puslapio leidėjas (angl. *Publisher*) (pvz.: didžiausi Lietuvoje-15min.lt ar lrytas.lt).

Internetinės reklamos mainai vyksta (angl. *Ad Exchange*) reklamos biržose, kurios leidžia pasiūlos pusei susijungti į vieną tinklą ir parduoti savo kanalais, kurie yra susiintegravę su įvairiomis reklamos pirkimo platformomis. Jie vadinami inventoriais (angl. *Inventory Source*), vienas didžiausių ir populiariausių – „DoubleClick“ – „Google“ įmonės inventorius [12].

## 1.2. Realus laiko aukcionas

Internetinė rinkodara vykdoma naudojant įvairius kanalus ir yra vienas iš greičiausiai besivystančių – realaus laiko aukcionas (angl. *real time bidding*), trumpinama RTB, dar literatūroje galima rasti terminą – automatinis pirkimas (angl. *programmatic*). [12]

Realaus laiko internetinės reklamos pirkimo ir padavimo metodas buvo pradėtas taikyti 2009 metais [6]. Vis dažniau jis yra pasirenkamas vietoje klasikinio internetinės reklamos pirkimo metodo, vadinamo vaizdinės reklamos pirkimu (angl. *Display advertising*). Šis internetinės reklamos pirkimas paremtas principu: perkama konkreti vieta konkrečiame internetiniame puslapyje. Perkama arba dienomis, savaitėmis, mėnesiais, arba sutariamas transakcijų skaičius – milijonas parodymų, šimtas tūkstančių paspaudimų ir pan. Naudojant šį reklamos pirkimo būdą, tarp kitų silpnybių didžiausia yra ta, kad nėra galimybės kontroliuoti ar pasirinkti norimos auditorijos. RTB metodas padeda spręsti vieną didžiausių internetinės reklamos pirkimo iššūkių – parinkti tikslinę auditoriją.

### 1.2.1. Realus laiko aukciono tipai

Dažniausiai naudojamas antros kainos, arba „Vickrey“ aukciono tipas [13]. Tai reiškia, jog tas pirkėjas, kuris siūlė didžiausią kainą- laimi, tačiau mokėti turi tik antrą didžiausią siūlytą kainą. Kainų siūlymai – konfidencialūs, trūkumas - sunku įvertinti tikrąją rinkos kainą, privalumas – apsaugoma reklamos pirkėjo (reklamuotojo) strategija.

Pirmos kainos aukcionas – vykdomas labai retai, laimi tas, kuris siūlė didžiausią kainą ir moka tą kainą, kurią pasiūlė paskutinę.

Antros kainos aukciono pavyzdys:

- X pirkėjas siūlo 0.01 EUR;
- Y pirkėjas siūlo 0.02 EUR;
- Z pirkėjas siūlo 0.025 EUR.

Gauname, jog pirkėjas Z laimi aukcioną ir moka 0.02 EUR už reklamos skydelio parodymą.

### 1.2.2. Realus laiko aukciono kainodaros modeliai

Internetinės reklamos pirkimo platforma turi įvairių būtų apmokestinti pirkėjus, tačiau būtina pabrėžti, kad leidėjams visada mokama už parodymą, o tiksliau CPM kainą. CPM (angl. *cost per mile*) yra kaina už tūkstantį parodymų:

$$CPM = \frac{Išlaidos}{Parodymai \cdot 1,000} \quad (1)$$

Vidutinė kampanijos parodymų kaina, skaičiuojama:

$$eCPM = \frac{\sum Išlaidos}{\sum Parodymai} \cdot 1,000 \quad (2)$$

Pateikiami keli kiti populiariausi pirkėjų kainodaros modeliai:

- CPC (ang. *Cost per Click*) – modelio principas yra optimizuoti reklamos pirkimą reklamos skydelių paspaudimo tikimybei įvykti (angl. *Clicks*). Taip pat kaip CPM modeliu, perkami yra parodymai, tačiau pirkimo kaina nustatoma didesnė ir efektyvesnė, kadangi optimizuojama paspaudimo tikimybės pagrindu. Algoritmo tikslas yra registruoti kuo daugiau vartotojų paspaudimų ant reklaminio skydelio, kartu sumokant kuo mažesnę kainą [49].

$$CPC = \frac{\Sigma I\dot{s}laidos}{\Sigma Paspaudimai} \quad (3)$$

- CPA (angl. *Cost per action*), arba CPL (angl. *cost per lead*)– šis modelis veikia tik tada, kai turima pakankamai surinktų kliento kampanijos duomenų, taip pat pirkėjas turi būti integravęs savo puslapyje sekimo taškus (angl. *tracking pixels*), kurių pagalba yra renkama informacija apie puslapio lankytojus bei jų elgseną puslapyje. Vartotojui paspaudus ant reklamos skydelio (arba tiesiogiai atidarius kliento puslapį, t.y. naršyklėje suvedus adresą), jis nukeliauja į reklamuojamo produkto puslapį, kuriame registruojama tiek techninė, tiek elgsenos informacija apie vartotoją. Jei kampanijos tikslas yra pritraukti kuo daugiau vartotojų, kurie užpildytų kliento registracijos formą, tai CPA algoritmas optimizuoja parodymų pirkimus tų vartotojų realaus laiko aukcione, kurie labiau linkę ne tik paspausti ant reklamos skydelio, bet ir nueiti į konkretų puslapį.

$$CPA = \frac{\Sigma I\dot{s}laidos}{\Sigma Konversijos} \quad (4)$$

- Konversija (angl. *conversion*) nusako kampanijos tikslą, šis gali būti įvairus – vartotojų apsilankymai reklamuojamo prekinio ženklo puslapyje po reklamos pamatymo, tame puslapyje įsigyta prekinio ženklo produkcija, tai taip pat gali būti registracija ar vartotojo nusipirkta prekė ir panašiai.
- CPMV (angl. *Cost per mile viewable*) – modelis, kuris optimizuoja pirkimus tik matomiems parodymams. „Matomumo“ kriterijai gali būti nustatomi reklamos pirkimo platformoje, arba naudojami standartiniai – parodymas buvo matomas, jei reklaminis skydelis buvo matomas ekrane bent 50 % ir bent 1.

$$CPMV = \frac{\Sigma I\dot{s}laidos}{\Sigma MatomiParodymai} \cdot 1,000 \quad (5)$$

Tam, kad veiktų šie ir keli sudėtingesni realaus laiko aukciono reklamos pirkimo algoritmai, kasdien per sekundės dalį gavus siūlymą pirkti reklamos vietą tinklapyje, internetinės reklamos pirkimo platforma turi įvertinti apie 100 milijardų paspaudimo, pirkimo, matomumo, auditorijos ir kitokių parametrų tikimybių. [12]

### 1.3. Internetinės reklamos technologijų įmonės pelno modeliai

Įmonės pagrindinė veikla – teikti programinės įrangos paslaugas, kurios leistų klientams ir internetinių svetainių savininkams bendradarbiauti, siekiant teisingai ir nešališkai fiksuoti reklaminių kampanijų eigą. Pagrindinį įmonės pelną sudaro mokestis už reklaminių skydelių saugojimą ir jų parodymų fiksavimą, kokybės analizę. Dažniausiai mokestis nustatomas kaip dalis registruotų transakcijų, pvz.:

- naudojant išankstinio susitarimo kampanijas (*TPAS*), apmokestinami yra parodymai;
- naudojant realaus laiko aukciono kampanijas (*DSP*) – apmokestinamos kliento išlaidos;
- papildomai apmokestinami *sunkieji* reklaminiai skydeliai, dar vadinami „turtingo turinio“, reklama (angl. *Rich Media Banner*).

Atsiskaitymas priklauso nuo pačios sutarties, kliento tipo ir naudojamų kompanijos produktų. Didžiausią pelną įmonei generuoja du pagrindiniai produktai – išankstinio susitarimo tarp internetinių svetainių ir reklamuotojų kampanijos (*TPAS*) ir realaus laiko aukcionuose (*RTB*) dalyvavusios kampanijos. Abu produktai yra apmokestinami naudojant skirtingus matavimo vienetų.

Globalūs klientai yra apmokestinami atsižvelgiant į tai, kuris pardavimų biuras pasirašė su jais bendradarbiavimo sutartį.

Išankstinio susitarimo kampanijų grynojo pelno prognozė einamajam mėnesiui pagrįsta kasdieniu tokių kampanijų registruotų parodymų kiekiu:

$$Pelnas = \sum(Parodymai) * Pelningumas \quad (6)$$

Realaus laiko aukciono kampanijų grynojo pelno prognozė einamajam mėnesiui pagrįsta kasdieniu tokių kampanijų registruotų išlaidų kiekiu:

$$Pelnas = \sum(Išlaidos) * Pelningumas \quad (7)$$

Pelno skaičiavimą sudaro du pagrindiniai komponentai: transakcijų kiekis, bei pelningumas. Abu šie parametrai yra kintantys laike, nepastovūs.

#### 1.4. Pelno prognozių tyrimai

Nenuspėjamas, dinamiškas rinkos pobūdis ir didelių duomenų prieinamumas suteikia finansų analitikams dideles galimybes, tačiau kartu tai tampa ir dideliu iššūkiu. Problemų tik daugėja, kai kalbama apie realaus laiko prekybos įmones, kur duomenys paprastai juda dideliu greičiu ir dideliais kiekiais.

Laiko eilučių duomenų prognozavimas yra svarbus ekonomikos, verslo ir finansų aspektas. Tradiciškai yra keletas būdų, kurių pagalba galima gana efektyviai prognozuoti sekančią laiko eilutės duomenų reikšmę, pvz., Autoregresija (*AR*), slenkantis vidurkis (*MA*), paprastas eksponentinis glodinimas (*ETS*) ir ypač *ARIMA* (angl. *Autoregressive Integrated Moving Average*), bei daug šio metodo variantų. Konkrečiai, *ARIMA* modelis parodė savo pranašumą tikslumu prognozuoti laiko eilutės ateities reikšmes. Tačiau turint pažangias kompiuterių skaičiavimo galimybes, kuriami pažangesni mašinių mokymosi algoritmai ir metodai, tokie kaip gilusis mokymasis (angl. *Deep Learning*), sukurti nauji algoritmai skirti prognozuoti laiko eilutės būsimas reikšmes. [15, 29]

Šiomis dienomis, finansų analitikai ir duomenų mokslininkai po truputi pradeda atsisakyti tradicinių autoregresinių metodų (*ARIMA*), kadangi darosi sudėtinga tinkamai parinkti parametrus, juos suprasti ir trūksta kompetencijų žinoti, kodėl vienose situacijose šie metodai veikia puikiai, o kitose – rezultatai būna prasti [14]. Vien autoregresinių metodų nebeužtenka, todėl būtina juos kombinuoti su kitais. Vis populiariesni darosi regresinių medžių ar neuroninių tinklų principu veikiantys metodai [16, 17, 18]. Tiesa, pastarieji metodai veikia tiksliau už *ARIMA* metodus tam tikrose situacijose [19]. Tačiau modernūs metodai reikalauja didesnių skaičiavimų resursų, laiko ir specialistų kompetencijų.

Finansų prognozės metu dažnai duomenys yra gaunami iš skirtingų šaltinių ir skirtingu formatu, kurie neturi tiesinių ryšių, todėl reikalingos įvairios duomenų transformacijos. Šioms problemoms spręsti, galima naudoti neuroninių tinklų metodus. Panaudotas heterogeninis daugiasluoksnis generalizuotas operacinis Perceptron (HeMLGOP) algoritmas, kuris laipsniškai mokosi heterogeninę neuronų architektūrą finansų prognozės problemai, su tam tikru disbalansu, spręsti. Metodo principas yra paremtas skirtingų klasių vidutinės kvadratinės paklaidos optimizavimu. Tyrimo metu, lyginant su kitais daugiamačiais regresiniais metodais, šio metodo rezultatai pranoko kitų atžvilgiu [25]. Šis metodas analizuojamos kompanijos trumpalaikiai pelno prognozei galėtų būti pritaikytas, tačiau dėl reikalingų itin didelių skaičiavimų resursų ir ilgo skaičiavimo laiko yra per brangus.

Finansų akcijų rinkos kainų prognozavimui buvo pritaikytas *sentiment* analizės metodas. Duomenys realiu laiku kaupiami naudojant su finansų rinka susijusias socialinių tinklų žinutes, komentarus po straipsniais, pranešančiais įvairias finansines žinias ir pan. Buvo tiriami du aspektai – vartotojų bendras nusiteikimas (teigiamas/neigiamas) ir emocijos (džiaugsmas, liūdesys ir pan). Panaudojus *Grangerio* priešastingumo metodą, įvertinta tokių komentarų įtaka/reikšmė akcijų rinkos kainos pokyčiams, rezultatai parodė, jog žvelgiant bendrai į visą rinką, tai neįtakoja akcijų kainų pokyčių, kadangi kainos pokyčiai yra gana sudėtingi ir dinamiški procesai. Tačiau rasta keletas laikotarpių ir akcijų eilučių, kurioms tiesioginis priešastingumas buvo stebėtinai, todėl nuspręsta, jog tokio metodo analizę galima būtų naudoti tik analizuojant kiekvieną atvejį atskirai [27]. Šis metodas pelno prognozei sudėtingai galėtų būti pritaikomas, tačiau pelno lygtis tiesiogiai priklauso nuo kliento išlaidų internetinės reklamos kampanijoms, todėl būtų įdomu patyrinėti vartotojų tikimybę paspausti ant reklaminio skydelio priklausomai nuo jo nusiteikimo/emocijos. Kuo didesnė ir tikslesnė tikimybė, jog vartotojas paspaus ant reklaminio skydelio, tuo kaina už parodymą galėtų būti didesnė. Didesnė kaina įtakotų didesnes klientų išlaidas, kurios didintų ir tiriamos įmonės pelną.

### **1.5. Matematinų modelių parametrų neapibrėžtumo ir jautrumo rezultatui tyrimai**

Matematinų modelių taikymas tampa dažnesnis ir naudojamas vis platesnėse srityse, ne tik versle ar moksliniuose tyrimuose, bet ir meteorologijoje (siekiant įvertinti klimato prognozės parametrų neapibrėžtumą ir galimą poveikį galutiniam modelio rezultatui), farmacijoje (tiriant tam tikros epidemijos atsiradimo prognozės modelio parametrų neapibrėžtumą), energetikoje ar technologijose [49, 23, 21, 24]. Toks populiarėjantis metodų naudojimas iškelia ir naujų uždavinių. Vienas tokių yra modelio rezultatų neapibrėžtumo tyrimas ir vertinimas. Pavyzdžiui farmacijoje parametrų neapibrėžtumo tyrimas naudojant *Bayes`o* statistinį modeliavimą, atliktas skaičiuojant epidemijos atsiradimo prognozes [22]. Tai reikalinga tam, jog yra situacijų, kai nėra tiksliai žinomi pradiniai parametrai, todėl bandoma atsakyti į klausimą, kokią įtaką tai turi galutinio modelio rezultatui.

Neapibrėžtumo analizė atliekama dviem būdais:

- kiekybiškai, skaičiuojant kiekybiškai galutinių rezultatų neapibrėžtumo dydį (ir, jei reikia, svarbiausiuose analizės etapuose);
- kokybiškai, apibūdinant ir (arba) grupuojant pagrindinius analizei būdingus neapibrėžtumus, nustatant jiems poveikio sistemai lygį [32].

Paprastai kiekybinė analizė laikoma patikimesnė, tačiau turi nemažai trūkumų. Pagrindiniai trūkumai yra tokie, jog ne visi neapibrėžtumo šaltiniai gali būti kiekybiškai įvertinti bet koku patikimumo

įverčiu, ypač tie, kurie susiję problemos apibrėžimu ar vertinimu pagrįstu sprendimu [32]. Todėl kiekybiniai vertinimai gali neobjektyviai apibūdinti skaičiavimo komponentus. Taip pat didžioji dalis vartotojų nėra susipažinę su kiekybiniam neapibrėžtumui vertinti naudojamais metodais, todėl kyla sunkumų efektyviai ir veiksmingai pristatyti vertinimo rezultatus. Tačiau daugeliu atveju, kokybinės priemonės yra naudojamos atskirai, arba kombinuojant su kiekybine analize.

### **1.5.1. Kiekybinė neapibrėžtumo analizė**

Kiekybinei neapibrėžtumo analizei atlikti sukurta keletas metodų. Skiriasi jie savo kompleksiskumu, tačiau naudojant bet kurį metodą yra galimybė įvertinti neapibrėžtumą kiekviename analizės etape ir parodyti neapibrėžtumo sklaidą visoje analizės eigoje. Taikomi metodai: jautrumo analizė [26], *Bayes`o* statistinis modeliavimas, Teiloro eilutės aproksimacija [30] ir *Monte Carlo* metodas.

#### **Jautrumo analizė**

Jautrumo analizė yra esminė priemonė visose modeliavimo (taikymo) srityse. Jos tikslas yra kiekybiškai įvertinti ir palyginti neapibrėžtų parametrų (ar kitų įvesties kintamųjų) įtaką konkretaus modelio rezultatams. Finansų analitikams tai yra vienas svarbiausių įrankių. Šis metodas leidžia lengvai pakeitus pradinis lygties parametrus greitai pamatyti, kaip pasikeis analizuojamo klausimo rezultatas. Jautrumo analizei atlikti yra daug skirtingų metodų (sklaidos diagramos vertinimas, koreliacijos koeficientų vertinimas, standartizuotos tiesinės regresijos metodas), tačiau jie paprastai apsiriboja vienu išvesties kintamuoju. Pagrindinis trūkumas jautrumo analizės, jog ji tampa ypatingai sudėtinga, jei neapibrėžtumas atsiranda tarp didelio kiekio kintamųjų vertinimo. Jautrumo parametrus įvertinti yra sukurta keletas įrankių, vienas tokių – R paketas *multisensi*.

#### **Teiloro eilutės aproksimacija**

Tai yra matematinis metodas, skirtas apytiksliai įvertinti neapibrėžtumo pasiskirstymą (mąstą) visame procese. Metodas lengvai pritaikomas dideliame duomenų kiekiui, skaičiavimo laikas ir kaštai taip pat maži.

#### **Bajeso statistinis modeliavimas**

Visi aukščiau aprašyti metodai taikomi deterministiniams modeliams. Didžioji dalis šių modelių parametrų buvo apskaičiuoti naudojant duomenis, bent jau iš pradžių modeliuose jie yra vertinami kaip žinomi ir fiksuoti objektai. Stochastiniai modeliai apskaičiuoja parametrus iš duomenų ir viename etape priderina modelį, tokiu būdu tiesiogiai įtraukdami neapibrėžtumą parametrų įvertiniuose dėl duomenų trūkumų. *Bayes`o* modeliavimas tęsiasi dar vienu žingsniu ir apima papildomą neapibrėžtį kitų šaltinių parametruose (išreikštų tikimybės pasiskirstymu) [33].

#### **Monte Carlo metodas**

Skaičiavimo algoritmas, imitacinio modeliavimo metodika paremta neapibrėžtumo ir jautrumo analizė yra paremta skirtingais modelio įvertinimais su pasirinktais tikėtiniais modelio įvesties parametrais. Gautų rezultatai yra naudojami modelio neapibrėžtumo įverčiui bei parametrus, kurie sukėlė šį neapibrėžtumą nustatyti. Šio metodo pagalba galima pakeisti brangiai kainuojančius bandymus kompiuterinėmis simuliacijomis, taip sutaupydami laiko tyrimams ir galėdami įvertinti galimus scenarijus.



*Monte Carlo* metodas apima du metodus – jautrumo analizę ir scenarijų analizę, pagrįstą tikimybių teorija. Naudojantis šiuo metodu gaunami imitacinio modeliavimo rezultatų pasiskirstymai [40].

### 1.5.2. Kokybinė neapibrėžtumo analizė

Kokybiniai neapibrėžtumo analizės metodai savo pobūdžiu yra mažiau formalūs nei kiekybiniai metodai. Viena vertus, tai yra didelis trūkumas, nes tai reiškia, jog ne visada lengva palyginti skirtingus tyrimus ar analizes. Kita vertus, šie metodai tampa daug lankstesni ir pritaikomi prie skirtingoms aplinkybėms. Todėl kokybiniai metodai gali būti rengiami pagal poreikį, kad juos būtų galima naudoti beveik bet kokiame analizės etape, beveik bet kokiame kontekste vertinant beveik bet kokį neapibrėžtumo aspektą.

Norint objektyviai ir tiksliai perteikti tokio tipo tyrimo rezultatus, jie turi atitikti tam tikrus kriterijus:

- Šie metodai turi būti pagrįsti aiškia neapibrėžtumo koncepcija ir aiškiais kriterijais;
- Jie turi būti atkuriami, t.y. turi gražinti tuos pačius rezultatus naudojant tuos pačius duomenis, esant tokioms pačioms sąlygoms;
- Turi būti lengvai interpretuojami, t.y. naudojami visoms pusėms suprantami terminai;
- Turi koncentruotis tik į neapibrėžtumo šaltinius, kurie turi įtakos galutiniam rezultatui ir gali turėti įtakos priimant sprendimus.

Keletas metodų yra naudojama tokio tipo neapibrėžtumo šaltinių analizei. Vienas gana naudingas būdas yra taikyti paprastus santykinius neapibrėžtumo rodiklius išreikštus „pasitikėjimo laipsniu“ [35]. Šis matavimo vienetas naudojamas apibūdinti neapibrėžtumo šaltinį, kuris paremtas ekspertų nuomone dėl modelio tikslumo, analizės ar išvadų. Idealiu atveju, „pasitikėjimo laipsnio“ apibrėžimai turėtų būti sudaromi remiantis ne vieno eksperto nuomone, bet naudojant tam sudarytą specialią ekspertų grupę, arba analizuojant turimus reprezentatyvius duomenis, kurie arba yra įtraukti į modelio sudarymą, arba jais remiantis buvo suformuluotos pradinės prielaidos.

Pasitikėjimo laipsnis	Pasitikėjimo laipsnio įvertis
Labai aukšto pasitikėjimo laipsnis	Bent jau 9 iš 10 tikimybė įvykti
Aukštas pasitikėjimo laipsnis	Apie 8 iš 10 tikimybė
Vidutinis pasitikėjimo laipsnis	Apie 5 iš 10 tikimybė
Žemas pasitikėjimo laipsnis	Apie 2 iš 10 tikimybė
Labai žemas pasitikėjimo laipsnis	Mažiau nei 1 iš 10 tikimybė

1 lentelė. Pasitikėjimo laipsnių kategorijų vertinimo kriterijai

Rezultatų pateikimas naudojant įverčių lenteles gali būti naudojamas pateikiant rezultatus tekstiniu formatu, tačiau pristatant gyvai, diskusijos metu, patogesnis yra grafinis pateikimas, pasitelkiant pasikliautinus intervalus.

Nagrinėjami pagrindiniai trys neapibrėžtumo šaltiniai:

- **Fizinis kintamumas.** Šis neapibrėžtumas dar vadinamas minimaliu arba neišvengiamu, kuris atsiranda dėl įvairių fizinių procesų ar kintamųjų natūralaus atsitiktinio kintamumo dėl daugelio priežasčių, pvz. aplinkos kintamumo, eksploatacijos variacijų, įvairių aplinkos procesų ir kiekybės kontroles. Toks neapibrėžtumas yra tiek sistemos savybėse (pvz. kaip greit įmonė gali reaguoti į rinkos pokyčius, sistemos veikimo techninės galimybės/stabilumas ir panašiai), tiek išorės įtaka sistemai (metų laikai, ekonominė situacija). Todėl matematiniais modeliais grįstoje prognozėje atsiranda neapibrėžtumo šaltiniai, nes modeliui sukurti naudojamos netikslios reikšmės, dėl ko vėliau modelio rezultatai tampa netikslūs. Tokie procesai inžinerinėje analizėje yra vadinami atsitiktiniais kintamaisiais, kuriems yra priskiriami statistiniai parametrai tokie kaip vidurkis, standartinis nuokrypis ir duomenų pasiskirstymo intervalai, apskaičiuoti remiantis istoriniais stebėjimais.
- **Duomenų neapibrėžtumas.** Šis neapibrėžtumas atsiranda dėl žinių ar informacijos trūkumo. Duomenų neapibrėžtumas atsiranda įvairiomis formomis, pvz. jei kiekybinis kintamasis laikomas atsitiktiniu, statistinių parametrų pasiskirstymas priklauso nuo turimų duomenų kiekio. Turint mažai duomenų, pasiskirstymo amplitudė tampa neaiški, todėl jie gali būti vertinami kaip atsitiktiniai dydžiai. Kita vertus, turimi duomenys gali būti nekokybiški arba netikslūs, todėl nėra paprasta šį neapibrėžtumą vertinti atsitiktiniais kintamaisiais. Taip pat turima informacija gali būti pateikta intervaline išraiška arba remiantis ekspertų nuomone. Dar vienas svarbus duomenų neapibrėžtumo šaltinis atsiranda dėl netikslų matavimo metodų (skirtingi įrankiai naudoja skirtingus kiekių kintamųjų apvalinimo metodus, skirtingus šaltinius valiutų kursų reikšmėms gauti ir panašiai).
- **Modelio paklaidos.** Šis neapibrėžtumas atsiranda dėl apytikslų matematinų modelių elgsenos skaičiavimų ir skaitinių parametrų aproksimacijos skaičiavimo metu. Dėl to atsiranda dviejų tipų klaidos – metodo parametrų aproksimacijos paklaidos ir modelio skaičiavimų paklaidos. Kompleksinės sistemos veikimo vertinimui naudojama daugybė analizės metodų, kiekvienas iš jų turi savo prielaidas ir aproksimaciją. Kadangi įvairių analizžių paklaidos yra gana sudėtingos, suvesti jas į vieną bendrą metodo paklaidą gana sudėtinga.

Neapibrėžtumo ir jautrumo analizė naudojama įvairiose srityse. Lentelėje pateikiamas sąrašas sričių, kuriose ši metodologija pritaikoma (žr. 2 lentelė). Sritys yra suskirstytos į pagrindines keturias kategorijas: sprendimų priėmimui (arba rekomendacijoms), esamos situacijos komunikacijai, sustiprinti sistemos supratimą ir modelių kūrimui.

1.	Sprendimų priėmimas arba rekomendacijų kūrimas
1.1.	Testavimo būdu ieškoti patikimiausio ir optimaliausio sprendimo
1.2.	Identifikuoti silpnąsias vietas, kurias veikiant, keistųsi optimalios strategijos įgyvendinimas
1.3.	Atrandant jautriausius parametrus, veikiančius įmonės veiklą
1.4.	Testuojant alternatyvius strategijos metodus
1.5.	Lyginant skirtingų sprendimų priėmimo rezultatus
2.	Komunikacijoje

2.1.	Padaryti rekomendacijas labiau patikimas, suprantamas ir rasti tinkamų argumentų
2.2.	Padeda sprendimų priėmėjams daryti prielaidas
3.	Sistemos supratimo stiprinimas
3.1.	Įvertinti ryšį tarp įvesties ir išvesties veiksnių
3.2.	Suprasti įvesties ir išvesties veiksnių poveikį vienas kitam
3.3.	Kurti hipotezes naujų strategijų ar rekomendacijų testavimui
4.	Modelio kūrimas
4.1.	Modelio galiojimo tikrinimas ir tikslumo vertinimas
4.2.	Ieškojimas klaidų modelyje
4.3.	Modelio suprastinimas
4.4.	Modelio tikslinimas
4.5.	Trūkstumų ar neteisingų reikšmių apdorojimas, vertinimas

2 lentelė. Neapibrėžtumo ir jautrumo tyrimų panaudojimo atvejai

Programavimo įrankio „R“ paketas *multisensi* yra sukurtas atlikti jautrumo analizę pagal daugiamačių modelį. Jo pagalba apskaičiuojami ir pateikiami grafiškai jautrumo rodikliai kiekviename išvesties kintamajame arba kintamųjų kombinacijoje, atsirandančių dėl matmenų mažinimo metodų. Pradinė *multisensi* versija buvo pagrįsta faktoriaus principu, paremtu modelio neapibrėžtumo parametrais, pagrindinių komponentų analizės taikymas modelio rezultatui, pirmųjų pagrindinių komponentų jautrumo *anova* analizė. Ši idėja buvo Campbell et al. [26], toliau ją tyrinėjo Lamboni et al. Tai *multisensi* pagrindu sukurta biblioteka, tačiau dabartinė versija apima alternatyvius metodus dimensijų mažinimui: *spline*, *bsplines* ir polinominė regresija. Be to, keli metodai dabar gali būti naudojami jautrumo analizei, įskaitant tuos, kurie įgyvendinti pakete *sensitivity*. [41]

## 2. Metodologija, metodų aprašymas

### 2.1. Pelno prognozės lygtis

Išlaidų vidutinės dienos reikšmė gaunama sudėjus visas einamojo mėnesio jau praėjusias dienas iki vakar ir padalinama iš praėjusių dienų kiekio (einamojo mėnesio).

$$I_{vid(d)} = \frac{1}{d} \sum_{i=1}^d (I_i) \quad (8)$$

Čia:  $I_{vid}$  – Vidutinės dienos išlaidos

$I$  – Dienos išlaidos

$d$  – diena

Kiekvienos dienos išlaidų dieninis vidurkis normalizuojamas, lyginant su paskutinės mėnesio dienos išlaidų vidurkiu.

$$I_{koeff(d)} = \frac{I_{vid(d)}}{I_{vid(max(d))}} \quad (9)$$

Čia:  $I_{koeff(d)}$  – Išlaidų dienos vidurkio koeficientas

$I_{vid}$  – Vidutinės dienos išlaidos

$d$  – diena

$max(d)$  – paskutinė mėnesio diena

Klientų išlaidų prognozė einamojo mėnesio pabaigai apskaičiuojama dalinant praėjusios dienos išlaidų vidurkį iš koeficiento, kuris nurodo, kokią dalį tos dienos vidutinės išlaidos sudaro paskutinės mėnesio dienos vidutinės dienos išlaidas. Gautą rezultatą dauginame iš mėnesio dienų skaičiaus ir turime bendrą klientų mėnesio išlaidų sumą.

$$I_{mèn} = \frac{I_{vid(d)}}{I_{koeff(d)}} \times d_{mèn} \quad (10)$$

Čia:  $I_{mèn}$  – einamojo mėnesio klientų prognozuojamos išlaidos

$d_{mèn}$  – dienų skaičius einamąjį mėnesį

Mėnesio pelningumas, t.y. kokia dalis klientų išlaidų sudarė įmonės pelną skaičiuojamas santykiu nuo išlaidų ir gauto pelno:

$$Pelningumas_{mèn} = \frac{Pelnas_{mèn}}{I_{mèn}} \quad (11)$$

Pelno prognozė einamajam mėnesiui paskaičiuojama sudauginus bendrą išlaidų prognozė iš pelningumo koeficiento:

$$Pelnas_{mèn} = \left( \frac{I_{vid(d)}}{I_{koeff(d)}} \times d_{mèn} \right) \times \frac{Pelnas_{mèn}}{I_{mèn}} \quad (12)$$

## 2.2. Slenkantis vidurkis

Slenkantis vidurkis yra ekstrapoliacijos metodas. Naudojantis šiuo metodu teigiama, kad prognozuojama laiko eilutės reikšmė  $(t+1)$  yra  $n$  kiekio praeities stebėjimų vidurkis.

$$MA_t = \frac{1}{n} \sum y_{t-n+i} = \frac{1}{n} (y_t + y_{t-1} + \dots + y_{t-n+1}) \quad (13)$$

## 2.3. Eksponentinis glodinimas

Slenkamojo vidurkio metodai fiksuoja esamą padėtį, tačiau nepratęsia sekos reikšmių pratęsti į ateitį. Šį prognozavimo uždavinį sprendžia eksponentinio glodinimo variantai.

Eksponentinio glodinimo metodai, glodina laiko eilutes naudojant vidurkio metodą stebėjimams. Suglodinta eilutė gali būti naudojama tiek proceso vaizdavimui (jeigu stebimas procesas su triukšmu), tiek prognozavimui (optimali prognozė yra prognozuojamo atsitiktinio dydžio vidurkis).

Eksponentinio glodinimo būdai:

- Paprastas eksponentinis glodinimas (angl. *simple exponential smoothing*, SES, *single exponential smoothing*) naudojamas laiko eilučių glodinimui, kai eilutės neturi trendo ar sezoniškumo komponentių. Šiam metodui siūloma naudoti tarpinį variantą — kai į prognozės lygtį įtraukiami visi stebėjimai, tačiau paskutiniams suteikiamas didesnis svoris, nei ankstesniems:

$$\hat{Y}_{T+1|T} = \alpha Y_T + \alpha(1 - \alpha)Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots \quad (14)$$

Lygtyje esantis parametras  $\alpha$  tenkina sąlygą  $0 \leq \alpha \leq 1$ . Kadangi svorių reikšmės eksponentiškai mažėja, todėl metodas ir vadinasi eksponentiniu glodinimu. Lygtį galime perrašyti ir kompaktiškesne forma (15):

$$\hat{Y}_{t+1|t} = \alpha Y_t + (1 - \alpha)\hat{Y}_{t|t-1} \quad (15)$$

Dvigubas eksponentinis glodinimas naudojamas tuomet, kai stebima laiko eilutė turi tendenciją. Šis metodas užrašomas viena prognozės (18) ir dviem glodinimo lygtimis (16,17):

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (16)$$

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \quad (17)$$

$$F_{t+m} = S_t + b_t m \quad (18)$$

Čia  $\alpha$  ir  $\beta$  glodinimo koeficientai

$S_t$  – suglodinta  $t$  stebėjimo reikšmė

$b_t$  – trendo suglodinta  $t$  reikšmė

$F_{t+m}$  – prognozės reikšmės  $m$  periodų į priekį

- Trigubas (Holt Winters) glodinimas taikomas tuomet, kai laiko eilutei būdingas trendas ir sezoniniai svyravimai. Sezoniniai svyravimai gali būti tiek adityvūs tiek multiplikatyvūs. Metodas užrašomas tokiomis formulėmis:

$$S_t = \alpha \frac{Y_t}{c_{t-L_t}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (19)$$

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \quad (20)$$

$$c_t = \gamma \frac{Y_t}{S_t} + (1 - \gamma)c_{t-L} \quad (21)$$

$$F_{t+m} = (S_t + b_t m)c_{t+m-L} \quad (22)$$

Čia:  $\alpha$  ir  $\beta$  glodinimo koeficientai

$S_t$  – suglodyta  $t$  stebėjimo reikšmė

$b_t$  – trendo suglodyta  $t$  reikšmė

$c_t$  – sezoniškumo įtakos faktorius

$L$  – periodų kiekis metuose

$F_{t+m}$  – prognozės reikšmės  $m$  periodų į priekį

#### 2.4. Croston`'s metodas

Pirmą kartą J.D. Croston šio metodo idėją publikavo 1972 metais. Straipsnyje jis pristatė naują metodologiją, kuri yra skirta prognozuoti laikinai paklausius produktus. Pagrindinė idėja:

- Įvertinti vidutinį paklausos lygį, kai ta paklausa vis kartojasi;
- Įvertinti vidutinį laiką tarp dviejų paklausos periodų;
- Prognozuoti produkto paklausą kaip paklausos lygį (kai tai vėl įvyksta), padaugintą iš įvykio tikimybės.

Pateikiamas prognozės skaičiavimo pavyzdys naudojant šį metodą:

1. Grafiškai arba statistiškai įvertiname, jog produkto paklausos periodiškumas yra vidutiniškai kas 5 periodus (pvz., tai yra 20 % tikimybė, jog bus produkto paklausos laikotarpis).
2. Taip pat įvertinama, jog esant produkto paklausos laikotarpiui, produkto paklausos lygis yra 10 vienetų.
3. Taigi paklausos prognozė kiekvienam sekančiam periodui:  $10 \times 0.2 = 2$ , kiekvienam periodui.

#### 2.5. ARIMA laiko eilučių prognozavimas

Laiko eilutė – to paties kintamojo reikšmės skirtingais laiko intervalais (dienomis, mėnesiais, ketvirčiais, metais ir t.t.). Pagrindinis laiko eilutės tyrimo tikslas - nustatyti duomenų kaitos dėsningumą (išskiriant sezoniškumą, tendenciją ir liekamąją dalį).

ARIMA (angl. Autoregressive Integrated Moving Average Process) modelių tikslas – prognozuoti nagrinėjamus reiškinius, kintančius laiko tėkmėje. Pagrindinė idėja – prognozės sudaromos panaudojant nagrinėjamo reiškinio pradinių duomenų ir modelio paklaidų pokyčius. ARIMA modelis

laiko eilutę išskaido į autoregresinį procesą (AR), kuris aprašo praeities įvykius, integruotą procesą (I), padedantį stabilizuoti duomenis, ir slenkančio vidurkio (MA) procesą, kuris vertina modelio paklaidų poveikį duomenims. Šis metodas taikomas, kai nagrinėjami nestacionarūs (integruoti) tiesiniai procesai.

Laiko eilutė laikoma stacionari siaurąja prasme, jeigu jos daugiamačiai pasiskirstymai nepriklauso nuo poslinkio laike (23). Integruotas autoregresijos – slenkančio vidurkio modelis ARIMA( $p, d, q$ ), užrašomas kaip ARMA( $p, q$ ) modelis statistinio rodiklio eilutės  $d$ -tosios eilės skirtumams (24). Jei  $d = 1$ , tai ARMA modelis yra užrašomas eilutei, kurios nariai (25):

$$F_{t_1, \dots, t_k}(\cdot) = F_{t_1 + \tau}(\cdot), \forall t_1, \dots, t_k, \tau \in T \quad (23)$$

$$\Delta^d Y_i(t) \quad (24)$$

$$Z_i(t) = \Delta^t Y_i(t) = \Delta^1 Y_i(t) - Y_i(t - 1) \quad (25)$$

Atsitiktinis procesas  $Z_i(t) \in I(d)$  vadinamas ARIMA( $p, d, q$ ) procesu, jei jo  $d$  eilės pokyčiai yra ARMA( $p, q$ ) procesas.

Matematinė ARIMA( $p, d, q$ ) modelio išraiška:

$$\Phi(L)(1 - L)^d Y_i(t) = \theta(L)\varepsilon_i(t) + \mu \quad (26)$$

Lygtyje esantys  $\Phi(L)$  ir  $\theta(L)$ , tai polinamai su laiko postūmio operatoriumi  $L$ .

$$\Phi(L) = (1 - \varphi_1 L - \varphi_2 L^2 \dots - \varphi_p L^p) \quad (27)$$

$$\theta(L) = (1 + \gamma_1 L + \gamma_2 L^2 + \dots + \gamma_q L^q) \quad (28)$$

Tokios lygties sprendinys visada yra nestacionarus procesas  $Y_i(t)$ , tačiau jį visada galima pakeisti d kartų diferencijuotu procesu.

## 2.6. Pelno prognozės sėkmingumo (tikslumo) vertinimas

- RMSE (angl. *root mean square error*) - Vidutinės kvadratinės paklaidos (RMSE) (arba kartais vidurkio kvadrato paklaida) yra dažniausiai naudojamos modelio arba paskaičiuotų prognozuojamų verčių ir nustatytų verčių skirtumams rasti. Šis matavimo metodas nurodo kokis yra skirtumas tarp prognozuotos reikšmės ir žinomos reikšmės arba kvadratinis vidurkis šių skirtumų.

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}} \quad (29)$$

- MAPE (angl. *mean absolute percentage error*) - vidutinė absoliutinė procentinė paklaida (arba vidutinis absoliutus procentinis nuokrypis) yra naudojama statistikoje kaip prognozės metodų tikslumo rodiklis, pavyzdžiui, įvertinti tendenciją arba naudojama kaip nuostolių funkcija regresijos problemoms spręsti. Šis rodiklis nusako santykinę prognozavimo tikslumą, juo galima palyginti skirtingų parametrų prognozes (prognozavimas labai tikslus, kai MAPE < 10 ir tikslus, kai 10 < MAPE < 20). Ji apskaičiuojama:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \quad (30)$$

Čia:  $y_t$  – faktiškos rodiklio reikšmės;  
 $\hat{y}_t$  – prognozuotos rodiklio reikšmės;  
 $n$  – prognozuojamų rodiklių kiekis.

- MASE (angl. *mean absolute scale error*) – vidutinė absoliuti skalės paklaida (MASE) yra prognozių tikslumo rodiklis. Tai yra vidutinė absoliuti prognozuojamų verčių paklaida, padalyta iš vidutinės absoliutinės paklaidos vieno etapo prognozės.

Nesezoninėms laiko eilutėms vidutinė absoliuti skalės paklaida skaičiuojama:

$$MASE = \frac{1}{T} \sum_{t=1}^T \left( \frac{|e_t|}{\frac{1}{T-1} \sum_{t=2}^T |Y_t - Y_{t-1}|} \right) = \frac{\sum_{t=1}^T |e_t|}{\frac{T}{T-1} \sum_{t=2}^T |Y_t - Y_{t-1}|} \quad (31)$$

$e_t$  – prognozės paklaida analizuojamu laiko momentu, skaičiuojama iš tikrosios reikšmės atėmus prognozuojamą;

$Y_t$  – faktiška rodiklio reikšmė.

Sezoninėms laiko eilutėms vidutinė absoliuti skalės paklaida skaičiuojama:

$$MASE = \frac{1}{T} \sum_{t=1}^T \left( \frac{|e_t|}{\frac{1}{T-m} \sum_{t=m+1}^T |Y_t - Y_{t-m}|} \right) = \frac{\sum_{t=1}^T |e_t|}{\frac{T}{T-m} \sum_{t=m+1}^T |Y_t - Y_{t-m}|} \quad (32)$$

Pagrindinis skirtumas su prieš tai aprašytu metodu, skirtu laiko eilutėms be sezono komponentės yra tas, jog vardiklyje paklaidos yra skaičiuojamos ne su ankstesne reikšme, bet su reikšme, esančia prieš sezoną  $m$ .

- R kvadratas (determinacijos koeficientas). Tai svarbiausia modelio tikimo duomenims charakteristika, kuri privaloma visuose regresijos modelių aprašymuose. Determinacijos koeficientas lygina skirtumus tarp prognozuojamų reikšmių, kai atsižvelgiama į regresijos modelį, su skirtumais tarp prognozuojamų reikšmių, kai į modelį neatsižvelgiama. Parodo, kuri vieno požymio bendro kitimo dalis gali būti paaiškinta kito požymio reikšmių kitimu.

Determinacijos koeficientas įgyja reikšmes intervale  $[0, 1]$ . Kuo koeficiento reikšmė didesnė, tuo modelis geriau tinka duomenims. Blogai, kai determinacijos koeficiento reikšmė yra mažesnė, nei 0,2, reiškia modelis nepaaiškina net 20 % turimų duomenų. Tačiau, jei determinacijos koeficientas viršija 0,8, tai modelis aprašo bent 80 % duomenų kas reiškia, jog skaičiuoti prognozes yra tikslinga. Tiesa, tai dar neužtikrina, kad visi kintamieji jame būtini, o pats modelis yra prasmingas.

$$R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2} \quad (33)$$

$S_\varepsilon^2$  – liekanų ( $\varepsilon$ ) dispersija;

$S_y^2$  – stebimų kintamojo ( $Y$ ) reikšmių dispersija.



## 2.7. Neapibrėžtumo tyrimai

### Standartinis nuokrypis (arba vidutinis kvadratinis nuokrypis)

Dydis, nusakantis atsitiktinio dydžio įgyjamų reikšmių sklaidą apie vidurkį (įprasta žymėti  $s$ , SD arba  $\sigma$ ). Standartinio nuokrypio dimensija yra lygi atsitiktinio dydžio dimensijai. Dėl šios priežasties šis dydis labiau tinkamas naudoti praktikoje nei dispersija, kurios matas lygus atsitiktinio dydžio dimensijos kvadratui.

Jeigu  $X$  yra atsitiktinis dydis, kurio vidurkis  $\mu$ , tai:

$$E[X] = \mu \quad (34)$$

### Dispersija

Statistinė imties charakteristika, atspindinti labiausiai tikėtiną eilinio matavimo vertės nukrypimą nuo aritmetinio vidurkio.

$$D(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (35)$$

### Koreliacijos koeficientai

Koreliacija (arba koreliacijos koeficientas) tikimybių teorijoje ir statistikoje yra statistinis ryšys tarp kintamųjų.

Koreliacijos koeficientas – koreliacijos stiprumo matas. Jeigu dviejų kintamųjų koreliacijos koeficientas lygus nuliui, tai tie kintamieji yra statistiškai nepriklausomi.

Koreliacijos koeficientų yra keletas. Žinomiausi yra:

$r_{xy}$  – Pirsono (*Pearson*) tiesinės koreliacijos koeficientas

$\rho_{xy}$  – Spirmeno (*Spearman*) ranginės koreliacijos koeficientas

Dviejų atsitiktinių dydžių  $X$  ir  $Y$ , kurių vidurkiai yra  $\mu_X$  ir taip:

$$\rho_{xy} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (36)$$

Koreliacijos koeficientų reikšmių skalės vertinimas:

Labai stiprus	Stiprus	Vidutinis	Silpnas	Labai silpnas	Nėra ryšio	Labai silpnas	Silpnas	Vidutinis	Stiprus	Labai stiprus
-1	nuo -1 iki -0,7	nuo -0,7 iki -0,5	nuo -0,5 iki -0,2	nuo -0,2 iki 0	0	nuo 0 iki 0,2	nuo 0,2 iki 0,5	nuo 0,5 iki 0,7	nuo 0,7 iki 1	1

3 lentelė. Koreliacijos koeficientų reikšmių skalė

### 3. Tiriamoji dalis

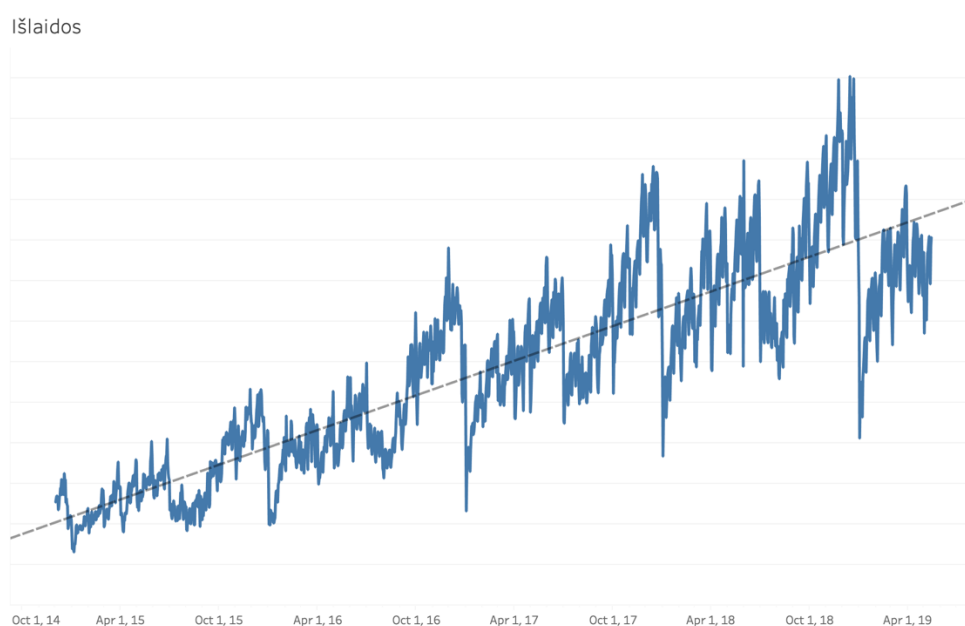
Duomenų analizei atlikti buvo naudojami du skirtingi duomenų rinkiniai. Pirmajame duomenų rinkinyje buvo naudojami vidiniai internetinės reklamos technologijų įmonės klientų transakcijų duomenys. Kadangi duomenys yra konfidencialūs ir ne tiesiogiai atskleidžiantys įmonės finansinę situaciją, tyrimui naudoti globalūs, agreguoti, dieniniai duomenys, parinktas tik pavyzdinis duomenų rinkinys. Analizuojamų duomenys rinkinys apima duomenis sukauptus nuo 2001 metų gruodžio pradžios iki 2018 metų pabaigos. Antrąjį duomenų rinkinį sudaro įmonės gauto pelno mėnesiniai duomenys, paskirstyti pagal produktą, sukaupti nuo 2015 metų pradžios iki 2018 metų pabaigos.

Pradiniame pirmajame duomenų rinkinyje iš viso buvo 161463 eilutės ir 4 stulpeliai. Duomenų rinkinį sudaro įmonės klientų transakcijų istoriniai duomenys. Antrajame – iš viso 7011 eilutės ir 4 stulpeliai. Abeji duomenų rinkiniai gauti iš įmonės vidinių duomenų šaltinių. Siekiant užtikrinti įmonės finansinių duomenų konfidencialumą, tiek pelno reikšmės, tiek transakcijų kiekis buvo anonimizuojamas naudojant santykinį daugiklį.

Vienas tiriamasis produktas pradėtas vystyti nuo kompanijos įsikūrimo pradžios (2001 metų pabaigos), todėl transakcijų registruota nuo periodo pradžios, tačiau antrasis produktas paleistas į rinką tik 2014 metų pabaigoje, todėl istorinių duomenų turima gerokai mažiau.

#### 3.1. Aprašomoji statistinė duomenų analizė

Atlikus tiriamos internetinės reklamos technologijų įmonės klientų registruotų išlaidų realaus laiko aukciono kampanijoms priklausomybę nuo laiko buvo pastebėtas bendras išlaidų augimas per visą laikotarpį (1 pav.). Taip pat matomas ryškus sezoninis išlaidų pokytis ties kiekvienų metų pabaiga ir naujų metų pradžia. Tokia tendencija yra paaiškinama šventiniu laikotarpiu suaktyvėjusia reklamos paklausa, o po naujų metų sumažėję pardavimai mažina ir reklamos kiekį.

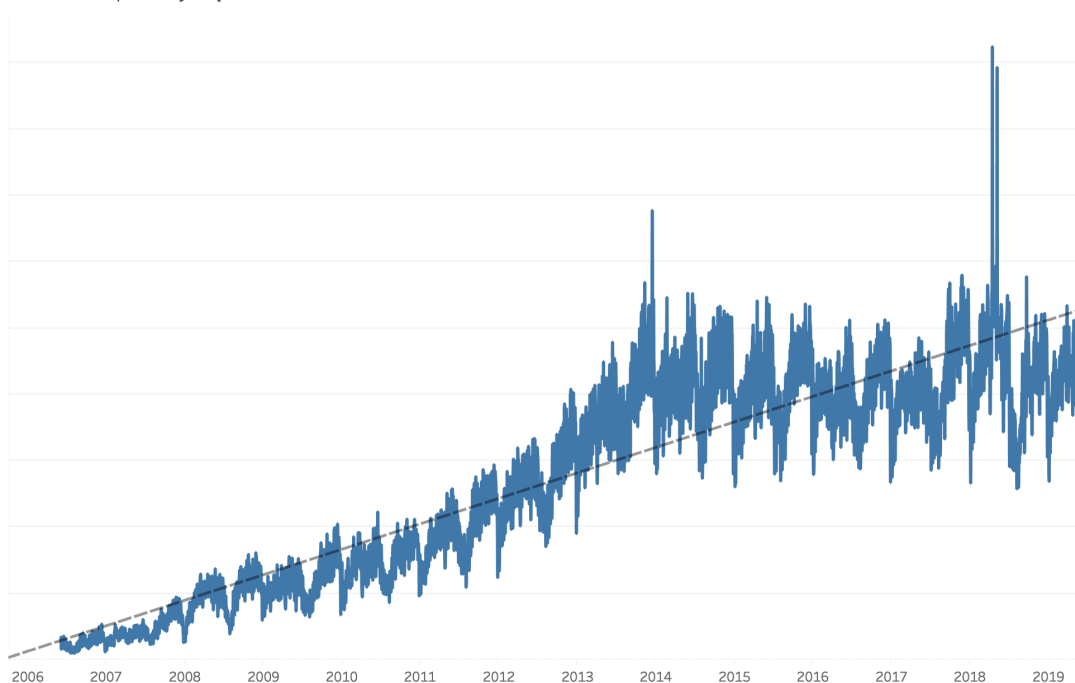


1 pav. Įmonės klientų išlaidos realaus laiko aukcione dalyvaujančiose kampanijose

Atliekant antrojo produkto žvalgomąją analizę pastebėta, jog registruotų parodymų kiekis tendencingai augo nuo 2006 metų vidurio iki 2014 metų pradžios. Vėliau registruotų transakcijų

kiekio augimas sustojo, tačiau visu periodu matomas ryškus sezoninis svyravimas (2 pav.). Neišvengta ir ryškių išskirčių, kurios yra paaiškinamos naujo kliento prijungimu ir testavimo laikotarpiu generuojamu dideliu kiekiu transakcijų.

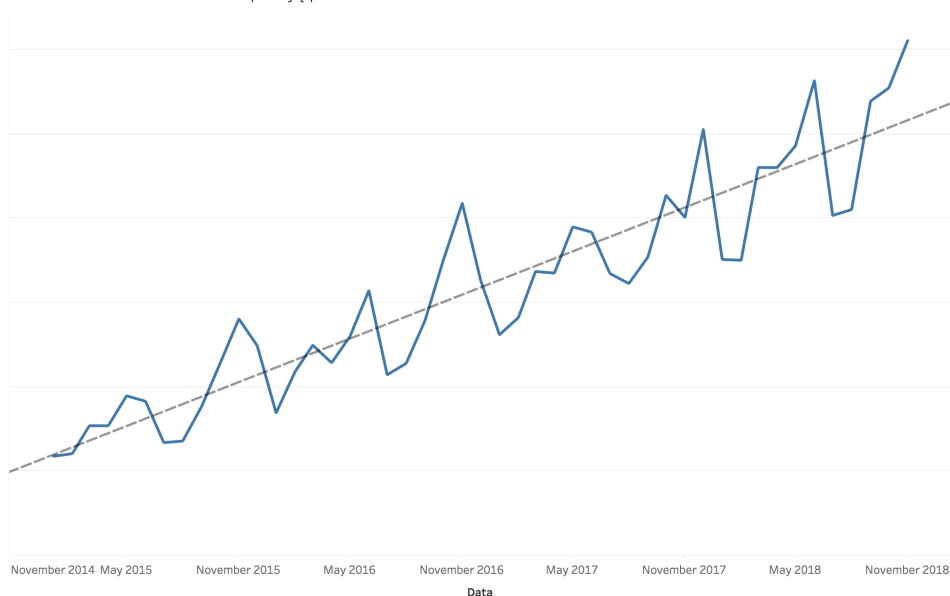
Reklamos parodymų kiekis



2 pav. Registruotų parodymų kiekis klientų, dalyvaujančių išankstinio susitarimo kampanijose

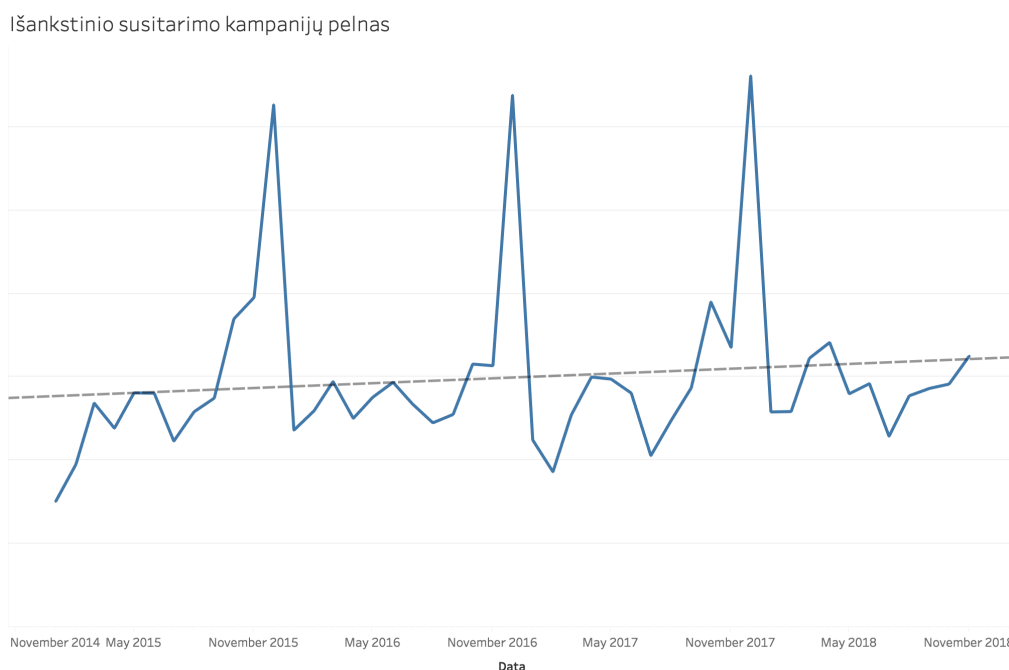
Tiriant įmonės pelno tendencijas išlaidas realaus laiko aukcionuose dalyvaujančių kampanijų, matomas taip pat auganti kryptis (3 pav. ir 4 pav.). Kadangi duomenys yra agreguoti mėnesio lygiu, išskiriamas metinis sezoniškumas. Kaip ir su transakcijų kiekiu, pelnas didžiausias reikšmes įgyja metų pabaigoje, krenta – metų pradžioje, bei metų viduryje.

Realaus laiko aukciono kampanijų pelnas



3 pav. Pelnas, gautas už realaus laiko aukciono kampanijas

Analizuojant pelną, gautą už išankstinio susitarimo kampanijas, matoma gana pastovi tendencija, lėtas pelno augimas. Tačiau kaip ir ankstesnio produkto pelno apžvalgoje, pastebėtas sezoninis pelno šuolis metų pabaigoje ir kritimas metų pradžioje.



4 pav. Pelnas, gautas už išankstinio susitarimo kampanijas

Akivaizdu, jog kylat transakcijų kiekiui, turėtų didėti ir pelnas, tačiau pelno skalė per visą laikotarpį padidėjo nežymiai, kai tuo tarpu transakcijų kiekis išaugo beveik 20 %. Pelno prognozės ir neapibrėžtumo tyrimą atliksime tik realaus laiko aukciono kampanijoms. Siekiant skaičiuoti pelno prognozę, už išankstinio susitarimo kampanijas, reikėtų ištraukti bent jau turėti informaciją apie transakcijų kiekį, kuris nebuvo apmokestintas. Šių duomenų kompanija nepateikė.

### 3.2. Realaus laiko aukciono pelno prognozė

Einamojo mėnesio pelno prognozė priklauso nuo transakcijų kiekio ir pelningumo rodiklio:

$$Pelnas = Išlaidos \times Pelningumas \quad (37)$$

Abu lygties kintamieji nėra pastovūs ir iš anksto žinomi, todėl juos reikia ne tik prognozuoti, bet ir įvertinti prognozės lygties parametrų neapibrėžtumą.

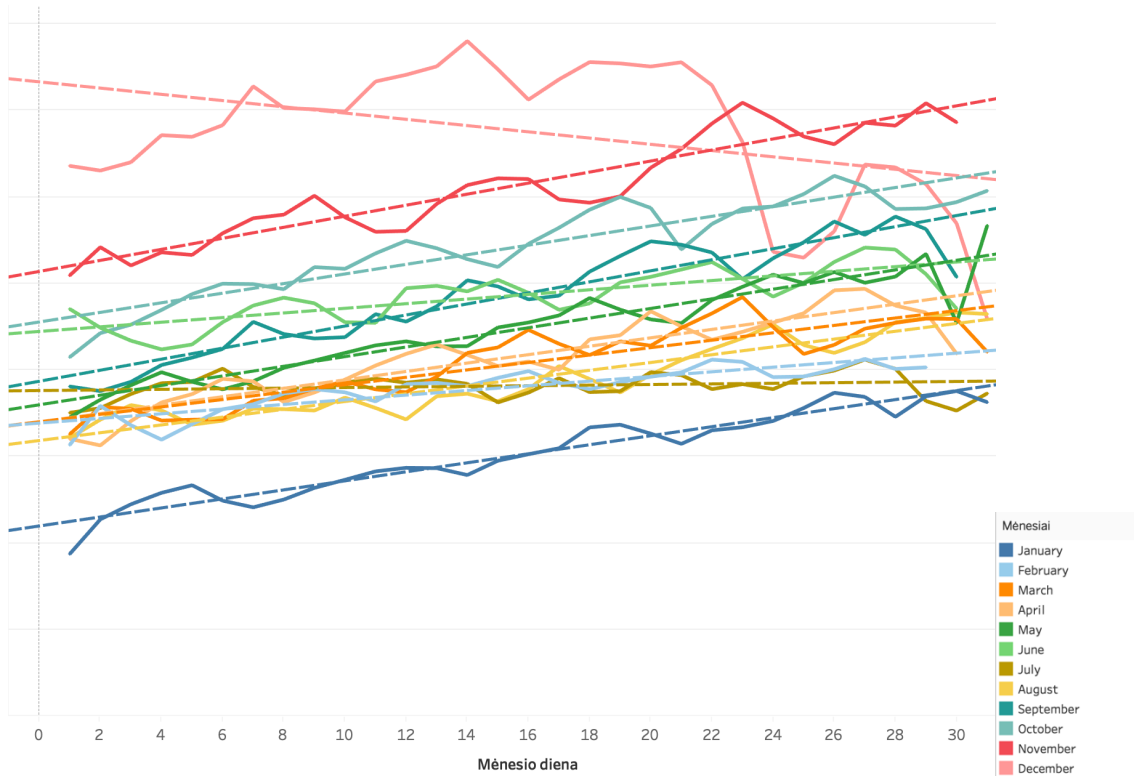
#### 3.2.1. Išlaidų prognozė

Šiame tyrime analizuojama pelno prognozė einamajam mėnesiui, kuri turi būti perskaičiuota ir atnaujinta realiu laiku, jog specialistai galėtų įvertinti pelno situaciją suplanuotų biudžetų atžvilgiu ir intis atitinkamų veiksmų.

Tiriant išlaidų kaitą mėnesio eigoje buvo pastebėta, jog tendencija yra labai panaši. Mėnesio pradžioje vidutinės dienos išlaidos būna mažesnės, o artėjant mėnesio pabaigai – didėja. Paskutinė mėnesio diena įprastai būna turinti didžiausias vidutinės dienos išlaidas per mėnesį.

Tokia mėnesio tendencija yra paaiškinama įprastu verslo principu. Mėnesio pradžioje klientai paleidžia naujas kampanijas, kurios dažniausiai trunka iki mėnesio ir su laiku jas keičia (optimizuoja), siekdami išleisti nustatytą kampanijos biudžetą optimaliai. Artėjant mėnesio pabaigai, likęs mėnesio biudžetas leidžiamas greičiau, kadangi kampanijos jau būna optimizuotos taip, jog pasiektų tikslią auditoriją.

Mėnesio vidutinės dienos išlaidos

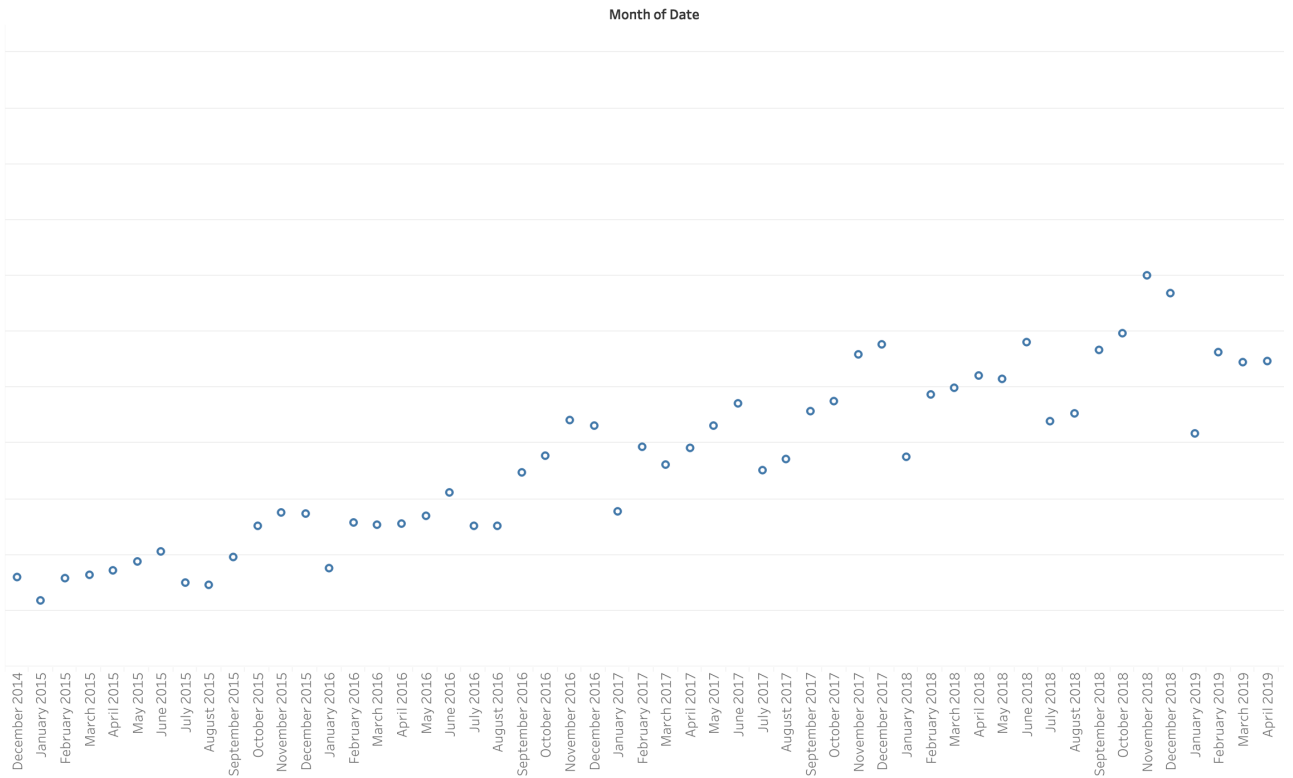


5 pav. Dieninis klientų išlaidų pokytis mėnesio eigoje

Bendra penkių metų mėnesių kasdienė išlaidų tendencija yra panaši – mėnesio eigoje didėjanti. Tačiau yra vienas mėnuo metuose (gruodis), kurio bendra išlaidų kryptis kitokia lyginant su kitais mėnesiais. Nors išlaidos gruodžio mėnesį būna didžiausios, bet jos sparčiai auga iki mėnesio 20 dienos, vėliau staigiai mažėja.

Didėjant faktinėms išlaidoms kiekvieną dieną, didėja ir vidutinės dienos išlaidos.

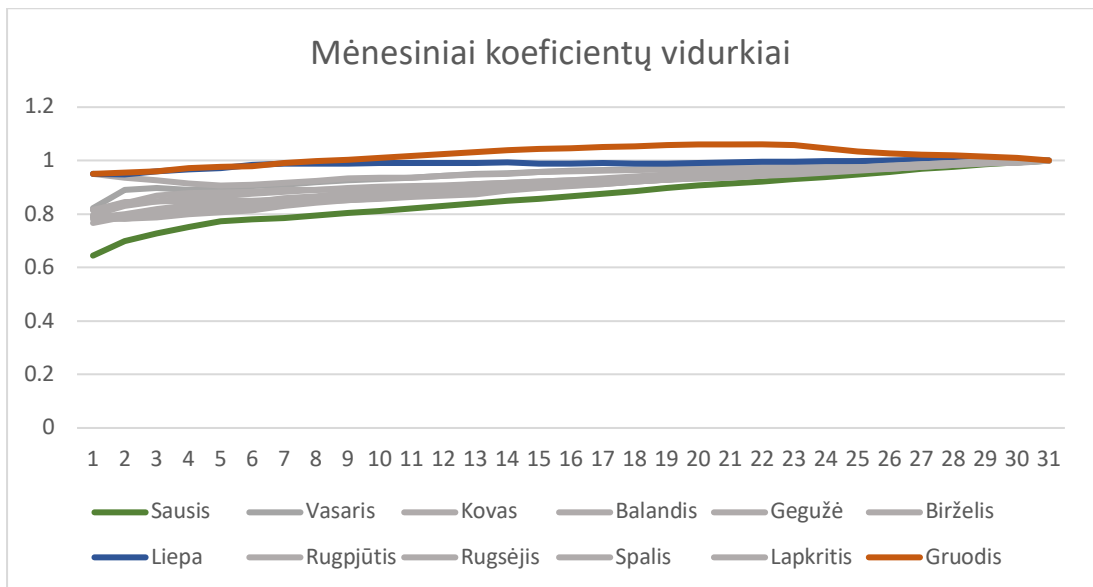
## Vidutinės dienos klientų išlaidos



6 pav. Klientų vidutinės dienos išlaidos

Kadangi kiekvieno mėnesio vidutinių dieninių išlaidų skalė gana stipriai skiriasi, skaičiuoti prognozę remiantis tik istoriniais faktiniais duomenimis nekorektiška. Siekiant prognozuoti registruotas kliento išlaidas mėnesio pabaigai, mėnesio dienos vidurkis normalizuojamas lyginant su paskutinės mėnesio dienos vidurkiu, siekiant rasti bendrą kiekvienam mėnesiui pritaikomą tendencijos kreivę.

Siekiant prognozuoti einamojo mėnesio klientų išlaidas, toliau analizuojamas kasdieninis vidutinių dieninių išlaidų pokytis remiantis paskaičiuotais koeficientais. Ankstesniuose grafikuose buvo išskirti keli mėnesiai, kurių metu kasdinių išlaidų tendencija buvo kitokia, lyginant su kitais mėnesiais. Ši hipotezė patikrinama ir naudojant paskaičiuotus koeficientus.

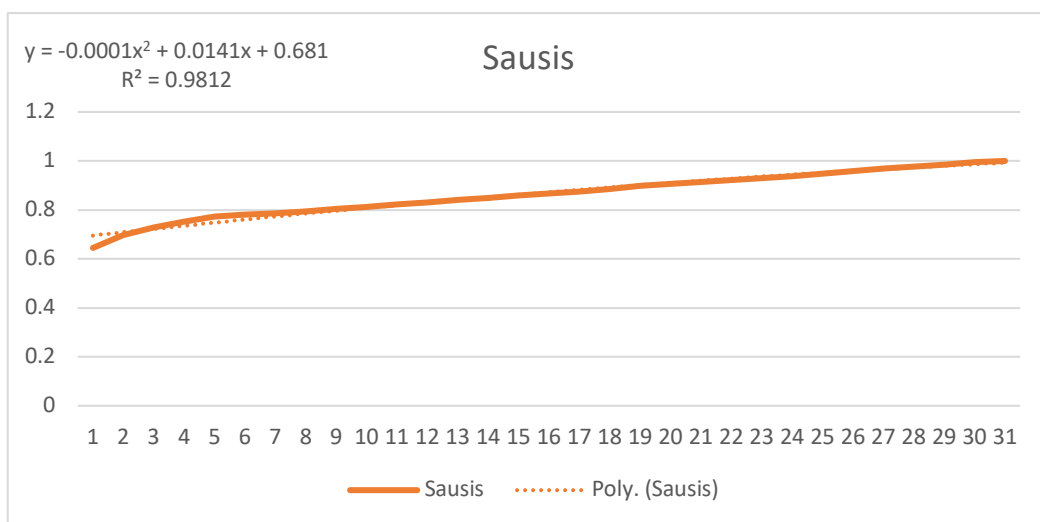


7 pav. Klientų vidutinių dienos išlaidų koeficientų mėnesiniai vidurkiai

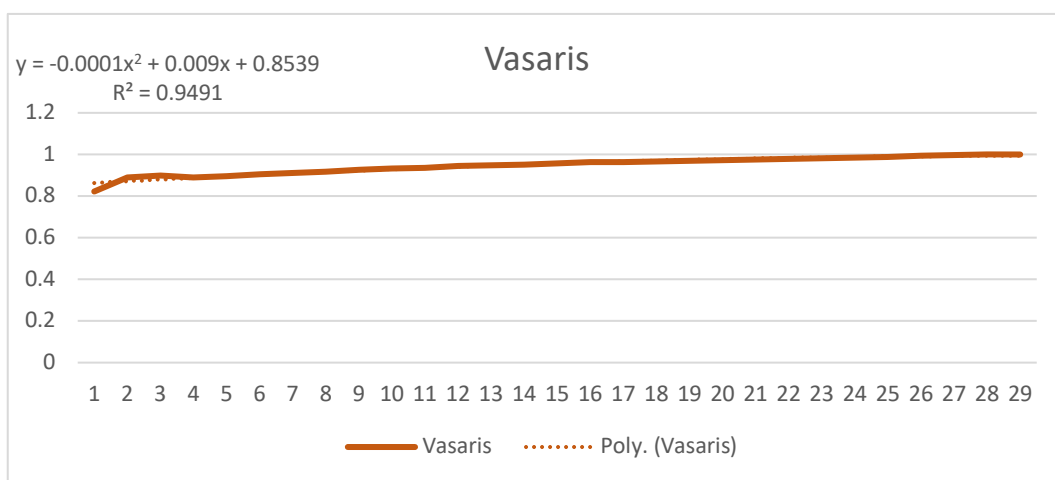
Vidutinių dieninių išlaidų pokytis mėnesio eigoje yra panašus visų mėnesių, tačiau išskiriami trys mėnesiai metuose (gruodis, sausis ir liepa), kuriems bendrą pokyčio formulę naudoti būtų nekorektiška. Todėl išlaidų prognozei skaičiuoti naudojamos penkios prognozės lygtys, atitinkamai kuriam mėnesiui prognozuojamos išlaidos (išskiriami mėnesiai su 30 ir 31 diena).

Prognozės pokyčio lygtis apskaičiuojama remiantis istoriniais koeficientų duomenimis, pritaikant polinominę, antro laipsnio lygtį. Lygties tinkamumas vertinamas remiantis  $r$  kvadrato statistika. Atitinkamai skirtingiems mėnesiams, skaičiuojamos keturios lygtys:

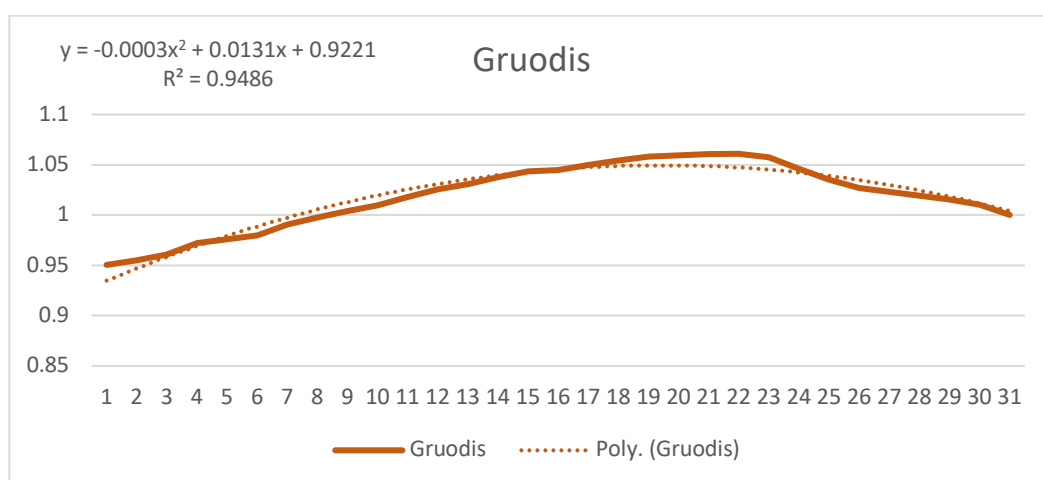
- Sausio mėnesio lygtis (8 pav.). Randama remiantis istoriniais sausio mėnesių koeficientų duomenimis;
- Vasario mėnesio lygtis (9 pav.). Randama remiantis istoriniais vasario mėnesių koeficientų duomenimis, taip pat išskiriami mėnesiai, turintys 29 dienas;
- Gruodžio mėnesio lygtis (10 pav.). Randama remiantis istoriniais gruodžio mėnesių koeficientų duomenimi;
- 30 dienų mėnesių lygtis (11 pav.). Randama remiantis keturių mėnesių koeficientų istoriniais vidurkiais – balandžio, birželio, rugsėjo ir lapkričio;
- 31 dienos mėnesių lygtis (12 pav.). Randama remiantis penkių mėnesių koeficientų istoriniais vidurkiais – kovo, gegužės, liepos, rugpjūčio ir spalio.



8 pav. Sausio mėnesio koeficientų kitimo lygtis

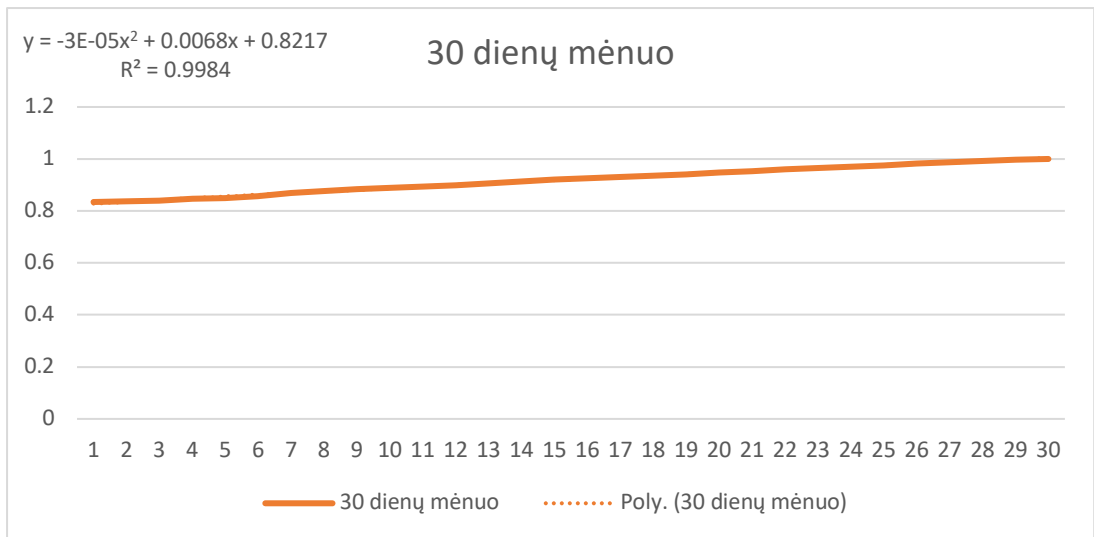


9 pav. Vasario mėnesio koeficientų kitimo lygtis

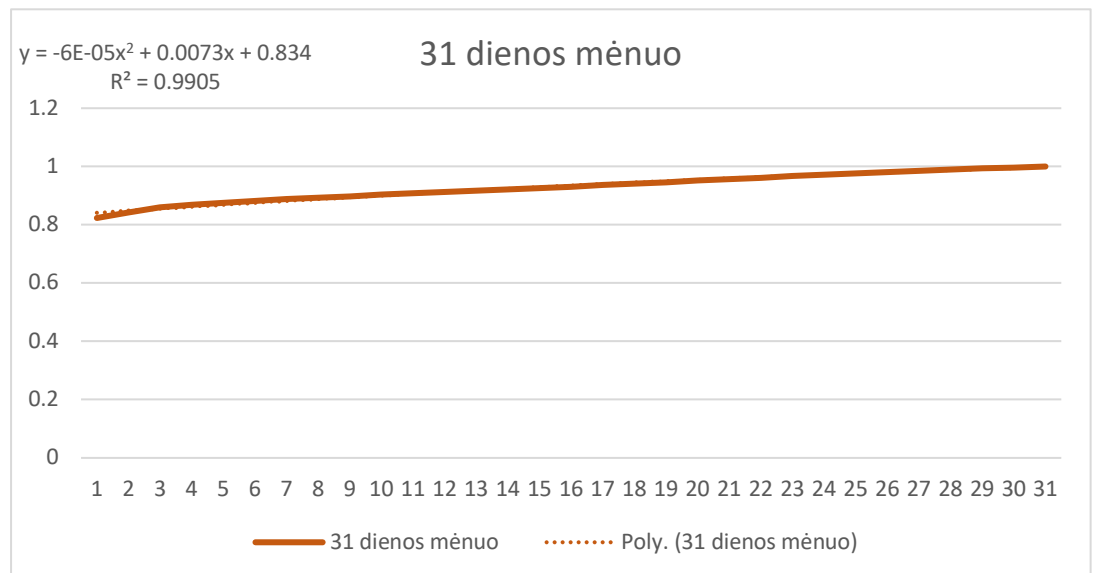


10 pav. Gruodžio mėnesio koeficientų kitimo lygtis





11 pav. 30 dienų mėnesių koeficientų kitimo lygtis



12 pav. 31 dienos mėnesių koeficientų kitimo lygtis

Modelių tinkamumas vertinamas remiantis determinacijos koeficientu, rezultatai pateikti lentelėje (žr. 4 lentelė). Gautas klientų mėnesio išlaidų prognozių lygtys:

- Sausio mėnuo:

$$I_{m\acute{e}n} = \frac{I_{vid(d-1)}}{-0.0001(d-1)^2 + 0.0141(d-1) + 0.681} \times 31 \quad (38)$$

- Vasario mėnuo:

$$I_{m\acute{e}n} = \frac{I_{vid(d-1)}}{-0.0001(d-1)^2 + 0.009(d-1) + 0.8539} \times 28 \quad (39)$$

- Gruodžio mėnuo:

$$I_{m\acute{e}n} = \frac{I_{vid(d-1)}}{-0.0003(d-1)^2 + 0.0131(d-1) + 0.9221} \times 31 \quad (40)$$

- 30 dienų mėnuo:

$$I_{m\acute{e}n} = \frac{I_{vid(d-1)}}{-0.00003(d-1)^2 + 0.0068(d-1) + 0.8217} \times 30 \quad (41)$$

- 31 dienos mėnuo:

$$I_{m\acute{e}n} = \frac{I_{vid(d-1)}}{-0.00006(d-1)^2 + 0.0073(d-1) + 0.834} \times 31 \quad (42)$$

Modelio pavadinimas	Modelio lygtis	R <sup>2</sup>
Sausio mėnuo	$-0.0001(x)^2 + 0.0141(x) + 0.681$	0.9812
Vasario mėnuo	$-0.0001(x)^2 + 0.009(x) + 0.8539$	0.9491
Gruodžio mėnuo	$-0.0003(x)^2 + 0.0131(x) + 0.9221$	0.9486
30 dienų mėnesiai	$-0.00003(x)^2 + 0.0068(x) + 0.8217$	0.9984
31 dienos mėnesiai	$-0.00006(x)^2 + 0.0073(x) + 0.834$	0.9905

4 lentelė. Klientų išlaidų dieninių vidutinių prognozių koeficientų modelių lygtys

Didžiausią reikšmę (arčiausiai 1) vertinant determinacijos koeficientą, įgijo mėnesinė lygtis, skirta prognozuoti mėnesius, turinčius 30 dienų. Modelis paaiškina beveik visas reikšmes. Išlaidų prognozės reikšmės turėtų būti gana tikslios. Mažiausia determinacijos koeficiento reikšmė stebima gruodžio mėnesio prognozės lygtyje, bet siekia beveik 0.95, o tai yra gana aukštas rodiklio rezultatas.

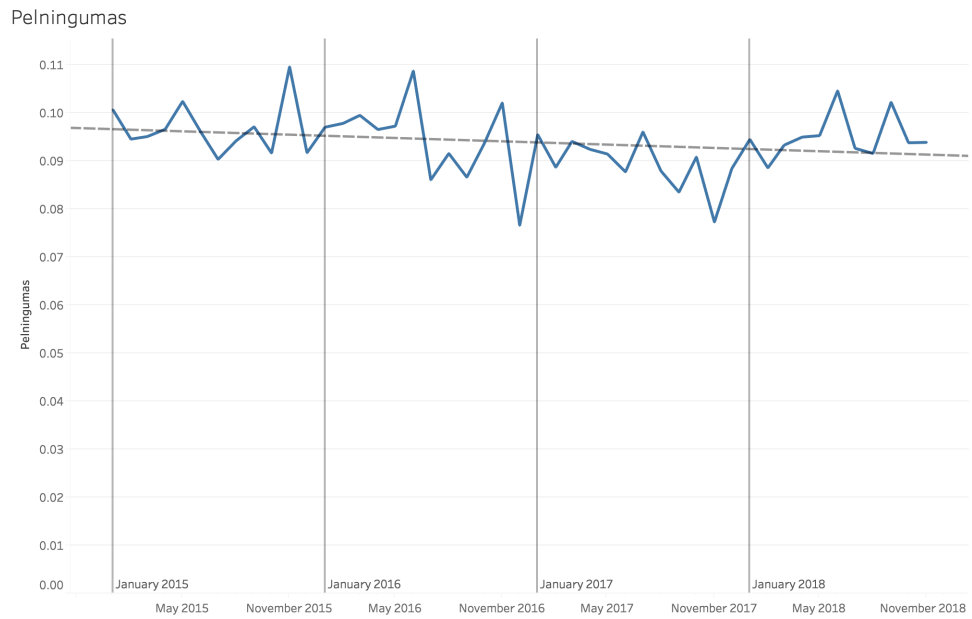
Tačiau nereikia pamiršti, jog rinka yra pakankamai dinamiška, veikiama įvairių išorinių faktorių, todėl tikėtis, jog galutinė išlaidų prognozė bus tiksli beveik 100 %, nereikėtų.

### 3.2.2. Pelningumo prognozė

Pelningumo duomenų rinkinį sudaro mėnesiniai duomenys laikotarpiu [2015-01-01 – 2018-11-01], be trūkstamų reikšmių. Pateikiamos pagrindiniai statistiniai įvertiniai:

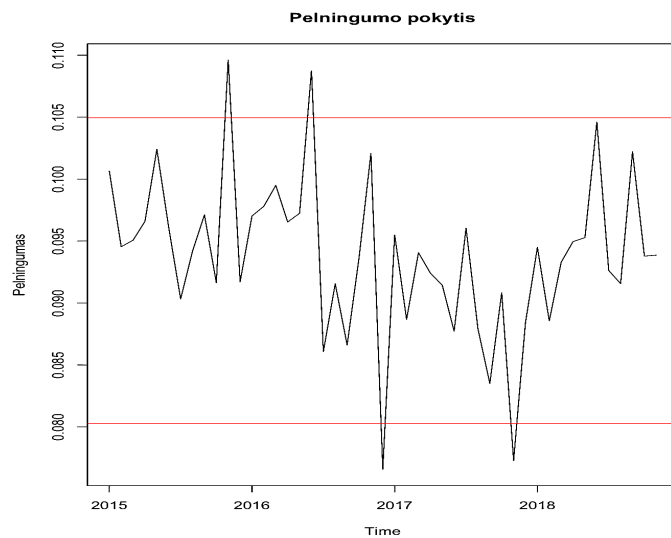
Minimali reikšmė	1 kvartilis	Mediana	Vidurkis	3 kvartilis	Maksimali reikšmė
0.07658	0.09113	0.09407	0.09388	0.09681	0.10960

Pelningumo rodiklis visu laikotarpiu kinta nuo 0.07658 iki 0.10960, nors pokyčio intervalas nėra didelis, tačiau atsižvelgiant į tai, jog šis rodiklis galutinėje pelno formulėje yra kaip daugiklis, esant didelei pinigų sumai, net tik vienos šimtosios dalies pokytis yra reikšmingas. Taip pat pastebėta, jog skirtumas tarp pirmo kvartilio ir maksimalios reikšmės (0.01847) yra panašus kaip tarp minimalios reikšmės ir pirmojo kvartilio (0.01455). Mediana yra artima vidurkiui, kas reiškia, jog dar negalima daryti išvadų apie dideles išskirtis, kadangi vidurkis jautrus išskirtims, o mediana – ne.



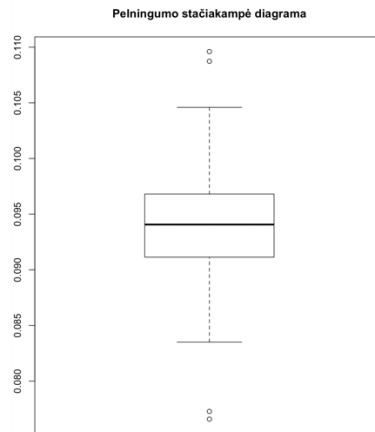
13 pav. Pelningumo kaita

Bendra pelningumo tendencija kintanti šiek tiek mažėjančia kryptimi. Tačiau grafiškai neišskiriama sezoninė kaita. Toliau nagrinėjami pelningumo rodikliai atliekant laiko eilutės analizę.



14 pav. Pelningumo pokytis su pažymėtais pirmu ir trečiu kvartiliais

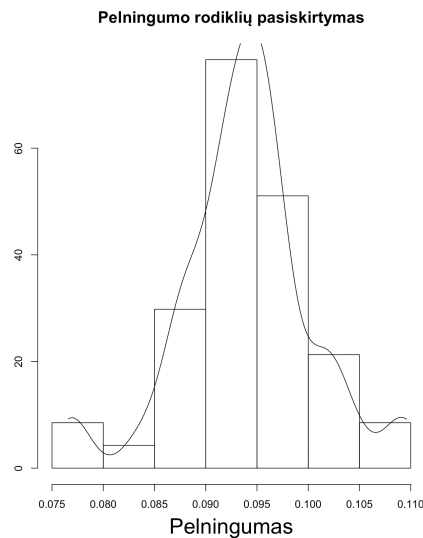
Raudonomis linijomis pažymėtas apatinis ir trečias viršutinis kvartiliai, taip pažymint ribą, kurią viršijus, matomos išskirtys. Pateiktame grafike matomi 4 taškai, kurie viršijo pažymėtus kvartilius. Patvirtinimui duomenys pavaizduoti pasitelkus *boxplot* funkciją:



15 pav. Pelningumo išskirtys

Matomi keturi žymėjimai, kurie žymi viršutines (teigiamas) ir apatines (neigiamas) išskirtis. Taip pat vizualiai galima teigti, jog skirstinys gana simetriškas, kadangi pirmas ir trečias kvartiliai vienodai nutolę nuo medianos. Tai tik patvirtino prieš tai pateikto grafiko rezultata.

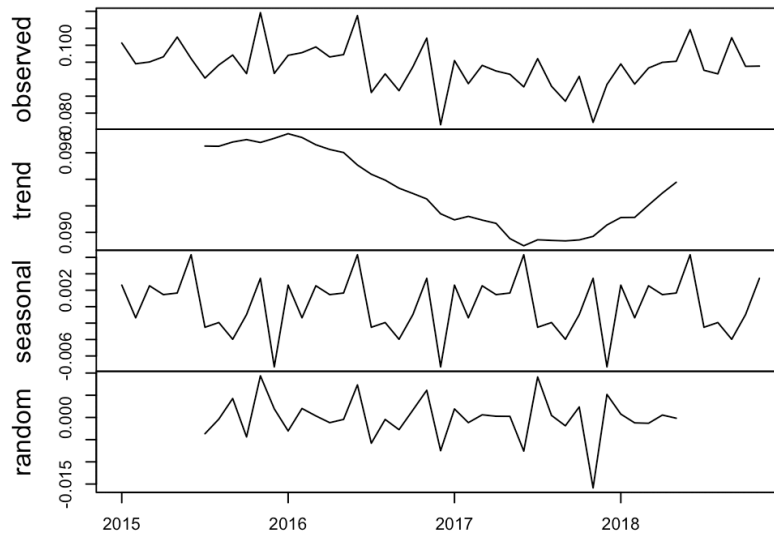
Toliau pateikiamas bendras mėnesinis pelningumo pasiskirstymas duotame laikotarpyje histograma:



16 pav. Pelningumo rodiklių tankio histograma

Didžiausias pelningumo rodiklių tankis (pasiskirstymas) yra intervale, į kurį patenka ir vidurkis, taip pat likusios reikšmės pasiskirsčiusios “varpo” forma aplink vidurkį, kas reiškia, jog vizualiai vertinant, galima teigti, jog skirstinys yra normalusis. Todėl didžiausia tikimybė, jog pelningumo rodiklis pateks į vidurkio intervalą.

Pelningumo reikšmių laiko eilutė išskaidoma į pagrindines komponentes naudojant adityvų metodą, nes vertinant pelningumo pokytį, neižvelgiama pokyčio intervalo didėjimas laikui bėgant.



17 pav. Laiko eilutės išskaidymas pagrindinėmis komponentėmis

Trendo komponentės kitimo sritis yra labai maža, palyginus su likusiomis (kinta nuo 0.090 iki 0.096, lyginant su (-0.008:0.006) sezoniškumas ir (-0.015:0.01) likutis), taip pat sezono periodiškumas matomas metinis (12 mėnesių).

Atlikus *Box-Pierce* testą laiko eilutės likutinėms reikšmėms ir patikrinus hipotezę, ar jos yra nepriklausomos viena nuo kitos, jos neatmetėme (18 pav. *p-value* > 0.05), todėl galima teigti (su 95% tikimybe), kad šios likutinės reikšmės nėra priklausomos nuo laiko vėlavimų (nėra autokoreliacijos).

#### Box-Pierce test

```
data: profitability_reminder
X-squared = 0.122, df = 1, p-value = 0.7269
```

18 pav. Laiko eilutės likučių autokoreliacijos testo rezultatas

Siekiant rasti tinkamiausią modelį paskaičiuoti pelningumo prognozę, testuojami penki metodai, jų gauti rezultatai pateikiami lentelėje.

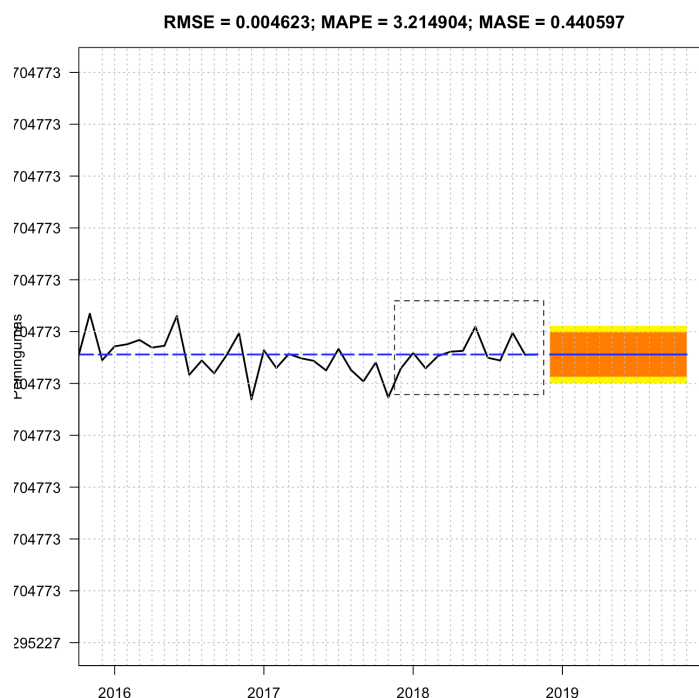
Nr.	Metodo pavadinimas	Prognozės įvertinys	RMSE	MASE	MAPE
1.	Mean(Vidurkis)	0.09387781	0.004623	0.440597	3.214904
2.	ETS(M,N,N)	0.0942196	0.007778	0.891995	6.465474
3.	Croston`s	0.09364631	0.006204	0.679294	4.911142
4.	ARIMA (1,1,1)	0.09564053	0.008941	1.087465	7.931347
5.	Slenkantis vidurkis (n=5)	0.094809748	0.006568		

5 lentelė. Pelningumo prognozių metodų rezultatai

## Vidurkis

Pelningumo prognozė, naudojant vidurkio metodą yra atliekama remiantis tik istoriniais pelningumo rodikliais.

$$\widehat{Pelningumas} = \frac{1}{n} (Pelningumas_t + Pelningumas_{t-1} + \dots + Pelningumas_{t-n+1}) \quad (43)$$



19 pav. Pelningumo prognozė vidurkio metodu

Grafiškai pateikti istoriniai pelningumo rodikliai ir prognozė ateinantiems 12 mėnesių. 80 % pasikliautino intervalo lygis pažymėtas oranžine spalva, geltona – 90 % pasikliautino intervalo lygis. Naudojant šį metodą, prognozė yra statiška, t.y. visos būsimos reikšmės yra vienodos, ilgalaikėms prognozėms jį naudoti būtų tikslinga, jeigu laiko eilučių tyrimo metodai nėra tikslūs ar negalimi. Šiuo atveju, vidutinė kvadratinė paklaida yra mažiausia lyginant su kitais metodais, tačiau reikalinga tik viena prognozavimo reikšmė, todėl metodas galėtų būti naudojamas.

### Slenkantis vidurkis

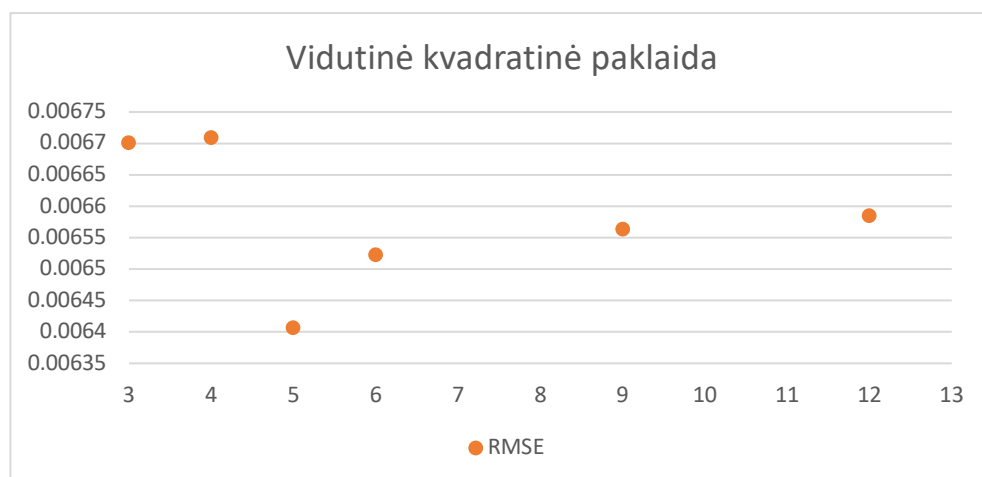
Slenkančio vidurkio metodų skaičiuojama remiantis paskutinių kelių buvusių taškų reikšmėmis sekančiam taškui prognozuoti. Atlikus keletą eilučių slenkančio vidurkio prognozės tyrimą (žr. 2 lentelė) ir paskaičiavus vidutines kvadratinės paklaidas kiekvieno metodo prognozuotoms reikšmėms, geriausi rezultatai (mažiausios paklaidų reikšmės) gauti skaičiuojant vidurkį naudojant paskutinių penkių mėnesių pelningumo reikšmes.

$$P_{m\acute{e}n} = \frac{1}{5} \sum_{i=1}^{m-5} (P_i) \quad (44)$$

Čia:  $P_{m\acute{e}n}$  – Pelningumo rodiklis einamąjį mėnesį;

$m$  – mėnuo;

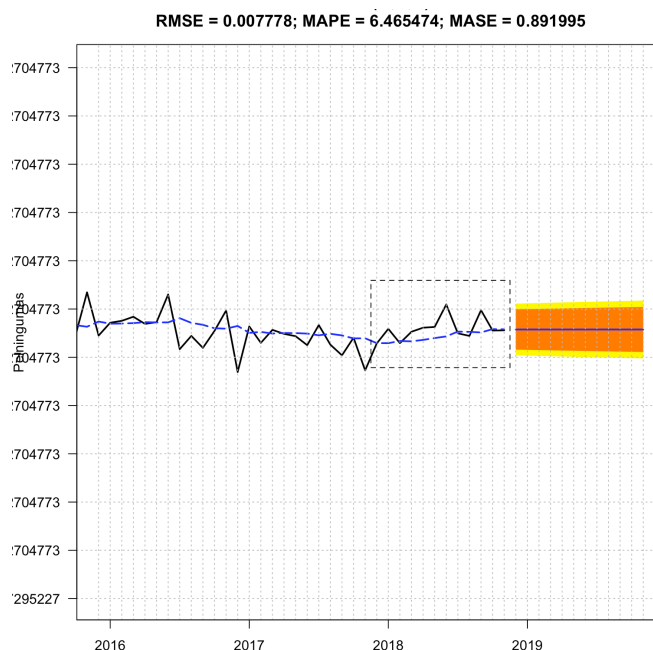
$P_i$  – prieš tai buvusių mėnesių pelningumo reikšmės.



20 pav. Slenkančio vidurkio vidutinių kvadratinų paklaidų įverčiai

### EkspONENTINIS GLODINIMAS

Apžvelgiama plačiau apie konkretų eksponentinio glodinimo atvejį. Gautas modelis ETS(M,N,N). Tai reiškia, kad turime multipliatyviuosius likučius, nėra trendo komponentės ir nėra sezono komponentės. Šiuo atveju modelis vadinamas paprastuoju eksponentiniu glodiniu.



21 pav. Pelningumo prognozė naudojant eksponentinio glodinimo metodą

Grafiškai pateikti pelningumo prognozės rezultatai, naudojant eksponentinio glodinimo metodą parodo jog šis metodas, kaip ir vidurkio, turi gana statinę ilgalaikę prognozę.

Įstačius visus komponentus į pelno lygtį, gauname pelno prognozės lygtį (30 dienų mėnesio pelnui prognozuoti):

$$Pelnas_{m\acute{e}n} = \left( \frac{I_{vid(d-1)}}{-0.00003(d-1)^2 + 0.0068(d-1) + 0.8217} * 30 \right) * \frac{1}{5} \sum_{i=1}^{m-5} (P_i) \quad (45)$$

### 3.3. Pelno prognozės neapibrėžtumo tyrimas

Neapibrėžtumo analizė šiame tyrime atliekama etapais:

- Pradinių duomenų neapibrėžtumo vertinimas;
- Neapibrėžtumo šaltiniai;
- Modelio parametrų jautrumo indeksų vertinimas.

#### 3.3.1. Pradinių duomenų neapibrėžtumas

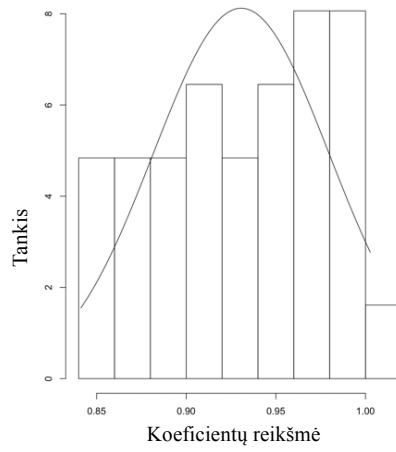
Šiame tyrime analizuojamos įmonės turimų pradinių duomenų neapibrėžtumas nėra stebimas, kadangi tiek išlaidų fiksavimas, tiek pelnas yra tiesiogiai stebima informacija. Išorės veiksniai taip pat dažniausiai neturi įtakos duomenų kokybei. Tais retais atvejais, sutrikus sistemos veiklai, gali atsirasti didesnių duomenų vėlavimų, bet mėnesio eigoje visos transakcijos yra agreguotos ir apdorotos teisingai, todėl duomenyse nestebimos trūkstamos reikšmės ar didelės nepaaiškinamos išskirtys.

#### 3.3.2. Neapibrėžtumo šaltiniai

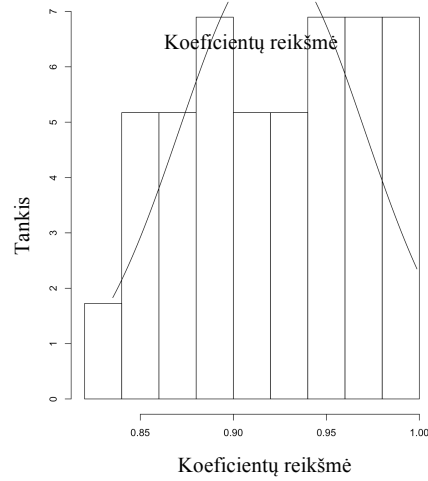
Neapibrėžtumas būdingas visiems finansų rinkos vertinimams, kuriuose matematiniai modeliai naudojami apibūdinti netiesiogiai stebėtą informaciją. Šio tyrimo atveju, atliekama pelno prognozė, todėl neapibrėžtumas atsiranda dėl klientų išlaidų dieninių vidurkių kaitos ir prognozei naudojamų vidurkių, t.y. prognozuojamo mėnesio dienos vidurkio koeficientui paskaičiuoti sudaromas modelis remiantis istoriniais duomenimis (vidurkio principu), kuris nepilnai atspindi realybę. Tokią išvadą galima daryti vertinant gautų koeficientų kitimo modelių deterministiniu koeficientu. Taip pat pelningumo rodiklis nėra pastovus kiekvieną mėnesį, todėl prognozei naudojamas vidurkių principu, kuris iš testuotų metodų turėjo mažiausią paklaidą. Vertinant esamą žinių apie modelio kiekvieno parametro reikšmės lygį, tikslinga naudoti visus prieinamus duomenis (informacijos šaltinius) ir jų pagrindu apibrėžti parametro reikšmių subjektyvų tikimybinį skirstinį [16 pav. Pelningumo rodiklių tankio histograma].

Gautų modelių [38, 39, 40, 41, 42], skirtų prognozuoti mėnesio išlaidas koeficientų skirstinių diagramos pateiktos grafiškai [22 pav., 23 pav., 24 pav., 25 pav., 26 pav.].

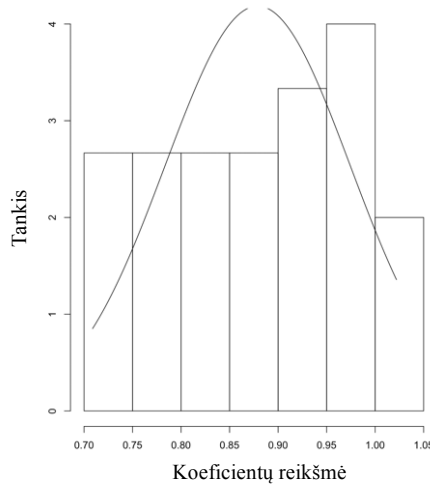




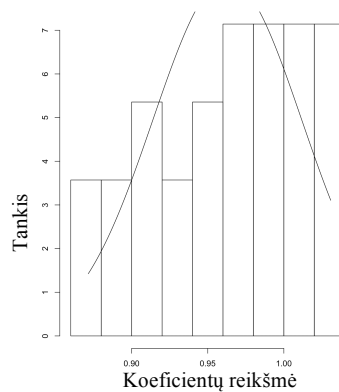
22 pav. 31 dienos mėnesių kelintų išlaidų koeficientų tankio diagrama



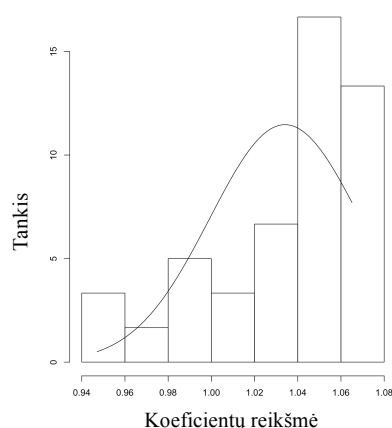
23 pav. 30 dienų mėnesių klientų išlaidų koeficientų tankio diagrama



24 pav. Sausio mėnesio klientų išlaidų koeficientų tankio diagrama



25 pav. Vasario mėnesio klientų išlaidų koeficientų tankio diagrama



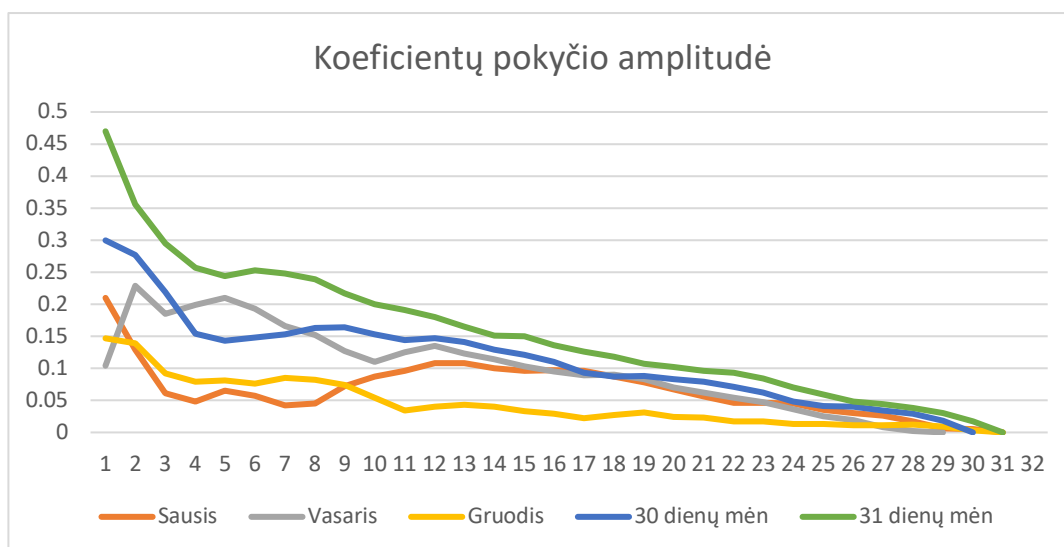
26 pav. Gruodžio mėnesio klientų išlaidų koeficientų tankio diagrama

Klientų dieninių vidutinių išlaidų prognozių modelių skirstinių apžvalga pateikiama lentelėje (žr. 2 lentelė). Didžiausias pasiskirstymo intervalas tarp prognozuojamų koeficientų stebimas sausio mėnesio prognozės rezultatuose. Kas reiškia, jog didžiausias neapibrėžtumo intervalas yra šio mėnesio prognozėje.

Metodas	Vidurkis	Standartinis nuokrypis	Minimali reikšmė	Maksimali reikšmė	Pokyčio intervalas
31 diena	0.9307	0.0491	0.8412	1.0026	0.1614
30 dienų	0.9207	0.0498	0.8352	0.9987	0.1635
Sausis	0.8789	0.0953	0.7088	1.022	0.3132
Vasaris	0.9629	0.0489	0.8715	1.0308	0.1593
Gruodis	1.0341	0.0348	0.9471	1.0651	0.118

6 lentelė. Klientų išlaidų koeficientų modelių skirstinių rezultatai

Vertinant mėnesių kiekvienos dienos vidutinių išlaidų koeficientų pokyčių amplitudę, įvertinama, kokia yra galima neapibrėžtumo sritis. Rezultatai pateikti grafiškai (žr. 27 pav. ). Išlaidų prognozės yra perskaičiuojamos kiekvieną dieną, siekiant stebėti suplanuotų biudžetų tikslų įgyvendinimą, todėl svarbu įvertinti ir kiekvienos dienos prognozės tikslumą. Stebint kiekvienos dienos koeficiento pokyčio amplitudę galima daryti prielaidą, kurią mėnesio dieną vertinti pelno prognozės rezultatus.



27 pav. Mėnesio dieninių vidutinių išlaidų koeficientų neapibrėžtumo srities amplitudė

Didžiausia koeficientų pokyčio amplitudė stebima mėnesio pradžioje, palaipsniui mažėjanti ir nuo vidurio stabilizuojasi. Toks didelis neapibrėžtumas mėnesio pradžioje aiškinamas pokyčiais rinkoje, metų laikų įtaka ir panašiai. Todėl išlaidų prognozės tiksliausios ir labiausiai atspindinčios realią situaciją būna nuo mėnesio vidurio.

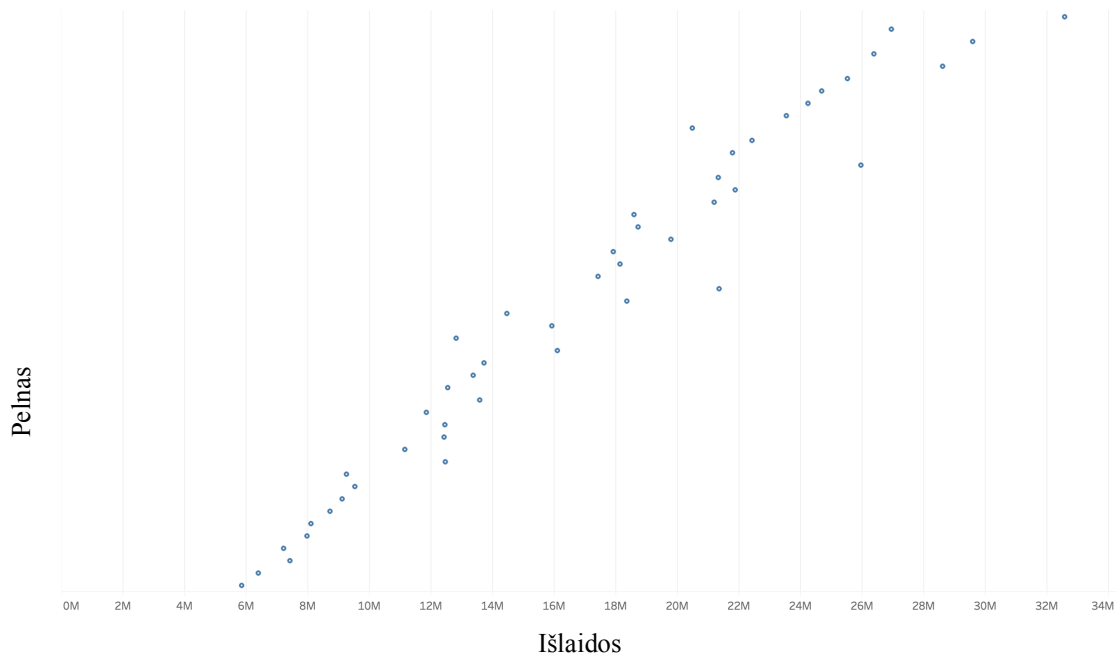
Neapibrėžtumas taip pat atsiranda ir prognozuojant pelningumo rodiklį. Pelningumas priklauso nuo veiksnių, kuriuos įvertinti būtų sudėtinga (dažniausiai tam būna naudojamos ekspertų žinios, arba informacija, gauta iš trečių šalių, pvz. tam tikroje šalyje yra planuojama prijungti naują klientą, kuris testavimo periodu generuos didelį kiekį transakcijų, už kurį bus mokamas statinis sutartinis minimalus mokestis, nepriklausomai nuo patirtų išlaidų).

Remiantis anksčiau pateikta pelningumo rodiklių pasiskirstymo diagrama (16 pav.), galima daryti prielaidą, jog didžiausia tikimybė pelningumo rodikliui yra patekti į vidurinio intervalo. Įvertinus tiek pelningumo, tiek išlaidų prognozės neapibrėžtumo šaltinius bei sklaidą, galima įvertinti kiekybinį parametru poveikį pelno prognozei.

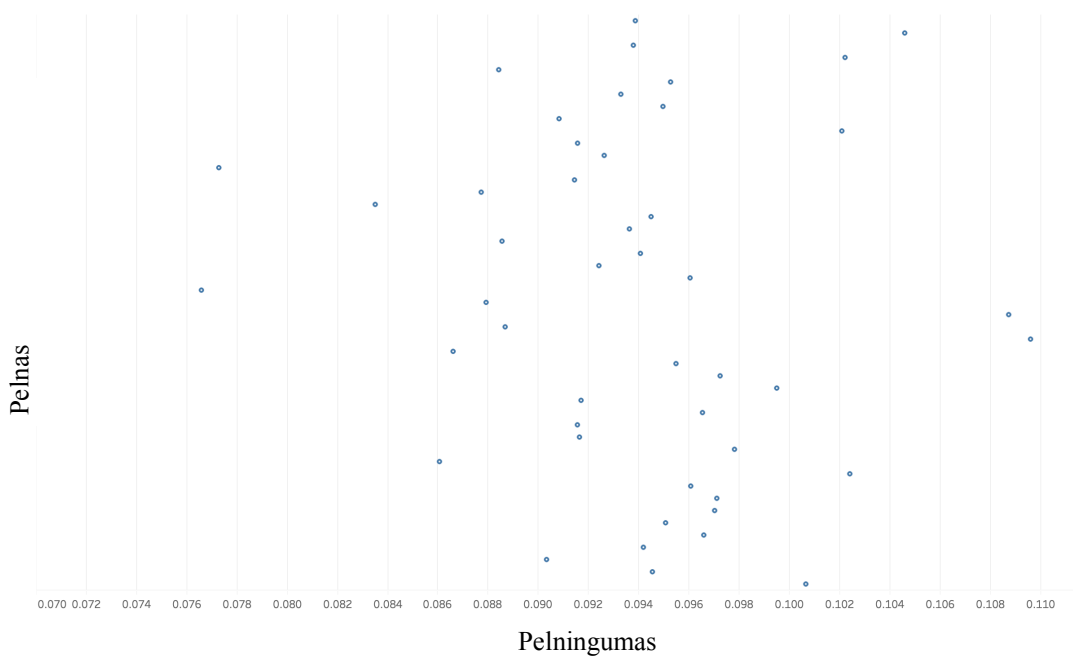
### 3.3.3. Modelio parametru jautrumo vertinimas

Pradiniame etape parametru jautrumą vertiname sklaido grafiku diagrama. Šiuo metodu nesunkiai galima pastebėti esamus tiesioginius ryšius tarp parametro ir modelio rezultato.

Pateiktame pelno prognozės ir išlaidų sklaidos grafike (29 pav.) matoma, jog didesnės išlaidų reikšmės susijusios su didesnio pelno prognozės reikšmėmis, o mažesnės – su mažesnėmis. Tačiau pelno prognozės ir pelningumo grafike (30 pav.) toks ryšys nestebimas. Tokios pirminės prielaidos visiškai nestebina, kadangi didžiausią svorį pelno prognozės lygtyje vis dėlto turi klientų išlaidų suma. Šią priklausomybę vertinsime kiekybiškai naudodami imties metodus.



28 pav. Pelno prognozės ir išlaidų sklaidos diagrama



29 pav. Pelno prognozės ir pelningumo rodiklio sklaidos diagrama

### Parametrų jautrumo vertinimai naudojant imties metodus

Parametrų jautrumo indeksų apskaičiavimui naudojant imties metodus, naudojami *Pearson* (koreliacijos koeficientas), *Kendall* ir *Spearman* (ranginiai metodai) koreliacijos koeficientai. Rezultatai pateikiami sekančioje lentelėje (žr. 7 lentelė).

	<i>Pearson</i>	<i>Kendall</i>	<i>Spearman</i>
Pelnas/Pelningumas	-0.07042576	-0.1045328	-0.1882516
Pelnas/Išlaidos	0.980799	0.9074931	0.9821924
Pelnas/Pelningumas ir Išlaidos	-0.0555006	-0.0970398	-0.0115171

7 lentelė. Koreliacijų koeficientų reikšmės

Kiekybiniai jautrumo indeksų įvertiniai tik patvirtino tai, ką buvo galima matyti pateikus rodiklius grafiškai. Pelningumo rodiklis turi nežymią neigiamą koreliaciją (neigiamą tiesinį ryšį) pelno skaičiavime, tačiau išlaidos turi labai stiprų teigiamą tiesinį ryšį su pelno prognozės reikšme. Tačiau šie koreliacijos koeficientai turi prasmę tik esant tiesinei priklausomybei ir nėra vertinama kelių parametų kombinacija, todėl buvo paskaičiuota koreliacija tarp abiejų parametų. Kadangi parametų tarpusavio sąveikos įtaka modelio rezultatui yra nežymi, todėl galima daryti išvadą, jog koreliacijos koeficientai yra tinkami naudoti jautrumo indeksams nustatyti.

Dar vienas jautrumo analizėje naudojamas imties metodas yra standartizuota tiesinė regresija. Pelno prognozės modelį  $(Pelnas = Išlaidos \times Pelningumas)$  išreikšime daugialype tiesine parametų funkcija:

$$Pelnas = F(Išlaidos_{m\acute{e}n}, Pelningumas_{m\acute{e}n}) = a + b_1 * Išlaidos + b_2 * Pelningumas \quad (46)$$

Kadangi koeficientai prie regresijos parametų turi būti normuoti parametų matavimo skalės, jie negali būti apskaičiuoti mažiausių kvadratų metodu. Normavimas atliekamas standartizuojant kiekvieną parametą ir modelio rezultatą:

$$\widehat{Išlaidos}_k = \frac{Išlaidos_k - E(Išlaidos)}{\sigma * Išlaidos} \quad (47)$$

$$\widehat{Pelningumas}_k = \frac{Pelningumas_k - E(Pelningumas)}{\sigma * Pelningumas} \quad (48)$$

$$\widehat{Pelnas} = \frac{Pelnas_k}{\sigma * Pelnas} \quad (49)$$

Čia  $k = 1, 2, \dots, n$ ;

$E(Išlaidos)$  - Išlaidų vidurkis;

$E(Pelningumas)$  – Pelningumo vidurkis;

$\sigma$  – standartinis nuokrypis;

$n$  – imties dydis.

Įstatę į sudarytas lygtis reikšmes, gauname standartizuotus regresijos koeficientus, kurie ir bus naudojami kaip parametų jautrumo indeksai. Standartizuota pelno regresijos lygtis:

$$\widehat{Pelnas} = 0.193 * Pelningumas + 1.03 * Išlaidos - 0.0000004067 \quad (50)$$

Gauta regresijos lygtis, kaip ir ankstesnėms regresijos lygtims, yra vertinama determinacijos koeficientu, kurio reikšmė šiuo atveju yra 0.9967. Šis koeficientas jautrumo analizėje nurodo modelio

rezultato neapibrėžtumo dalį, kurią galima paaiškinti modelio parametrų neapibrėžtumu. Likusi nepaaiškinta dalis galėtų būti analizuojama dispersijos išskaidymo metodais.

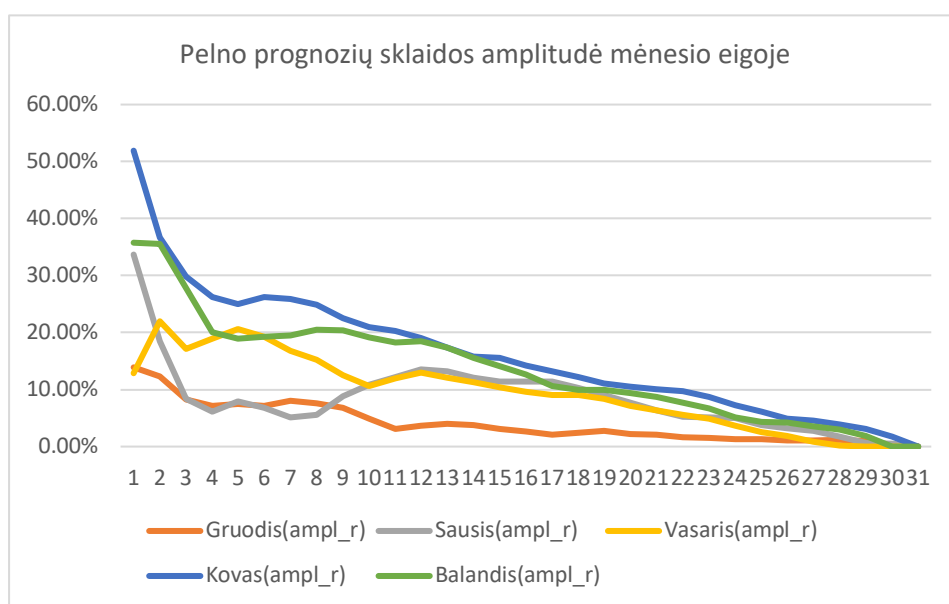
Iš gautos standartizuotos lygties galima daryti tokią interpretaciją – jeigu pelningumo parametą padidintumėm dydžiu, lygiu dispersijos ir parametro dydžio sandaugai:

$$\sigma * Pelningumas$$

atitinkamai pelno prognozė padidėtų vidutiniškai dydžiu, lygiu dispersijos, beta koeficiento ir pelno sandaugai:

$$\sigma * Pelnas * 0,193$$

Pelno prognozių neapibrėžtumo intervalai mėnesio eigoje pateikti sekančiame paveiksle (žr. 30 pav.).



30 pav. Pelno prognozės neapibrėžtumo intervalo kaita mėnesio eigoje

## Tyrimo apžvalga ir išvados

1. Susipažinus su internetinės reklamos technologijų tyrimais įvardinti pagrindiniai pajamų šaltiniai – gaunami iš interneto reklamos užsakovų išlaidų reklaminėms kampanijoms. Įvertinus pelno lygtį, toliau atliekamas pelno prognozės neapibrėžtumo tyrimas.
2. Atlikus mokslinių straipsnių analizę finansų modeliavimo temomis, buvo nustatyta, jog dažniausiai finansų prognozėms naudojami laiko eilučių metodai (autoregresiniai slenkančio vidurkio, eksponentinio glodinimo). Šie metodai buvo pasirinkti prognozuoti pelningumo koeficientą.
3. Apžvelgus publikuotus mokslinius straipsnius neapibrėžtumo tyrimo temomis, atliktas kiekybinis ir kokybinis neapibrėžtumo tyrimas tiek išlaidų, tiek pelningumo prognozei vertinti, bei pelno prognozės rodiklių jautrumo vertinimas.
4. Apžvelgus ir įvertinus analizuojamos kompanijos duomenis, įvertintos klientų išlaidų tendencijos skirtingų mėnesių eigoje, mėnesiai suskirstyti pagal tendencijų kryptis ir sudarytos mėnesio išlaidų prognozės lygtys pritaikant polinominės regresijos metodus. Modelių tiesiškumui vertinti naudotas determinacijos koeficientas.
5. Įvertinus mėnesinius pelningumo koeficientus, buvo pasirinkti laiko eilučių prognozių modeliai (ARIMA, ETS), taip pat vidurkio, slenkančio vidurkio ir *Croston's* metodai. Įvertinus modelių tikslumą remiantis RMSE, MAPE ir MASE statistikomis, prognozei parinktas 5 reikšmių slenkančio vidurkio metodas.
6. Naudojant prognozuotų rodiklių ir reikšmių neapibrėžtumo intervalus ir amplitudžių kitimo tendenciją, įvertintas rodiklių neapibrėžtumas. Einamojo mėnesio išlaidų prognozės dalyje, neapibrėžtumas tirtas kasdienių vidutinių išlaidų koeficientams, naudojant jų dienos pasiskirstymo intervalus. Nustatyta, jog iki mėnesio vidurio pokyčio amplitudė skirtingomis dienomis kinta nuo 10 % iki 50 %. Vidutiniškai amplitudės kitimo sritis yra mažesnė nei 10 % tik nuo mėnesio vidurio, todėl po šio termino rekomenduojama naudoti prognozės rezultatą verslo sprendimams priimti.
7. Pritaikius jautrumo analizę paremtą imties sklaidos metodais, paaiškėjo, jog didžiausią efektą pelno prognozės modelio rezultatams turėjo klientų išlaidų kiekis. Bendra pelno lygties kintamųjų koreliacija parodė, kad pelningumo ir išlaidų tarpusavio priklausomybės nėra, todėl daroma įtaka rezultato neapibrėžtumui yra nedidelė.

## Literatūros sąrašas

1. Financial Forecasting, Analysis and Modelling: A Framework for Long-Term Forecasting, Michael Samonas (2018)
2. Unconstraining Methods for Revenue Management Systems under Small Demand
3. K. Campbell, M. McKay & B. Williams – “Sensitivity analysis when model outputs are functions”, Reliability Engineering and Systems Safety 91 (2006)
4. Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District
5. Modelling customer online behaviours with neural networks: applications to conversion prediction and advertising retargeting. Yanwei Cui, Rogatien Tobossi, and Olivia Vigouroux (2018)
6. Ad Exchanges: Research Issues, S. Muthukrishnan (2009)
7. Internet Advertising: An Interplay among Advertisers, Online Publishers, Ad Exchanges and Web Users, Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, Jun Wang (2012)
8. Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting, Jun Wang, Weinan Zhang, Shuai Yuan (2017.07)
9. Deep Landscape Forecasting for Real-time Bidding Advertising. Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang, Weinan Zhang, Yong Yu (2019)
10. Alphabet Inc. pelnas. Internetinės reklamos pelnas sudaro 85% viso alphabet inc. pelno. Šaltinis: Alphabet Inc. Q3 2018 Results. (2019)
11. Optimal Bidding, Allocation and Budget Spending for a Demand Side Platform Under Many Auction Types, Alfonso Lobos, Paul Grigas, Zheng Wen, Kuang-chih Lee (2018)
12. Real-time Bidding for Online Advertising: Measurement and Analysis, Shuai Yuan, Jun Wang, Xiaoxue Zhao (2013)
13. The Vickrey Auction with a Single Duplicate Bidder Approximates the Optimal Revenue, Hu Fu, Christopher Liaw, Sikander Randhawa (2019)
14. Why Are the ARIMA and SARIMA not Sufficient, Shixiong Wang, Chongshou Li and Andrew Lim (2019)
15. Forecasting economic and financial time series: Arima vs. LSTM, Sima Siami Namin, Akbar Siami Namin (2018)
16. Curriculum Learning in Deep Neural Networks for Financial Forecasting, Allison Koenecke, Amita Gajewar (2019)
17. Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting, Rajat Sen , Hsiang-Fu Yu , and Inderjit Dhillon (2019)
18. Time Series Analysis of the Southern Oscillation Index using Bayesian Additive Regression Trees, S. van der Merwe (2009)
19. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks, Michael J Kane, Natalie Price, Matthew Scotch and Peter Rabinowitz (2014)
20. Fair Online Advertising, L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi (2019)
21. Matematinų modelių parametrų jautrumo ir rezultatų neapibrėžtumo statistiniai tyrimo metodai, Vytis Kopustinskas, Robertas Alzbutas, Juozas Augutis (2007)
22. Monitoring and prediction of an epidemic outbreak using syndromic observations, Alex Skvortsov and Branko Ristic (2011)
23. The impact of pharmacybernetic in reducing medication error, Dr. Muhammad Shahzad Aslam (2012)
24. Event-triggered Online Proactive Network Association to Mobile Edge Computing for IoT, Jian Zhang , Qimei Cui , Xuefei Zhang , Kwang-Cheng Chen , and Ping Zhang (2019)



25. Data-driven Neural Architecture Learning for Financial Time-series Forecasting, Dat Thanh Tran, Juho Kannianen, Moncef Gabbouj and Alexandros Iosifidis (2018)
26. K. Campbell, M. McKay & B. Williams – “Sensitivity analysis when model outputs are functions”, Reliability Engineering and Systems Safety 91 (2006)
27. Market Trend Prediction using Sentiment Analysis: Lessons Learned and Paths Forward, Andrius Mudinas, Dell Zhang, Mark Levene (2019)
28. Uncertainty analysis methods, Sankaran Mahadevan (2009)
29. Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods, Xin-Yao Qian (2017)
30. Comparison Between Bayesian and Frequentist Tail Probability Estimates, Nan Shena, B'arbara Gonz'aleza , Luis Ra'ul Pericchi (2019)
31. Bayesian time-aligned factor analysis of paired multivariate time series, Arkaprava Roy, Jana Schaich-Borg, David B Dunson (2019)
32. Implementation of Qualitative Uncertainty Guidance: A Worked Example, Ruth Salway, Gavin Shaddick (2012)
33. Bayes and Empirical Bayes Methods for Data Analysis, Bradley P. Carlin and Thomas A. Louis (1996)
34. Actual cost-per-click (CPC): Definition. Interaktyvi prieiga per [<https://support.google.com/google-ads/answer/6297?hl=en>], žiūrėta 2019-04-03.
35. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties, Michael D. Mastrandrea, Christopher B. Field, Thomas F. Stocker, Ottmar Edenhofer, Kristie L. Ebi, David J. Frame, Hermann Held, Elmar Kriegler, Katharine J. Mach, Patrick R. Matschoss, Gian-Kasper Plattner, Gary W. Yohe, and Francis W. Zwiers (2010)
36. Uncertainty-Aware Data Aggregation for Deep Imitation Learning, Yuchen Cui, David Isele, Scott Niekum, Kikuo Fujimura (2019)
37. Predicting financial market crashes using ghost singularities, Damian Smug, Peter Ashwin, Didier Sornette (2018)
38. Assessing and communicating climate change uncertainties, Saskia C. van Pelt (2014)
39. A quick guide to multisensi, an R package for multivariate sensitivity analyses, Caroline Bidot, Herv'e Monod, Marie-Luce Taupin (2018)
40. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach, Andrej-Nikolai Spiess, and Natalie Neumeyer (2010)
41. Sensitivity analyses: a brief tutorial with R package pse, Chalom, A., Mandai, C.Y., Prado, P.I. (2013)

## Priedai

### 1 priedas. Pelningumo rodiklio prognozė naudojant slenkančio vidurkio metodą (MA3, MA4, MA5)

Mėnuo	Pelningumas	Prognozė(n=3)	Paklaida(n=3)	Prognozė(n=4)	Paklaida(n=4)	Prognozė(n=5)	Paklaida(n=5)
2015m01	0.1007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m02	0.0945	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m03	0.0951	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m04	0.0966	0.0968	-0.0002	0.0000	0.0000	0.0000	0.0000
2015m05	0.1024	0.0954	0.0070	0.0967	0.0057	0.0000	0.0000
2015m06	0.0961	0.0980	-0.0020	0.0972	-0.0011	0.0979	-0.0018
2015m07	0.0903	0.0984	-0.0080	0.0975	-0.0072	0.0969	-0.0066
2015m08	0.0942	0.0963	-0.0021	0.0963	-0.0022	0.0961	-0.0019
2015m09	0.0971	0.0935	0.0036	0.0957	0.0014	0.0959	0.0012
2015m10	0.0916	0.0939	-0.0022	0.0944	-0.0028	0.0960	-0.0044
2015m11	0.1096	0.0943	0.0153	0.0933	0.0163	0.0939	0.0157
2015m12	0.0917	0.0994	-0.0077	0.0981	-0.0064	0.0966	-0.0049
2016m01	0.0970	0.0976	-0.0006	0.0975	-0.0005	0.0968	0.0002
2016m02	0.0978	0.0994	-0.0016	0.0975	0.0003	0.0974	0.0004
2016m03	0.0995	0.0955	0.0040	0.0990	0.0005	0.0976	0.0019
2016m04	0.0965	0.0981	-0.0016	0.0965	0.0000	0.0991	-0.0026
2016m05	0.0972	0.0979	-0.0007	0.0977	-0.0005	0.0965	0.0007
2016m06	0.1087	0.0978	0.0110	0.0978	0.0110	0.0976	0.0111
2016m07	0.0861	0.1008	-0.0148	0.1005	-0.0144	0.1000	-0.0139
2016m08	0.0916	0.0973	-0.0058	0.0971	-0.0056	0.0976	-0.0061
2016m09	0.0866	0.0955	-0.0088	0.0959	-0.0093	0.0960	-0.0094
2016m10	0.0936	0.0881	0.0056	0.0932	0.0004	0.0940	-0.0004
2016m11	0.1021	0.0906	0.0115	0.0895	0.0126	0.0933	0.0088
2016m12	0.0766	0.0941	-0.0175	0.0935	-0.0169	0.0920	-0.0154
2017m01	0.0955	0.0908	0.0047	0.0897	0.0058	0.0901	0.0054
2017m02	0.0887	0.0914	-0.0027	0.0919	-0.0033	0.0909	-0.0022
2017m03	0.0941	0.0869	0.0072	0.0907	0.0034	0.0913	0.0028
2017m04	0.0924	0.0927	-0.0003	0.0887	0.0037	0.0914	0.0010
2017m05	0.0914	0.0917	-0.0003	0.0927	-0.0012	0.0894	0.0020
2017m06	0.0877	0.0926	-0.0049	0.0916	-0.0039	0.0924	-0.0047
2017m07	0.0960	0.0905	0.0055	0.0914	0.0046	0.0909	0.0052
2017m08	0.0879	0.0917	-0.0038	0.0919	-0.0040	0.0923	-0.0044
2017m09	0.0835	0.0906	-0.0071	0.0908	-0.0073	0.0911	-0.0076
2017m10	0.0908	0.0892	0.0017	0.0888	0.0020	0.0893	0.0015
2017m11	0.0773	0.0874	-0.0101	0.0896	-0.0123	0.0892	-0.0119
2017m12	0.0884	0.0839	0.0046	0.0849	0.0035	0.0871	0.0013
2018m01	0.0945	0.0855	0.0090	0.0850	0.0095	0.0856	0.0089
2018m02	0.0886	0.0867	0.0018	0.0878	0.0008	0.0869	0.0016
2018m03	0.0933	0.0905	0.0028	0.0872	0.0061	0.0879	0.0054

2018m04	0.0950	0.0921	0.0029	0.0912	0.0038	0.0884	0.0066
2018m05	0.0953	0.0923	0.0030	0.0928	0.0024	0.0919	0.0033
2018m06	0.1046	0.0945	0.0101	0.0930	0.0116	0.0933	0.0113
2018m07	0.0926	0.0983	-0.0057	0.0970	-0.0044	0.0953	-0.0027
2018m08	0.0916	0.0975	-0.0059	0.0969	-0.0053	0.0961	-0.0046
2018m09	0.1022	0.0963	0.0060	0.0960	0.0062	0.0958	0.0064
2018m10	0.0938	0.0955	-0.0017	0.0977	-0.0040	0.0973	-0.0035
2018m11	0.0939	0.0959	-0.0020	0.0950	-0.0012	0.0970	-0.0031
2018m12		0.0966		0.0954		0.0948	
RMSE			0.0068		0.0069		0.0066

## 2 priedas. Pelningumo rodiklio prognozė naudojant slenkančio vidurkio metodus (MA6, MA9, MA12)

Mėnuo	Pelningumas	Prognoze(n=6)	Paklaida(n=6)	Prognoze(n=9)	Paklaida(n=9)	Prognoze(n=12)	Paklaida(n=12)
2015m01	0.1007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m02	0.0945	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m03	0.0951	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m04	0.0966	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m05	0.1024	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m06	0.0961	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2015m07	0.0903	0.0976	-0.0072	0.0000	0.0000	0.0000	0.0000
2015m08	0.0942	0.0958	-0.0017	0.0000	0.0000	0.0000	0.0000
2015m09	0.0971	0.0958	0.0013	0.0000	0.0000	0.0000	0.0000
2015m10	0.0916	0.0961	-0.0045	0.0963	-0.0047	0.0000	0.0000
2015m11	0.1096	0.0953	0.0143	0.0953	0.0143	0.0000	0.0000
2015m12	0.0917	0.0965	-0.0048	0.0970	-0.0053	0.0000	0.0000
2016m01	0.0970	0.0958	0.0013	0.0966	0.0004	0.0967	0.0004
2016m02	0.0978	0.0969	0.0009	0.0967	0.0011	0.0964	0.0015
2016m03	0.0995	0.0975	0.0020	0.0962	0.0033	0.0966	0.0029
2016m04	0.0965	0.0979	-0.0013	0.0965	0.0000	0.0970	-0.0005
2016m05	0.0972	0.0987	-0.0015	0.0972	0.0000	0.0970	0.0002
2016m06	0.1087	0.0966	0.0121	0.0976	0.0112	0.0966	0.0122
2016m07	0.0861	0.0995	-0.0134	0.0989	-0.0128	0.0976	-0.0116
2016m08	0.0916	0.0976	-0.0061	0.0982	-0.0067	0.0973	-0.0057
2016m09	0.0866	0.0966	-0.0100	0.0962	-0.0096	0.0970	-0.0104
2016m10	0.0936	0.0945	-0.0008	0.0957	-0.0020	0.0962	-0.0025
2016m11	0.1021	0.0940	0.0081	0.0953	0.0068	0.0963	0.0058
2016m12	0.0766	0.0948	-0.0182	0.0958	-0.0192	0.0957	-0.0191
2017m01	0.0955	0.0894	0.0061	0.0932	0.0023	0.0944	0.0010
2017m02	0.0887	0.0910	-0.0023	0.0931	-0.0044	0.0943	-0.0056
2017m03	0.0941	0.0905	0.0036	0.0922	0.0019	0.0936	0.0005
2017m04	0.0924	0.0918	0.0007	0.0905	0.0019	0.0931	-0.0007
2017m05	0.0914	0.0916	-0.0001	0.0912	0.0002	0.0928	-0.0013

2017m06	0.0877	0.0898	-0.0021	0.0912	-0.0035	0.0923	-0.0046
2017m07	0.0960	0.0916	0.0044	0.0913	0.0047	0.0905	0.0055
2017m08	0.0879	0.0917	-0.0038	0.0916	-0.0037	0.0914	-0.0034
2017m09	0.0835	0.0916	-0.0081	0.0900	-0.0065	0.0911	-0.0076
2017m10	0.0908	0.0898	0.0010	0.0908	0.0000	0.0908	0.0000
2017m11	0.0773	0.0896	-0.0123	0.0903	-0.0130	0.0906	-0.0133
2017m12	0.0884	0.0872	0.0012	0.0890	-0.0006	0.0885	-0.0001
2018m01	0.0945	0.0873	0.0072	0.0884	0.0061	0.0895	0.0050
2018m02	0.0886	0.0871	0.0015	0.0886	-0.0001	0.0894	-0.0008
2018m03	0.0933	0.0872	0.0061	0.0883	0.0050	0.0894	0.0039
2018m04	0.0950	0.0888	0.0062	0.0889	0.0060	0.0893	0.0056
2018m05	0.0953	0.0895	0.0058	0.0888	0.0065	0.0895	0.0057
2018m06	0.1046	0.0925	0.0121	0.0896	0.0150	0.0899	0.0147
2018m07	0.0926	0.0952	-0.0026	0.0920	0.0007	0.0913	0.0014
2018m08	0.0916	0.0949	-0.0033	0.0922	-0.0006	0.0910	0.0006
2018m09	0.1022	0.0954	0.0068	0.0938	0.0085	0.0913	0.0109
2018m10	0.0938	0.0969	-0.0031	0.0953	-0.0015	0.0928	0.0009
2018m11	0.0939	0.0967	-0.0028	0.0952	-0.0013	0.0931	0.0008
2018m12		0.0964		0.0958		0.0945	
RMSE			0.0068		0.0070		0.0068

### 3 priedas. Klientų išlaidų dienos vidurkių mėnesių vidurkiai (Sausis – Birželis)

Diena	Sausis	Vasaris	Kovas	Balandis	Gegužė	Birželis
1	0.6447	0.8221	0.7967	0.7921	0.7874	0.9507
2	0.6982	0.8908	0.8311	0.7827	0.8007	0.9369
3	0.7272	0.8980	0.8477	0.7956	0.8195	0.9254
4	0.7517	0.8898	0.8507	0.8091	0.8340	0.9146
5	0.7734	0.8951	0.8487	0.8190	0.8411	0.9068
6	0.7808	0.9054	0.8489	0.8318	0.8456	0.9094
7	0.7855	0.9109	0.8556	0.8442	0.8493	0.9160
8	0.7937	0.9188	0.8602	0.8476	0.8550	0.9242
9	0.8037	0.9257	0.8664	0.8523	0.8598	0.9328
10	0.8121	0.9322	0.8758	0.8587	0.8665	0.9351
11	0.8216	0.9369	0.8821	0.8670	0.8732	0.9363
12	0.8312	0.9440	0.8853	0.8770	0.8808	0.9423
13	0.8412	0.9495	0.8907	0.8879	0.8887	0.9476
14	0.8487	0.9524	0.8998	0.8976	0.8936	0.9512
15	0.8582	0.9566	0.9085	0.9048	0.9010	0.9565
16	0.8671	0.9619	0.9197	0.9115	0.9070	0.9608
17	0.8751	0.9646	0.9273	0.9156	0.9132	0.9625
18	0.8866	0.9667	0.9335	0.9224	0.9209	0.9637
19	0.8972	0.9696	0.9408	0.9288	0.9279	0.9670

20	0.9062	0.9728	0.9459	0.9378	0.9333	0.9697
21	0.9138	0.9752	0.9525	0.9454	0.9370	0.9729
22	0.9224	0.9792	0.9597	0.9511	0.9427	0.9775
23	0.9308	0.9828	0.9684	0.9567	0.9488	0.9812
24	0.9389	0.9857	0.9740	0.9628	0.9558	0.9833
25	0.9478	0.9893	0.9770	0.9685	0.9616	0.9859
26	0.9584	0.9932	0.9807	0.9765	0.9697	0.9904
27	0.9686	0.9973	0.9855	0.9845	0.9765	0.9949
28	0.9759	0.9996	0.9896	0.9913	0.9832	0.9984
29	0.9853	1.0000	0.9938	0.9980	0.9901	0.9999
30	0.9939		0.9982	1.0000	0.9920	1.0000
31	1.0000		1.0000		1.0000	

#### 4 priedas. Klientų išlaidų dienos vidurkių mėnesių vidurkiai (Liepa – Gruodis)

Diena	Liepa	Rugpjūtis	Rugsėjis	Spalis	Lapkritis	Gruodis
1	0.9493	0.8165	0.7808	0.7668	0.8142	0.9505
2	0.9484	0.8397	0.7824	0.7908	0.8461	0.9550
3	0.9602	0.8693	0.7886	0.8008	0.8485	0.9606
4	0.9679	0.8797	0.8006	0.8139	0.8612	0.9723
5	0.9727	0.8808	0.8076	0.8286	0.8675	0.9762
6	0.9842	0.8839	0.8149	0.8458	0.8752	0.9796
7	0.9878	0.8899	0.8311	0.8600	0.8836	0.9908
8	0.9894	0.8913	0.8418	0.8675	0.8896	0.9980
9	0.9887	0.8919	0.8517	0.8781	0.8989	1.0040
10	0.9901	0.8972	0.8569	0.8850	0.9033	1.0096
11	0.9909	0.9000	0.8644	0.8933	0.9053	1.0182
12	0.9906	0.9013	0.8685	0.9032	0.9066	1.0254
13	0.9917	0.9055	0.8748	0.9115	0.9111	1.0307
14	0.9924	0.9091	0.8879	0.9176	0.9168	1.0381
15	0.9896	0.9094	0.8979	0.9210	0.9222	1.0434
16	0.9889	0.9114	0.9054	0.9265	0.9272	1.0451
17	0.9901	0.9192	0.9124	0.9326	0.9299	1.0499
18	0.9891	0.9249	0.9211	0.9400	0.9325	1.0544
19	0.9879	0.9284	0.9300	0.9482	0.9359	1.0581
20	0.9905	0.9330	0.9397	0.9550	0.9412	1.0595
21	0.9940	0.9392	0.9485	0.9572	0.9472	1.0609
22	0.9949	0.9453	0.9565	0.9617	0.9548	1.0610
23	0.9967	0.9524	0.9613	0.9667	0.9632	1.0574
24	0.9975	0.9614	0.9668	0.9708	0.9702	1.0460
25	0.9987	0.9672	0.9728	0.9755	0.9756	1.0353
26	0.9997	0.9708	0.9794	0.9817	0.9796	1.0271
27	1.0023	0.9764	0.9843	0.9874	0.9854	1.0229

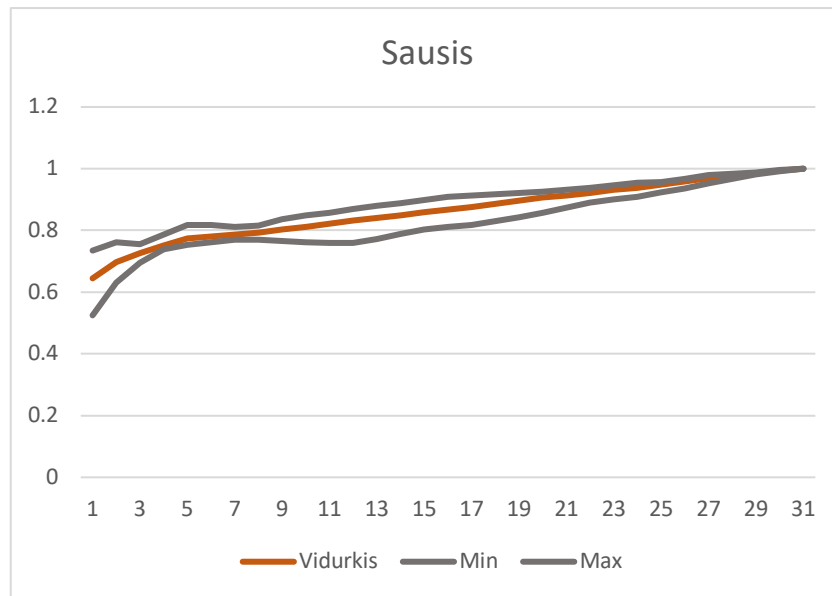
28	1.0045	0.9829	0.9913	0.9909	0.9901	1.0192
29	1.0034	0.9888	0.9977	0.9942	0.9959	1.0156
30	1.0009	0.9941	1.0000	0.9971	1.0000	1.0104
31	1.0000	1.0000		1.0000		1.0000

**5 priedas. Mėnesių vidurkių koeficientai**

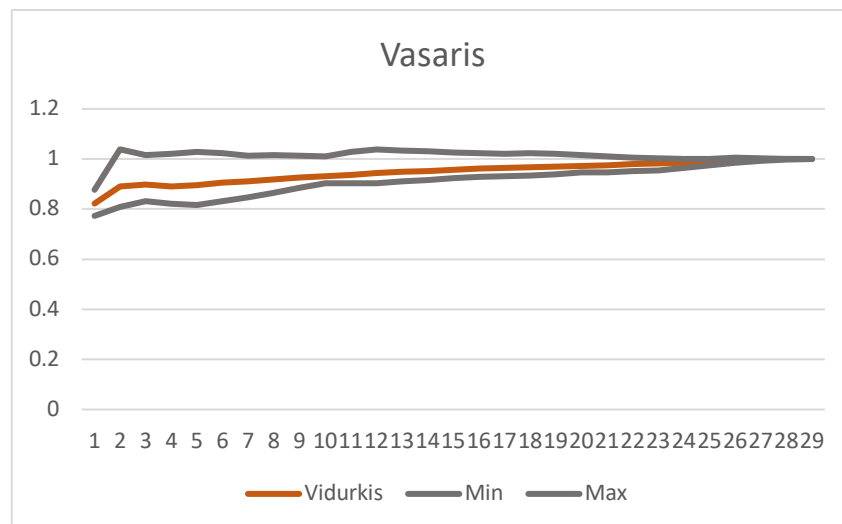
Diena	30 dienų mėnuo	31 dienos mėnuo
1	0.8344	0.8233
2	0.8370	0.8421
3	0.8395	0.8595
4	0.8464	0.8692
5	0.8502	0.8744
6	0.8578	0.8817
7	0.8687	0.8885
8	0.8758	0.8927
9	0.8839	0.8970
10	0.8885	0.9029
11	0.8933	0.9079
12	0.8986	0.9122
13	0.9053	0.9176
14	0.9134	0.9225
15	0.9203	0.9259
16	0.9262	0.9307
17	0.9301	0.9365
18	0.9349	0.9417
19	0.9404	0.9467
20	0.9471	0.9516
21	0.9535	0.9560
22	0.9600	0.9609
23	0.9656	0.9666
24	0.9708	0.9719
25	0.9757	0.9760
26	0.9815	0.9805
27	0.9873	0.9856
28	0.9928	0.9902
29	0.9979	0.9941
30	1.0000	0.9965
31		1.0000

**6 priedas.** Dieninių išlaidų pokyčio amplitudė mėnesio eigoje

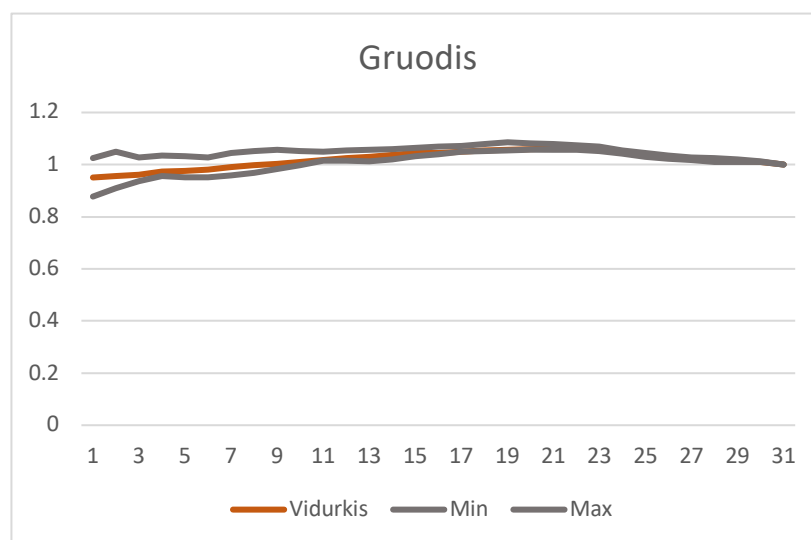
Diena	Sausis	Vasaris	Gruodis	30 dienų mėnuo	31 dienų mėnuo
1	0.2099	0.1041	0.1468	0.2996	0.4699
2	0.1290	0.2288	0.1392	0.2768	0.3557
3	0.0612	0.1852	0.0919	0.2187	0.2950
4	0.0483	0.1987	0.0794	0.1540	0.2570
5	0.0651	0.2098	0.0809	0.1427	0.2436
6	0.0577	0.1929	0.0759	0.1476	0.2527
7	0.0420	0.1665	0.0851	0.1530	0.2476
8	0.0455	0.1521	0.0822	0.1627	0.2393
9	0.0720	0.1271	0.0740	0.1639	0.2167
10	0.0871	0.1098	0.0539	0.1531	0.2005
11	0.0965	0.1247	0.0346	0.1445	0.1912
12	0.1085	0.1347	0.0398	0.1467	0.1801
13	0.1082	0.1232	0.0429	0.1406	0.1650
14	0.1006	0.1140	0.0403	0.1294	0.1513
15	0.0964	0.1035	0.0334	0.1210	0.1500
16	0.0968	0.0952	0.0296	0.1105	0.1364
17	0.0959	0.0894	0.0226	0.0932	0.1265
18	0.0872	0.0901	0.0272	0.0875	0.1179
19	0.0777	0.0829	0.0309	0.0879	0.1075
20	0.0669	0.0705	0.0238	0.0832	0.1025
21	0.0565	0.0620	0.0228	0.0787	0.0964
22	0.0463	0.0537	0.0177	0.0713	0.0930
23	0.0460	0.0471	0.0170	0.0626	0.0839
24	0.0452	0.0360	0.0132	0.0485	0.0703
25	0.0350	0.0251	0.0133	0.0409	0.0596
26	0.0303	0.0189	0.0114	0.0401	0.0483
27	0.0267	0.0081	0.0115	0.0347	0.0447
28	0.0173	0.0017	0.0125	0.0292	0.0385
29	0.0060	0.0000	0.0095	0.0186	0.0305
30	0.0048		0.0030	0.0000	0.0169
31	0.0000		0.0000		0.0000



31 pav. Sausio mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis

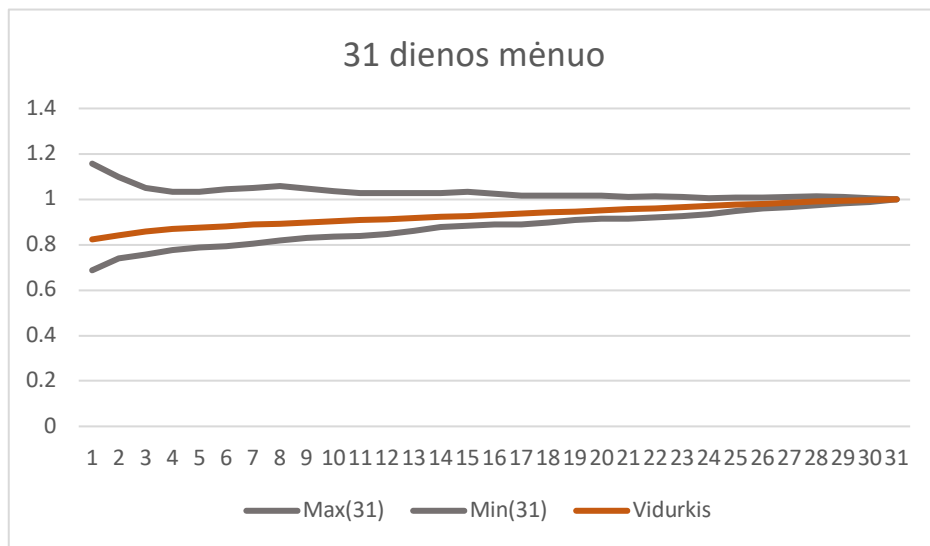


32 pav. Vasario mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis

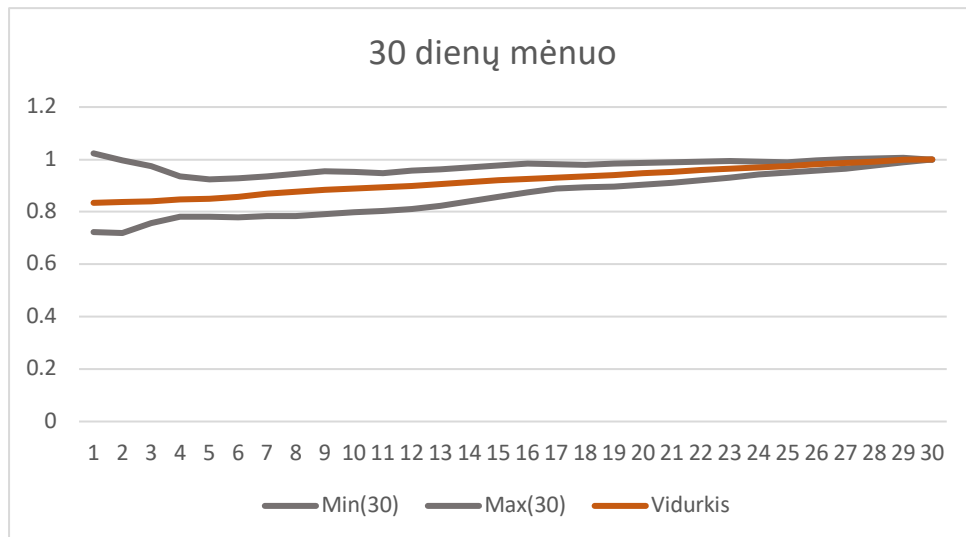


33 pav. Gruodžio mėnesio klientų dieninių išlaidų koeficientų neapibrėžtumo sritis





34 pav. 31 dienos mėnesių klientų dieninių išlaidų koeficientų neapibrėžtumo sritis



35 pav. 30 dienų mėnesių klientų dieninių išlaidų koeficientų neapibrėžtumo sritis

## 7 priedas. Programinis R kodas

```
require(goftest)
days_31 <- rnorm(31, mean=0.9307, sd=0.0491)
CvM_31 <- cvm.test(days_31, "pnorm", mean=0.9307, sd=0.0491)
ad_31 <- ad.test(rnorm(31, mean=0.9307, sd=0.0491))
```

```
CvM_31$p.value
ad_31$p.value
```

```
days_30 <- rnorm(30, mean=0.9207, sd=0.0498)
CvM_30 <- cvm.test(days_30, "pnorm", mean=0.9207, sd=0.0498)
ad_30 <- ad.test(rnorm(30, mean=0.9207, sd=0.0498))
```

```
CvM_30$p.value
ad_30$p.value
```

```

days_01 <- rnorm(31, mean=0.8789, sd=0.0953)
CvM_01 <- cvm.test(days_01, "pnorm", mean=0.8789, sd=0.0953)
ad_01 <- ad.test(rnorm(31, mean=0.8789, sd=0.0953))

CvM_01$p.value
ad_01$p.value

days_02 <- rnorm(29, mean=0.9629, sd=0.0489)
CvM_02 <- cvm.test(days_02, "pnorm", mean=0.9629, sd=0.0489)
ad_02 <- ad.test(rnorm(29, mean=0.9629, sd=0.0489))

CvM_02$p.value
ad_02$p.value

days_12 <- rnorm(31, mean=1.0341, sd=0.0348)
CvM_12 <- cvm.test(days_12, "pnorm", mean=1.0341, sd=0.0348)
ad_02 <- ad.test(rnorm(29, mean=0.9629, sd=0.0489))

CvM_12$p.value
ad_02$p.value

results <- cbind(CvM_12$p.value,CvM_02$p.value CvM_01$p.value)

#Person correlation
# set data folder, file and data type
data_fold <- './data_source/'
data_file <- 'priklausomybe'
data_file2 <- 'priklausomybe_st'

# read data from csv
my_data2 <- read.csv(paste(data_fold,data_file2, '.csv',sep=""))

class(my_data)

data2 <- cbind(my_data$date,my_data$Peln)
data4 <- cbind(my_data2$pelnas_st, my_data2$pelningumas_st, my_data2$islaidos_st)

row.has.na <- apply(data4, 1, function(x){any(is.na(x))})
sum(row.has.na)
filtered <- data4[!row.has.na,]

filtered[,1]

cor(filtered[,2],filtered[,1], method = c("pearson", "kendall", "spearman"))

cor(filtered[,1], filtered[,2], method = "kendall")
cor(filtered[,1], filtered[,3], method = "spearman")
cor(filtered[,2], filtered[,3], method = "pearson")

fit <- lm(filtered[,1] ~ filtered[,2])
fit2 <- lm(filtered[,1] ~ filtered[,3])

```

```

model <- lm(filtered[,1] ~ filtered[,2] + filtered[,3])
summary(filtered[,2])

mean(filtered[,3])
sd(filtered[,3])

pnorm(27.4, mean=50, sd=20)
pnorm(27.4, 50, 20)

da2 <- pnorm(c(25,50,75), mean = 11, sd = 2)
pnorm(1200,1000,200)

set.seed(124)
pelnas <- rnorm(100, 1586300, 641550.9)
pelnas[1:10]

pelningumas <- rnorm(100, 0.09387781, 0.006520686)
pelningumas[1:10]

islaidos <- rnorm(100, 17018346, 6957789)
islaidos[1:10]

data_normal_2 <- rbind(pelnas,pelningumas,islaidos)

transposed_2 <- t(data_normal_2)

# Laiko eilučių prognozės metodai
library(AnomalyDetection)
library(imputeTS)
library(Hmisc)
library(pastecs)

TIME <- Sys.time()
if("pacman" %in% rownames(installed.packages()) == FALSE) {install.packages("pacman")}
require(pacman)
pacman::p_load(fpp, xts, forecastHybrid, prophet)

# set data folder, file and data type
data_fold <- './data_source/'
data_file <- 'pelningumas'
data_type <- 'month'

# read data from csv
my_data <- read.csv(paste(data_fold,data_file, '.csv',sep=""))

head(my_data)
plot(my_data[,2])
summary(my_data)
#describe(my_data)
#stat.desc(my_data)

```

```

ylabel <- colnames(my_data)[2]
xlabel <- sprintf("Time (%s)",data_type)
data_freq <- switch(data_type,"week" = 52, "month" = 12, "quarter" = 4, "day" = 365, 0)

# set forecasting horizon
fh <- switch(data_type,"week" = 28, "month" = 12, "quarter" = 8, "day" = 365, 0)
fhstlm <- 1
fw <- 3 * fh # (data_freq * 2) # historical data to show

# set interval forecast limits, neapibrėžtumo intervalai
interval_forecast <- 'level=c(80,90)' # level=c(80,90) or fan=TRUE ?

sdN <- 3
dP <- 6

# function to estimate grid lines in forecast plot
tickSize <- function(range,minCount){
  logMaxTick <- log10(range/minCount)
  exponent <- floor(logMaxTick)
  mantissa <- 10^(logMaxTick-exponent)
  af <- c(1,2,5) # allowed factors
  mantissa <- af[findInterval(mantissa,af)]
  return(mantissa*10^exponent)
}

# load data with date labels, create train set
single_time_series <- my_data[,ylabel]
notNA <- !is.na(single_time_series)
staNA <- min(which(notNA == TRUE))
endNA <- max(which(notNA == TRUE))
single_time_series <- single_time_series[staNA:endNA]

data_init <- as.numeric(unlist(strsplit(as.character(my_data[staNA,1]), substr(data_type,1,1))))

ts_full <- ts(single_time_series,start=data_init,frequency=data_freq)
ts_obs <- time(ts_full) # observations
ts_short <- window(ts_full,end=ts_obs[length(ts_obs)-fh])
class(my_data)

# settings for Y and X axis
# ylimits <- c(floor(floor(mean(ts_full)-
3*sd(ts_full))/10)*10,ceiling(ceiling(mean(ts_full)+4*sd(ts_full))/10)*10)
ylimits <- c(min(my_data[,2])-mean(my_data[,2]),max(my_data[,2])+mean(my_data[,2]))
# ylimits <- c(0.07,0.11)
ystep <- tickSize(range(ylimits)[2]-range(ylimits)[1],5) # y axis step
yticks <- seq(ylimits[1],ylimits[2],by=ystep)
xlimits <- c(ts_obs[1],ts_obs[length(ts_obs)])

# plot time series
plot(ts_full,ylab=ylabel)
tsdisplay(ts_full,main=ylabel)
boxplot(ts_full, main="Pelningumo stačiakampė diagrama") # boxplot

```

```

plot(ts_full, ylab = "Pelningumas", type="l", main="Pelningumo pokytis")
abline(h=0.08027, col="red")
abline(h=0.10494, col="red")

hist(ts_full, xlab="Pelningumas", main="Pelningumo rodiklių pasiskirtymas", prob=TRUE,
cex.lab=2, cex.main=1.5)
lines(density(ts_full, from = min(ts_full), to = max(ts_full)))

summary(my_data)
# autocorrelation test.
adf.test(ts_full)

decompose_addit_prof <- stl(ts_full, s.window="periodic", t.window=5, robust=TRUE)
profitability_reminder <- ts_full$time.series[, 'reminder']
profitability_reminder <- decompose_addit_prof$time.series[,3]

plot(decompose_addit_prof)
require(imputeTS)
acf(stl_remainder_182)
kadangi nulines hipotezes neatmetame (h0 - priklausomybe nuo laiko), tai autokoreliacijos nera.
Box.test(profitability_reminder, type = c("Box-Pierce", "Ljung-Box"))

#vidurkis <- meanf(ts_full, h = 2, level = c(80, 95), fan = FALSE, lambda = NULL,biasadj =
FALSE, bootstrap = FALSE, npaths = 5000)
#vidurkis_2 <- meanf(ts_full, h = 2, level = c(80, 95), fan = FALSE, lambda = NULL,biasadj =
FALSE, bootstrap = TRUE, npaths = 1000)

# plot histogram
hist(ts_full,col=7,main=paste('Histogram of',ylabel)); par(new=TRUE);
M <- mean(ts_full); SD = sd(ts_full)
x <- seq(-4,4,length=200)*SD + M
plot(x,dnorm(x,M,SD),type="l",lty=2,axes=FALSE,xlab="",ylab=")

# additional plots
uYears <- unique(floor(time(ts_full)))
nYears <- length(uYears)
seasonplot(ts_full,year.labels=TRUE, year.labels.left=TRUE, ylab=ylabel, col=rainbow(nYears),
pch=20)
monthplot(ts_full,ylab=ylabel,col.base=2)

# time-series decomposition
plot(stl(ts_full,s.window='periodic',robust=TRUE),main=paste('Decomposition by Loess
of',ylabel))
plot(decompose(ts_full, type="multiplicative"))
plot(decompose(ts_full, type="additive"))

# define univariate time-series models
ts_models_full <- c(
  sprintf('fc_full <- meanf(ts_full,%s,h=%d)',interval_forecast,fh),
  sprintf('fc_full <- forecast(auto.arima(ts_full,seasonal=F),%s,h=%d)',interval_forecast,fh),
  sprintf('fc_full <- forecast(auto.arima(ts_full,seasonal=T),%s,h=%d)',interval_forecast,fh),

```

```

    sprintf('fc_full <- forecast(bats(ts_full),%s,h=%d)',interval_forecast,fh),
    sprintf('fc_full <- forecast(ets(ts_full),%s,h=%d)',interval_forecast,fh),
    sprintf('fc_full <- croston(ts_full,h=%d)',fh),
)
ts_models_short <- gsub("_full","_short",ts_models_full)
ts_models_method <- NULL
ts_models_line <- NULL
ts_models_forecast <- matrix(NA,length(ts_models_full),fh)
ts_short_RMSE <- NULL
ts_short_MAPE <- NULL
ts_short_MASE <- NULL

# run univariate time-series models
for (i in seq(length(ts_models_full))) {
  cat(sprintf('\n\n\n[%d] %s\n',i,ts_models_full[i]))
  eval(parse(text=ts_models_short[i]));
  eval(parse(text=ts_models_full[i]));
  short_obs <- time(ts(c(fc_full$fitted[(length(fc_full$fitted)-
fw+1):length(fc_full$fitted)],fc_full$mean),start=ts_obs[length(ts_obs)-
fw+1],frequency=data_freq))
  print(fc_full$model)
  print(fc_full$mean)
  ts_models_forecast[i,] <- fc_full$mean
  true_short <- window(ts_full,start=ts_obs[length(ts_obs)-fh+1])
  acc <- accuracy(fc_short,true_short)
  cat('\n')
  print(acc[,c("MAE", "RMSE", "MAPE", "MASE", "Theil's U")])
  ts_short_RMSE[i] <- acc['Test set','RMSE']
  ts_short_MAPE[i] <- acc['Test set','MAPE']
  ts_short_MASE[i] <- acc['Test set','MASE']
  # create time-series plot with forecast and errors

plot(fc_full,ylim=ylimits,xlim=c(short_obs[1],short_obs[length(short_obs)]),axes=FALSE,lwd=2,co
ol=1,ann=FALSE,shadecols='oldstyle')
lines(fc_full$fitted,col=4,lty=5,lwd=2)
#ts_models_method[i] <- gsub("\\s+$"," ",fc_full$method)
mm <- length(fc_short$method)
if (mm > 1) {
  ts_models_method[i] <- paste(mm,"pack",sep="-")
} else {
  ts_models_method[i] <- strsplit(gsub("\\s+$"," ",fc_short$method),split="\n")[[1]][1]
  ts_models_method[i] <- fc_short$method
  ts_models_method[i] <- gsub(" model| method","",ts_models_method[i])
  ts_models_method[i] <- gsub("BATS",ts_models_method[i])
}
ts_models_line[i] <- sub('fc_full <- ',",",ts_models_full[i])
title(paste(sprintf("%d.",i),ts_models_method[i],sprintf(paste("\nRMSE = %5.',dP,'f; MAPE =
%5.',dP,'f; MASE = %5.',dP,'f,sep=""",acc['Test set','RMSE'],acc['Test set','MAPE'],acc['Test
set','MASE']),sep=""),xlab=xlabel,ylab=ylabel)
xticks <- unique(round(c(ts_obs,time(fc_full$mean))))
axis(1,at=xticks,labels=TRUE)

```

```

axis(2,yticks,las=1)
abline(h=yticks, v=short_obs, col="gray", lty=3);
rect(time(ts_short)[length(ts_short)]+1/data_freq/2, min(true_short)-ystep/2,
time(fc_short$mean)[length(fc_short$mean)]+1/data_freq/2, max(true_short)+ystep/2,
lty="dashed")
box()
}

# sort models by RMSE
idx <- order(ts_short_RMSE)
topModels <-
data.frame(cbind(ts_models_method[idx],round(ts_short_RMSE[idx],digits=dP),round(ts_short_M
APE[idx],digits=dP),round(ts_short_MASE[idx],digits=dP),ts_models_line[idx],ts_models_forecas
t[idx,]))

# add date on forecast
if (data_freq==4) {
timeLabel <- as.character(as.yearqtr(time(fc_full$mean)))
} else {
timeLabel <- as.character(as.yearmon(time(fc_full$mean)))
}
colnames(topModels) <- c('Method name','RMSE','MAPE','MASE','Function runline',timeLabel)

# print model table
sep_line <- '====='
cat(sprintf('\n\n\n%s[ %d %ss ahead forecasting ]%s\n\nTOP-%d forecasting methods for \'%s\'
time series\n\n',sep_line,fh,data_type,sep_line,length(ts_models_full),ylabel))
print(cbind(idx,topModels[,1:4]))
cat('\n')

summary(eval(parse(text=ts_models_full[14])))

# save model table
# if resulting .csv doesn't open nicely - change write.csv into write.csv2
write.csv2(topModels,file=paste(data_file,'_',fh,'_steps_forecast.csv',sep=""))
print(Sys.time() - TIME)

# pl - pelnas
revenue <- function(pl,)
data_file <- 'ratios30'
data_file <- 'ratios01'
data_file <- 'ratios02'
data_file <- 'ratios12'

data_fold <- './data_source/'
data_file <- 'ratios'
ratios_data <- read.csv(paste(data_fold,data_file, '.csv',sep=""))
class(ratios_data)
data2<-data.matrix(ratios_data, rownames.force = NA)

```

```

hist(data2, probability=TRUE, main="Išlaidų dieninio vidurkio koeficientų prognozės
pasiskirstymas")
xx <- seq(min(data2), max(data2), length=100)
lines(xx, dnorm(xx, mean=mean(data2), sd=sd(data2)))

min(data2)
max(data2)
mean(data2)
sd(data2)
max(data2)-min(data2)

require(goftest)
days_31 <- rnorm(31, mean=0.9307, sd=0.0491)
CvM_31 <- cvm.test(days_31, "pnorm", mean=0.9307, sd=0.0491)
ad_31 <- ad.test(rnorm(31, mean=0.9307, sd=0.0491))

CvM_31$p.value
ad_31$p.value

days_30 <- rnorm(30, mean=0.9207, sd=0.0498)
CvM_30 <- cvm.test(days_30, "pnorm", mean=0.9207, sd=0.0498)
ad_30 <- ad.test(rnorm(30, mean=0.9207, sd=0.0498))

CvM_30$p.value
ad_30$p.value

days_01 <- rnorm(31, mean=0.8789, sd=0.0953)
CvM_01 <- cvm.test(days_01, "pnorm", mean=0.8789, sd=0.0953)
ad_01 <- ad.test(rnorm(31, mean=0.8789, sd=0.0953))

CvM_01$p.value
ad_01$p.value

days_02 <- rnorm(29, mean=0.9629, sd=0.0489)
CvM_02 <- cvm.test(days_02, "pnorm", mean=0.9629, sd=0.0489)
ad_02 <- ad.test(rnorm(29, mean=0.9629, sd=0.0489))

CvM_02$p.value
ad_02$p.value

days_12 <- rnorm(31, mean=1.0341, sd=0.0348)
CvM_12 <- cvm.test(days_12, "pnorm", mean=1.0341, sd=0.0348)
ad_02 <- ad.test(rnorm(29, mean=0.9629, sd=0.0489))

CvM_12$p.value
ad_02$p.value

results <- cbind(CvM_12$p.value, CvM_02$p.value, CvM_01$p.value)

#Pearson correlation
# set data folder, file and data type
data_fold <- './data_source/'

```



```

data_file <- 'priklausomybe'
data_file2 <- 'priklausomybe_st'

# read data from csv
my_data2 <- read.csv(paste(data_fold,data_file2, '.csv',sep=""))

class(my_data)

data2 <- cbind(my_data$date,my_data$Pelna)
data4 <- cbind(my_data2$pelnas_st, my_data2$pelningumas_st, my_data2$islaidos_st)

row.has.na <- apply(data4, 1, function(x){any(is.na(x))})
sum(row.has.na)
filtered <- data4[!row.has.na,]

filtered[,1]

cor(filtered[,2],filtered[,1], method = c("pearson", "kendall", "spearman"))

cor(filtered[,1], filtered[,2], method = "kendall")
cor(filtered[,1], filtered[,3], method = "spearman")
cor(filtered[,2], filtered[,3], method = "pearson")

fit <- lm(filtered[,1] ~ filtered[,2])
fit2 <- lm(filtered[,1] ~ filtered[,3])

model <- lm(filtered[,1] ~ filtered[,2] + filtered[,3])
summary(filtered[,2])

mean(filtered[,3])
sd(filtered[,3])

pnorm(27.4, mean=50, sd=20)
pnorm(27.4, 50, 20)

da2 <- pnorm(c(25,50,75), mean = 11, sd = 2)
pnorm(1200,1000,200)

set.seed(124)
pelnas <- rnorm(100, 1586300, 641550.9)
pelnas[1:10]

pelningumas <- rnorm(100, 0.09387781, 0.006520686)
pelningumas[1:10]

islaidos <- rnorm(100, 17018346, 6957789)
islaidos[1:10]

data_normal_2 <- rbind(pelnas,pelningumas,islaidos)

transposed_2 <- t(data_normal_2)

```