



KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

Karolina Liaudanskaitė

TELEKOMUNIKACIJŲ BENDROVĖS VARTOTOJŲ
ELGSENOS TYRIMAI

Baigiamasis magistro projektas

Vadovai

Doc. dr. Tomas Ruzgas
Prof. dr. Rimantas Gatautis

KAUNAS, 2017

KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

TELEKOMUNIKACIJŲ BENDROVĖS VARTOTOJŲ
ELGSENOS TYRIMAI

Baigiamasis magistro projektas

Didžiųjų verslo duomenų analitika (kodas 621G12002)

Vadovai

Doc. dr. Tomas Ruzgas

Prof. dr. Rimantas Gatautis

Recenzantai

Doc. dr. Mindaugas Kavaliauskas

Prof. dr. Rita Kuvykaitė

Projektą atliko

Karolina Liaudanskaitė



KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

(Fakultetas)

Karolina Liaudanskaitė

(Studento vardas, pavardė)

Didžiųjų verslo duomenų analitika, 621G12002

(Studijų programos pavadinimas, kodas)

„Telekomunikacijų bendrovės vartotojų elgsenos tyrimai“
AKADEMINIO SAŽININGUMO DEKLARACIJA

20 17 m. birželio 02 d.
Kaunas

Patvirtinu, kad mano, **Karolinos Liaudanskaitės**, baigiamasis projektas tema „Telekomunikacijų bendrovės vartotojų elgsenos tyrimai“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

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

(vardą ir pavardę įrašyti ranka)

(parašas)

TURINYS

IŽANGA.....	7
1. LITERATŪROS APŽVALGA	8
1.1. Vartotojų elgsenos samprata.....	8
1.2. SOR koncepcijos taikymas vartotojų elgsenos tyrimuose.....	10
1.3. Didžiųjų duomenų panaudojimas vartotojų elgsenos tyrimų kontekste.....	12
1.4. Duomenų tyrybos sistemų apžvalga	17
2. TYRYMO METODAI	21
2.1. Aprašomoji statistika	21
2.2. Suderinamumo hipotezių tikrinimo metodai	22
2.3. Koreliacinė analizė	23
2.4. Dispersinė analizė.....	24
2.5. Požymių priklausomumo lentelių tyrimas	25
2.6. Logistinė regresija	26
2.7. Faktorinė analizė.....	27
2.8. Diskriminantinė analizė.....	29
3. TYRIMO REZULTATAI	30
3.1. Darbe naudojami duomenys ir statistinės analizės modeliai	30
3.2. Aprašomoji statistika	31
3.3. Suderinamumo hipotezių tikrinimas.....	32
3.4. Koreliacinė analizė	33
3.5. Dispersinė analizė.....	36
3.6. Požymių priklausomumo lentelių tyrimas	40
3.7. Logistinė regresija	42
3.8. Faktorinė analizė.....	44
3.9. Diskriminantinė analizė.....	45
IŠVADOS.....	48
LITERATŪRA	49
PRIEDAS 1	51
PRIEDAS 2	52
PRIEDAS 3	53

Liaudanskaitė, Karolina. Telekomunikacijų bendrovės vartotojų elgsenos tyrimai. *Magistro* baigiamasis projektas / vadovai doc. dr. Tomas Ruzgas ir prof. dr. Rimantas Gatautis; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Mokslo kryptis ir sritis: fiziniai mokslai, matematika

Reikšminiai žodžiai: vartotojų elgsena, SOR, didieji duomenys, statistika

Kaunas, 2017. 55 p.

SANTRAUKA

Magistro baigiamajame projekte apžvelgiama vartotojų elgsenos, didžiųjų duomenų bei jų tyrybos sistemų mokslinė literatūra siekiant išanalizuoti telekomunikacijų bendrovės vartotojų elgseną. Darbe pateikiama analizuojamų statistinės analizės metodų apžvalga: aprašomosios statistikos, suderinamumo hipotezių tikrinimo, koreliacinės analizės, dispersinės analizės, požymių priklausomumo lentelių tyrimo, logistinės regresijos, faktorinės analizės bei diskriminantinės analizės metodai. Šių statistinės analizės metodų pagalba buvo siekiama daugiau sužinoti apie turimus duomenis, tai yra išsiaiškinti priklausomybes tarp kintamųjų bei kaip ir kokie faktoriai veikia vartotojų elgseną. Visų metodų realizavimui buvo pasirinkta SAS programinė įranga.

Liaudanskaitė, Karolina. Telecommunications Company's Consumers Behaviour Research. *Master's thesis* / supervisors assoc.prof. Ph.D. Tomas Ruzgas and assoc. prof. Ph.D. Rimantas Gatautis. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: Physical Sciences, mathematics

Key words: consumers behaviour, SOR, big data, statistics

Kaunas, 2017. 55 p.

SUMMARY

Master's thesis provides an overview of scientific literature which includes publications about consumers behaviour, big data and data science platforms. It gives a better understanding of the telecommunications company's consumers behavior. There are also presented these statistical analysis methods: descriptive statistics, hypothesis testing, correlation analysis, analysis of variance, frequency tables analysis, logistic regression, principal components analysis and discriminant analysis. These methods were used to find out about the data, correlation between variables and how and which factors influence telecommunications company's consumers behaviour the most. SAS software was used in this thesis.

IŽANGA

Dabartiniame pasaulyje vartojimas yra vienas pagrindinių šalies ekonomiką veikiančių faktorių, tad vartotojų vaidmuo itin svarbus tiek bendrai šalies ekonomikai, tiek konkretiems verslo subjektams. Didėjant vartotojų skaičiui, augant jų poreikims, besikeičiant gyvenimo būdui - vartojimas tampa vis sudėtingesne problema. Norint plėtoti verslą, gauti didesnes pajamas ir pelną, įgyti konkurencinį pranašumą, būtina patenkinti vartotojų poreikius ir lūkesčius. Šiam tikslui įgyvendinti aktualu atlikti vartotojų elgsenos tyrimus, atskleidžiančius vartotojų elgsenos ypatumus. Pasinaudojant vartotojų elgsenos tyrimų rezultatais galima nusistatyti konkrečią strategiją bei siūlyti tam tikras prekes ar paslaugas tikslinei vartotojų grupei siekiant maksimalaus vartotojų pasitenkinimo teikiamomis paslaugomis ar prekėmis. Iš tiesų visame pasaulyje labai gerai suprantama vartotojų elgsenos tyrimų svarba siekiant patenkinti savo vartotojų poreikius, o didieji duomenys yra puikus naujas įrankis šiam tikslui pasiekti, kadangi vartotojai kiekvieną dieną palieka informaciją, kuri gali būti tikslingai panaudota priimant sprendimus susijusius su didesniu vartotojų lūkesčių patenkinimu.

Šiame magistrinio baigiamajame projekte analizuojami telekomunikacijų bendrovės duomenys apie jų vartotojus.

Darbo tikslas – atlikti telekomunikacijų bendrovės vartotojų elgsenos tyrimus.

Darbo tikslui pasiekti iškelti **uždaviniai**:

1. Pateikti vartotojų elgsenos, didžiųjų duomenų ir jų tyrybos tematika atliktų tyrimų analizę;
2. Išsirinkti reikiamus metodus vartotojų elgsenos tyrimui atlikti;
3. Atlikti telekomunikacijų bendrovės vartotojų elgsenos tyrimą;

Darbe naudoti tyrimo **metodai**:

- Mokslinės literatūros analizė ir apibendrinimas;
- Statistinės informacijos rinkimas ir duomenų analizė;
- Suderinamumo hipotezių tikrinimas, logistinė regresija, koreliacinė analizė, dispersinė analizė, faktorinė analizė, diskriminantinė analizė, požymių priklausomumo lentelių tyrimas, pasinaudojant SAS programine įranga.

1. LITERATŪROS APŽVALGA

1.1. Vartotojų elgsenos samprata

Mokslinėje literatūroje galima rasti daugybę vartotojo elgsenos apibrėžimų. Dažnai vartotojų elgsena apibūdinama kaip žmonių poreikiai, skatinantys pirkti [1]. Pasak Kotler'io ir Armstrong'o (1994) tiriant vartotojų elgseną siekiama išsiaiškinti kaip žmonės perka, ką jie perka, kada perka ir kodėl jie perka [2]. Solomon'o (1995) pateikiamas apibrėžimas apibūdina vartotojų pirkimo elgseną kaip individualių klientų, jų grupių ar organizacijų paslaugų ir prekių pasirinkimo, pirkimo ir naudojimo procesą siekiant patenkinti savo poreikius ir norus [3]. Panašų apibrėžimą siūlo Schiffman'as ir Kanuk'as (2000). Autoriai vartotojų elgseną aprašo kaip elgesį, kurį vartotojai išreiškia, kai renkasi ir perka prekes ar paslaugas panaudodami savo turimus išteklius, siekiant patenkinti savo poreikius [4]. Anot Blackwell'o, Miniard'o ir Engel'o (2006) vartotojų pirkimo elgsena yra sudėtingas ir dinamiškas klausimas, kuris negali būti paprastai ir lengvai apibrėžtas [5]. Stallworth'o (2008) manymu vartotojų elgsena tai kompleksinė veikla, apimanti paslaugų ir prekių pirkimą bei naudojimą, kurį lemia vartotojų emociniai ir psichologiniai poreikiai [6]. Kotler'is ir Keller'is (2011) teigia, kad vartotojų elgsena tai mokslas apie tai kaip individualūs klientai, grupės ar organizacijos renkasi, perka, naudoja ir disponuoja idėjomis, prekėmis ir paslaugomis savo poreikių ir norų patenkinimui. Tai apima vartotojų veiksmus ir po šiais veiksmais paslėptus motyvus. Marketingo specialistai tikisi, jog išsiaiškinę vartotojų priežastis pirkti tam tikras prekes ar paslaugas jie tiksliai galės apibrėžti kuris produktas reikalingas rinkoje, kuris jau nebebus vartojamas ir patį geriausią būdą tai prekei pateikti [7].

Nors pateikiama išties daug vartotojų elgsenos apibrėžimų, jie visi veda prie bendros išvados, jog vartotojų elgsena tai vartotojų prekių ar paslaugų rinkimosi, pirkimo ir naudojimosi procesas siekiant patenkinti savo poreikius. Svarbu paminėti, jog šis procesas itin permainingas kadangi nuolat besikeičiantys fiziniai ir psichologiniai vartotojų poreikiai keičia ir jų elgseną.

Chand'as (2016) išskyrė pagrindinius veiksnius, darančius įtaką vartotojų elgsenai [8]:

1. Marketingo veiksniai, apimantys produkto dizainą, kainą, reklamą, pakuotę, paskirstymą.
2. Asmeniniai veiksniai, pavyzdžiui, amžius, lytis.
3. Psichologiniai veiksniai, kurie apima pirkimo motyvus, suvokimą ir požiūrį į patį produktą.
4. Situaciniai veiksniai, tokie kaip fizinė ir socialinė pirkimo aplinka bei laikas.
5. Socialiniai veiksniai – socialinis statusas, išsilavinimas ir pajamų lygis.
6. Kultūriniai veiksniai – religija, papročiai.

Autoriai (Kotler, Keller, 2011) teigia, jog paprastas poreikis tampa motyvu, kai jis pasidaro pakankamai intensyvus, kad priverstų vartotoją veikti. Toliau, pateikiamos trys žinomiausios žmogaus motyvacijos teorijos [7]. *Zigmundo Froido teorija*. Pasak Zigmund Freud'o žmogaus

elgseną lemia nesąmoningi psichologiniai veiksniai, kurių asmuo negali visiškai suprasti. Asmuo, kuris domisi konkrečiais prekiniais ženklais reaguoja ne tik į įvardytas produkto galimybes, bet ir kitas, mažiau sąmoningai suprantamas užuominas, pavyzdžiui, formą, dydį, svorį, medžiagą, spalvą ar prekinį ženklą. Mokslininkai tiriantys motyvacinius veiksnius dažnai ima kelių dešimčių vartotojų interviu siekiant atskleisti gilesnius motyvus, kurie paskatino pirkimą. Šiam tikslui pasiekti naudojama daugybė metodų, tokių kaip žodžių asociacijos, sakinio pabaigimas, nuotraukų interpretavimas. *Maslovo teorija*. Abraham'as Maslow'as siekė paaiškinti, kodėl žmonės turi tam tikrų poreikių tam tikru metu. Atsakydamas į šį klausimą autorius pateikė žmogaus poreikių išdėstymą hierarchine tvarka nuo svarbiausių iki mažiausiai svarbių. Pirmiausia turi būti patenkinti žmogaus fiziologiniai, saugumo, socialiniai, savivertės poreikiai ir tik tada savęs realizavimo. (žr. 1.1.1 pav.). *Herzbergo teorija*. Frederick'as Herzberg'as sukūrė dviejų veiksmų teoriją, kuri atskiria faktorius sukeliančius nepasitenkinimą ir faktorius sukeliančius pasitenkinimą. Nepasitenkinimą sukeliančių veiksmų nebuvimas nepakankama motyvacija pirkti. Pavyzdžiui, kompiuteris, parduodamas be garantijos sukeltų nepasitenkinimo jausmą, tačiau tas pats kompiuteris parduodamas su garantija pasitenkinimo jausmo nesukeltų, nes vartotojas tai priimtų kaip natūralų dalyką. Pasak šios teorijos, pardavėjai turėtų daryti viską, kad būtų išvengta vartotojo nepasitenkinimą keliančių veiksmų (prastas aptarnavimas, kokybės politikos nebuvimas) nors tai ir negarantuos pardavimo. Be to, itin svarbu identifikuoti veiksmus, sukeliančius didžiausią pasitenkinimą ir juos pateikti vartotojui.

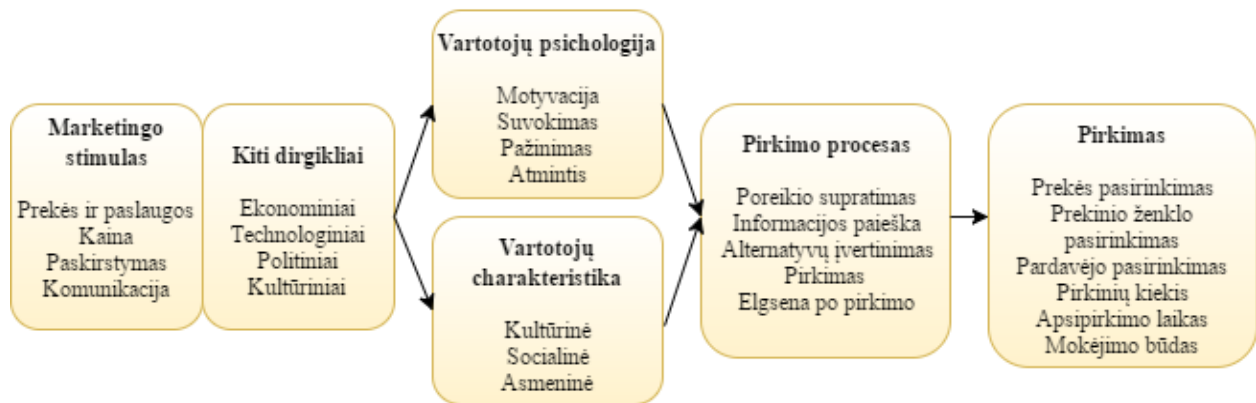


1.1.1 pav. Maslow'o poreikių hierarchija (Kotler, Keller, 2009)

Egan'as (2011) mano, jog geras vartotojų elgsenos supratimas yra indėlis į šalies ekonomikos būklę. Autorius teigia, kad prekių ir produktų kokybė yra itin gera tose šalyse, kurios skiria itin daug dėmesio vartotojų elgsenos supratimui. Tai savo ruožtu padidina produktų ir paslaugų konkurencingumą tarptautinėje rinkoje bei didina šalies eksporto potencialą [9]. Tačiau, kaip teigia Kotler'is ir Keller'is (2011), nepaisant įdedamų didelių pastangų suprasti vartotojų elgseną, labai sunku nustatyti tikslias priežastis, kodėl vartotojas perka ir teikia pirmenybę vienam produktui ar

paslaugai. Manoma, kad vartotojai linkę priimti sprendimus remiantis savo emocine būseną net patys to neįsąsmonėdami [7].

Atspirties taškas padedantis suprasti vartotojų elgseną yra stimulo – reakcijos modelis (Kotler, Keller, 2009) pateiktas 1.1.2 pav [7]. Marketingo ir aplinkos stimulus kartu su daugybe psichologinių veiksnių ir individualių žmogaus charakterio ypatybių paveikia vartotojo sąmonę ir sąlygoja sprendimo priėmimą bei apsisprendimą pirkti. Marketingo specialistų uždavinys yra suprasti kas vyksta vartotojo sąmonėje tarp stimulo ir galutinio apsisprendimo pirkti.



1.1.2 pav. Vartotojų elgsenos modelis (sudaryta pagal Kotler, Keller, 2009)

Taigi, labai svarbu suprasti vartotojų pirkimo elgseną ir suprasti būdus kaip klientai renkasi produktus ar paslaugas. Tokie faktai labai svarbūs gamintojams bei paslaugų teikėjams, kadangi gali suteikti konkurencinį pranašumą. Pasinaudojant vartotojų elgsenos tyrimų rezultatais galima nusistatyti konkrečią strategiją bei siūlyti tam tikras prekes ar paslaugas tikslinei klientų grupei siekiant patenkinti jų poreikius ir lūkesčius. Dėl šių priežasčių toliau buvo nagrinėtas stimulo – organizmo – reakcijos modelis.

1.2. SOR koncepcijos taikymas vartotojų elgsenos tyrimuose

1982 m. Donovanas ir Rossiter atliko pirmąjį empirinį bandymą siekiant sužinoti apie mažmeninės prekybos aplinkos poveikį vartotojų elgsenai. Jų bandymas iš aplinkos psichologijos modelio peraugo į stimulo-organizmo-reakcijos modelio pagrindą (Eroglu, Machleit, Davis, 2003) [10]. Aplinkos poveikio užuominos buvo laikomos dirgikliu, vartotojų emocinės reakcijos – organizmu, o priartėjimo arba vengimo elgsena - reakcija. Šis tyrimas padėjo empiriškai įrodyti, jog mažmeninės prekybos aplinka iš tiesų labai stipriai veikia pirkimo elgseną. Sekant šio bendro tyrimo pėdomis, vėliau buvo telkiamas dėmesys į konkrečius aplinkos aspektus, tokius kaip muzika, apšvietimas, kvapas ar spalvos.

Autoriai (Daugherty, Epp, Hoskins, Johnson, Mailton, McHugh, 2010) teigia, jog SOR modelis yra skirtas apibūdinti paslaugos aplinkos poveikį vartotojų elgsenai bei išskiria toliau išvardintus stimulo, organizmo bei reakcijos komponentus [11].

Stimulus:

- Pasinaudojant penkiais žmogaus pojūčiais (rega, klausa, uoslė, skonis, lytėjimas) patraukti vartotojo dėmesį.

Organizmas:

- Dirgiklių/stimulo gavėjas. Apima darbuotojus ir vartotojus.

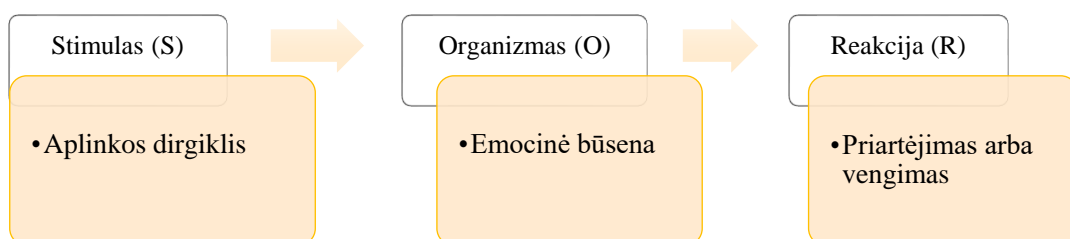
Reakcijų rinkinys:

- Vartotojų reakcija ir elgesys kaip atsakas į stimulą/dirgiklį.

Trys emocinės būsenos:

1. Pasitenkinimas/nepasitenkinimas (suteiktos paslaugos pasitenkinimo lygis).
 2. Sujaudinimas/ nesujaudinimas (sujaudinimo laipsnis).
 3. Valdymas/nuolankumas (vartotojų kontrolės ir gebėjimo laisvai veikti laipsnis paslaugų aplinkoje)
- Priartėjimas/vengimas.

Peng'as, Kim'as (2014) SOR teoriją apibrėžia kaip santykį tarp pirkimo vertybių, požiūrio, vartotojų gebėjimo reguliuoti savo emocinius pirkinius, ketinimus pakartotinai pirkti, aplinkos dirgiklių poveikį pakartotinam pirkimui bei pačią vartotojų elgseną. Prekybos aplinka veikia kaip dirgiklis, stimulus (S), kuris turi įtakos vartotojui, tai yra organizmui (O) ir sukelia priėmimo arba atmetimo reakciją (R) (žr. 1.2.1 pav.). Visa tai apima pačios parduotuvės paiešką, ketinimus pirkti bei pirkti pakartotinai. Ši teorija analizuoja aplinkos veiksnius (spalva, apšvietimas, muzika, perpildymas, kvapas, išdėstymas), kurie veikia vartotojų vidinę būseną ir daro įtaką galutiniam vartotojo apsisprendimui, reakcijai. Taipogi SOR modelis buvo panaudotas siekiant nustatyti, kad internetinių parduotuvių vartotojų vidinė motyvacija skiriasi ir kad šie motyvai turi didelį poveikį vartotojų pasitenkinimui [12].



1.2.1 pav. SOR modelio struktūra (sudaryta pagal Peng, Kim, 2014).

Chang'as, Eckman'as, Yan'as (2011) teigia, jog klasikiniame SOR modelyje stimulus apibrėžiamas kaip veiksnys, kuris daro įtaką vidinei individo būsenai ir taip jį gali paveikti savo naudai. Kuomet vartotojų elgsena yra nusakoma SOR modeliu, stimulus apibrėžiamas kaip išorinis veiksnys ir susideda ne tik iš marketingo komplekso kintamųjų, bet ir kitų aplinkos veiksnių. Tuo tarpu reakcija (R) SOR modelyje reprezentuoja galutinį vartotojo reakcijos rezultatą, kuris gali būti arba priartėjimas, arba vengimas. Priartėjimo elgsena reiškia teigiamus veiksmus, kurie apima

ketinimą likti parduotuvėje, toliau naršyti internetiniame puslapyje ar sprendimo priėmimą – pirkti. Taipogi akcentuojama, jog parduotuvės aplinka susideda iš daugybės elementų ir charakteristikų, kurios sukuria tam tikrą atmosferą organizmui. Parduotuvės aplinką stengiamasi suprojektuoti taip, kad parduotuvės elementai sukeltų teigiamas emocijas, jausmus vartotojui ir taip didintų pirkimo tikimybę. Taip pat pabrėžiama pardavimo aplinkos charakteristikų derinių svarba vartotojo elgsenai, o ne tik atskiri elementai, kadangi sujungti mažmeninės parduotuvės aplinkos elementai reprezentuoja aplinką, dizainą, socialinius veiksnius, kurie teigiamai veikia vartotojo emocinę būseną bei skatina pirkti. Siekiant geriau išanalizuoti aplinką 1986m. Baker'is pasiūlė taikyti bendrą tipologiją, kuri apima socialinius veiksnius (darbuotojai, vartotojai, parduotuvėje esantys žmonės), dizaino veiksnius (vizualiniai ženklai tokie kaip išdėstymas, tvarka, švarumas, spalva) ir aplinkos veiksnius (nevaizdinės užuominos, pavyzdžiui, kvapas ir garsas). 1992m. Bittner'io pasiūlyta aplinkos tipologija buvo labiau orientuota į paslaugas mažmeninės prekybos kontekste ir plačiau apima tokius aplinkos veiksnius: penkis žmogaus pojūčius, išdėstymą, funkcionalumą, ženklus, simbolius bei signalus per kuriuos norima perduoti informaciją vartotojui [13]. Autorių (Eroglu, Machleit, Davis, 2003) teigimu, internetinės mažmeninės prekybos parduotuvės taipogi sukuria aplinką, kuri paveikia pirkėjo reakciją. Nors ši aplinka stokoja taktinių bei uoslės signalų, internete pardavėjas gali manipuliuoti ne tik vizualiniais, bet ir klausos ženklais, kurie gali sukelti efektingą internetinių svetainių lankytojų reakciją [10]. Panašiai kaip tradicinių parduotuvių taip ir interneto svetainių aplinka (spalvos, grafika, išdėstymas, dizainas) gali suteikti informaciją apie pardavėją, prekių ar paslaugų kokybę, tikslinę rinką, o taip pat pardavėjo reagavimą apsilankymo svetainėje metu.

Iš tiesų, stimulo – organizmo – reakcijos koncepcija išties sudėtinga, kadangi mus supančioje aplinkoje egzistuoja daugybė dirgiklių. Nustatymui kaip konkretus stimulus (spalvų išdėstymas, parduotuvės planas, garso lygis, apšvietimas, reklama) paveikia vartotojo emocinį atsaką, reikalinga atlikti daugybę tyrimų bei bandymų.

1.3. Didžiųjų duomenų panaudojimas vartotojų elgsenos tyrimų kontekste

Iš tiesų pati duomenų sąvoka labai plati, pagal Svilar, Chakraborty, Kanioura (2013) į šią sąvoką įeina visų klientų informacija, statistika ar atitinkama verslo informacija, kurią įmonė kaupia ne vienerius metus. Didieji duomenys yra didėjantys duomenų rinkiniai, kuriuos bendrovė įgyja ir kurie gali būti analizuojami siekiant atskleisti įvairias tendencijas bei asociacijas, kurios yra naudingos verslui. Didieji duomenys yra tokios apimties, kad jų neįmanoma išanalizuoti rankiniu būdu ar apdoroti standartiniais metodais, tačiau jų panaudojimas įvairiose srityse, tame tarpe ir vartotojų elgsenos kontekste, vis auga [14].

Per pastarąjį šimtmetį įvyko precedento neturintys įvykiai skaitmeniniame pasaulyje, kurie sukėlė itin didelius pokyčius tiek ekonomikos, tiek finansų, tiek marketingo srityse. Su didėjančia vidutiniąja klase, išaugo nuolat socialiniais tinklais besinaudojančių žmonių skaičius. Internet Live Stats duomenimis 2016 m. 46% pasaulio gyventojų turėjo prieigą prie interneto ryšio, kai prieš 20 metų šis skaičius nesiekė nei 1%. Žmonių besinaudojančių socialiniais tinklais skaičius taip pat auga. Tokie pokyčiai keičia ir žmonių gyvenimo būdą, elgseną. Iš tiesų, šiandien internetu naudojamosi kiekviename žingsnyje, nuo paprasčiausios orų prognozės peržiūros iki kelionių bilietų pirkimo, renginių organizavimo ar el. pašto tikrinimo. Naudodamiesi šiomis paslaugomis, paliekama daugybė informacijos, kuri gali būti sumaniai panaudota priimant sprendimus, kadangi tokia informacija gaunama greitai ir realiu laiku. Visame pasaulyje labai gerai suprantama vartotojų elgsenos tyrimų svarba siekiant patenkinti savo vartotojų poreikius, o didieji duomenys yra puikus įrankis šiam tikslui pasiekti (Stoicesu, 2015). Pasinaudodamos turima informacija įmonės gali siūlyti savo prekes ar paslaugas tik tiems, kam jų reikia ir tik tada, kai jų reikia [15].

Pasak Marr'o (2016) įprastai buvo manoma, jog tam tikra prasme marketingas yra viena meno formų, tačiau pastaruoju metu didieji duomenys įnešė didelę dalį mokslo į šią sritį. Marketingo specialistai tampa vis labiau priklausomi nuo įvairių duomenų nei bet kada anksčiau. Nors duomenys bei jų analitika niekada nepakeis marketingo srityje reikalingo kūrybiškumo, kuris slepiasi po geriausiomis marketingo reklamomis, didieji duomenys tapo neįtikėtiniu įrankiu suteikiant įžvalgas dar geresniems sprendimams ateityje [16].

Vojtovič, Navickas, Gruzauskas (2016) mano, jog didžiųjų duomenų koncepcija gali būti apibūdinta keliomis charakteristikomis. Pirmasis elementas yra kiekis. Paskutiniaisiais dešimtmečiais įmonių duomenys smarkiai išaugo, tą patį galima pasakyti ir apie klientus. Mažmenine prekyba užsiimančios įmonės labai dažnai renka informaciją apie vartotojus, todėl atsiradusios didelės prekybos istorijos duomenų bazės reikalauja naujų serverių bei vis didesnių debesų kompiuterijos galimybių. Kita problema yra tai, jog nėra pakankamai informacijos, kuri galėtų būti naudojama duomenų tyrybai kadangi vartotojai bijo, kad jų informacija gali nutekėti ar ja bus manipuluojama, nors ir surinkta papildoma informacija galėtų būti naudojama siekiant teikti geresnes paslaugas. Svarbu pabrėžti, kad reikalingas ne tik didelis informacijos kiekis, bet reikalinga ir kokybinė informacija. Jei įmonės turi tam tikrą informaciją apie savo klientų profesiją, pomėgius, gyvenimo būdą, jie išanalizuodami šią informaciją gali teikti/pasiūlyti asmenines paslaugas ir produktus skirtus būtent jiems. Šis analizės tipas vis svarbesnis šiandienos pasaulyje, kadangi duomenų įvairovė vis didėja. Autoriai taip pat pateikia pagrindinius duomenų tyrybos metodus skirtus didiesiems duomenims (žr. 1.3.1 lentelę). Aprašomoji statistika, praeities informacijai taikomi duomenų tyrybos algoritmai apibūdina sėkmės ar nesėkmės priežastis. Prognozavimo analizei naudojama užsakymų ir klientų pirkimų istorija ir bandoma prognozuoti

jų elgesį. Šis metodas gali būti naudojamas kuriant geresnes marketingo strategijas. Svarbu paminėti, jog metodai, kurie naudojami analizuoti ir pateikti išvalgas gali būti sujungiami trimis etapais. Pirmasis etapas yra duomenų rinkimas. Antrasis etapas - duomenų tyryba, trečiasis - rezultatų interpretacija, aiškinimas. Siekiant tinkamai surinkti informaciją galima naudoti keletą metodų [17].

1.3.1 lentelė. Duomenų analizės metodai (sudaryta pagal Vojtovič, Navickas, Gruzauskas, 2016)

Algoritmas	Metodo principas	Tikslas
Sprendimų medžiai	Konstruoja sprendimų medį grindžiamą regresijos analize tarp įvesties duomenų ir antrinių duomenų.	Prognozuoja vartotojų elgseną; Padeda priimti geresnius sprendimus.
Logistinė regresija	Aprašo priklausomybę tarp įvesties duomenų ir antrinių duomenų ir suformuoja regresijos modelį.	Prognozės mažmeninės prekybos nuolaidų metu; Įvertina svarbius veiksnius.
Tiesinė regresija		
Laiko eilutės	Statistinė priklausomybė vertinama tarp įėjimo duomenų ir antrinių duomenų.	
Klasteriai	Duomenys yra sugrupuojami pagal apskaičiuotus atstumus ar vertinant skirtingų grupių tikimybes.	Klientų segmentavimas pagal marketingo kanalo identifikaciją; Įmonės struktūros koregavimas pagal identifikuotus segmentus.
Sekos klasteriai (Sequence clustering)	Taikomas tas pats principas kaip ir klasterizavime; Be to dar atsižvelgiama ir į įvykio tikimybę bei seką.	Identifikuoja vartotojų elgseną pagal jų pirkimo istoriją.
Asociatyvumo taisyklės	Identifikuoja duomenų derinius ir tikimybes jiems nutikti.	Naudojama pirkinų krepšelio analizei.
Neuroniniai tinklai	Metodas naudodamasis įvesties duomenimis apskaičiuoja įmanomus derinius. Tada apskaičiuoti duomenys vėl analizuojami galimoms kombinacijoms.	Prognozuoja klientų elgesį; Padeda geresniems sprendimų priėmimams; Priskiria klientus konkrečiam segmentui.

Marr (2016) pateikia sąrašą pasiūlymų kaip didieji duomenys ir jų analizė, gali pagerinti marketingo sprendimus [16]:

- Gali padėti aiškiau apibrėžti savo idealų klientą. Anksčiau, marketingo specialistams dažnai tekdavo nuspėti savo idealaus vartotojų segmento demografinę padėtį. Pasinaudojant didžiais duomenimis galima tiksliai sužinoti daugybę detalių apie vartotojus, įskaitant tokius dalykus kaip, puslapiuose, kuriuose vartotojas dažniausiai lankosi, kokiais socialiniais kanalais naudojasi ir net kuriuos mygtukus puslapyje jis paspaudžia.
- Optimizuoti klientų lojalumą. Duomenys gali suteikti išvalgų apie tai kokie yra klientai, kur jie yra, ko jie nori, kaip dažnai jie perka, kada ir kaip jie nori, kad su jais būtų susisiekti, ir daug kitų svarbių veiksnių. Įmonės taip pat gali analizuoti kaip vartotojai sąveikauja su jų interneto svetaine, ar net fizine parduotuve, taip pagerinant vartotojo patirtį.
- Padidinti klientų lojalumą bei juos išlaikyti. Daugelis mažmeninių įmonių yra įdiegę lojalumo kortelės sistemas tam, kad galėtų sekti kliento pirkimus, tačiau šios sistemos taip

pat gali padėti stebėti, koks stimulus, paskatos ir kurios akcijos yra pačios efektyviausios skatinant vartotoją ar tam tikrą tikslinę vartotojų grupę pirkti.

- Marketingo optimizavimas. Naudojantis tradicinėmis reklamos priemonėmis sunku sekti jų poveikį vartotojų sprendimams ir investicijų grąžą. Didieji duomenys ir jų analitika gali padėti įmonėms priimti optimalius marketingo sprendimus daugybėje skirtingų kanalų ir nuolat optimizuoti marketingo pastangas atliekant testavimus, matavimus bei analizę.
- Biudžeto valdymas. Kadangi didieji duomenys leidžia įmonėms stebėti ir optimizuoti savo vykdomas marketingo kampanijas, vadinasi jie gali geriau paskirstyti savo rinkodaros biudžetą aukščiausiai investicinei grąžai.
- Personalizavimas realiu laiku. Didieji duomenys kartu su „machine learning“ algoritmais suteikia galimybę individualizuoti savo pasiūlymus individualiems klientams realiu laiku. Įmonės gali individualizuoti kokius produktus ir akcijas konkretus klientas matys, pavyzdžiui, atsiųsti individualų pasiūlymą į vartotojo telefono numerį kai tik šis atsiduria tam tikroje apibrėžtoje fizinėje vietoje.
- Prognozavimas. Didžiųjų duomenų analitika gali būti taikoma, siekiant sukurti modelį, kuris gali sėkmingai prognozuoti vartotojo elgseną ir pardavimus.
- Rinkos tyrimai. Įmonės gali atlikti kokybinius ir kiekybinius rinkos tyrimus daug greičiau ir pigiau nei bet kada anksčiau. Internetinės apklausos, kurios padeda išsiaiškinti tikslinės rinkos klientų nuomones ir atsiliepimus, kurie gali būti lengviau išanalizuoti analitikų.
- Reputacijos valdymas. Pasinaudodamos didžiais duomenimis, įmonės gali lengvai stebėti savo prekinio ženklo vardo paminėjimą daugelyje įvairių svetainių, per socialinius kanalus rasti nefiltruotas nuomones ir atsiliepimus apie pačią įmonę ir produktus. Taip pat galima pasinaudoti socialine žiniasklaida teikiant klientų aptarnavimo paslaugas ir taip didinant prekinio ženklo patikimumą.
- Konkurentų analizė. Nauji socialinės stebėsenos įrankiai leidžia lengvai rinkti ir analizuoti duomenis apie konkurentus ir jų marketingo pastangas. Įmonės, kurios naudoja šią informaciją be abejonės turės papildomą konkurencinį pranašumą.

Didieji duomenys dažniausiai apibrėžiami kaip itin didelės apimties, įvairovės ir greičio duomenys. Įvairūs šaltiniai siūlo įvesti ir ketvirtą dimensiją - teisingumo. Dėl didelės duomenų apimties ir įvairovės sunku identifikuoti naudingus ir tikslus duomenis iš neišvalytų duomenų (angl. dirty data), todėl duomenų teisingumas viena sudėtingiausių užduočių. Didžiausia problema yra tai, kad neišvalyti duomenys gali lemti dideles klaidas ir neteisingus rezultatus, kurie gali turėti įtakos ir greičio dimensijai. Taigi, toliau pateikiama didžiųjų duomenų privalumų ir trūkumų lentelė (žr. 1.3.2 lent.) [15].

1.3.2 lentelė. Didžiųjų duomenų panaudojimo trūkumai ir privalumai (Stoicesu, 2015)

Privalumai	Trūkumai
Apimtis	
Fiksuojant visą vartotojų veiklą, galima susidaryti išsamų vartotojo paveikslą, kuris padėtų suprasti jo elgseną bei pagerinti vartotojo pirkimo patirtį.	Kuo didesnė duomenų apimtis, tuo sudėtingesnis jų apdorojimas
Ivairovė	
Duomenys iš įvairių šaltinių padeda identifikuoti paslėptus modelius ir netikėtas koreliacijas	Grėsmė privatumui ir asmens duomenų saugumui
Greitis	
Realaus laiko duomenys padeda išspręsti realaus laiko problemas	Dažnai nėra aišku kas kaupia duomenis ir kaip juos naudoja
Teisingumas	
Išvalyti ir tikslūs duomenys gali būti panaudojami siekiant išgauti naują informaciją	Netikslūs duomenys gali lemti neteisingas išvadas ir negiamus rezultatus

Klasikinės matavimo sistemos buvo skirtos analizuoti vartotojų elgesio ketinimus, o ne faktinį elgesį, kadangi tokia informacija buvo paprastai gaunama. Iki šiol, rinkti duomenis apie realų vartotojų elgesį buvo nepraktiška, tačiau atsiradusios socialinės žiniasklaidos priemonės radikaliai pakeitė turimus vartotojų elgsenos duomenis. Elektroninės prekybos svetainės pakeičia realias parduotuves ir fiksuoja kiekvieną vartotojo atliktą žingsnį. Įprasti telefono pokalbiai su draugais apie naujausius pirkinius pakeičiami Twitter'io žinutėmis, kurios gali būti analizuojamos visų, kas seka vartotoją toje paskyroje. Iš tiesų, visos technologijos sukurtos šiomis dienomis (telefonai, kompiuteriai, automobiliai, šaldytuvai) gamina terabaitus ir petabaitus duomenų. Informacija yra išgaunama iš visur, iš automobilių stovėjimo aikštelės, interneto paieškų, Facebook'o, išmaniųjų telefonų, todėl kiekvienu savo veiksmu vartotojai palieka skaitmeninį pėdsaką, kuris gali būti įrašomas, saugomas ir po to analizuojamas. Tuo metu kai vartotojai gali pasakyti ką jie mano, duomenų tyrėjai analizuodami vartotojų elgsenos duomenis, gali nustatyti kaip elgiamasi iš tikrųjų [15].

Smith'as (2016) teigia, jog motyvus veikiančius vartotojų elgseną geriausiai galima suprasti naudojantis mokslinių tyrimų ir duomenų analizės deriniu, pavyzdžiui [18]:

- Vartotojų atsiliepimai. Padeda išryškinti bendras problemas ir vartotojų norus.
- Klausimų - atsakymų (Q&A) svetainės. Tokios svetainės gali suteikti įžvalgas apie vartotojams labiausiai rūpimus klausimus ir problemas, susijusias su preke, paslauga ar produkto prekiniu ženklu.
- Apklausos. Interaktyvios apklausos padeda greitai sužinoti atsakymus į konkrečius klausimus.
- Raktažodžio (angl. keyword) paieška gali padėti sužinoti kas konkrečiai domina vartotoją, jo susidomėjimo lygį bei kalbą, kurią vartotojas naudoja.
- Verslo klasės žiniatinklio analizė - Google Analytics. Ši analizė padeda nustatyti lankytojų srautą, interesus, demografinę padėtį.

- Konkurentų analizės įrankiai gali suteikti naudingą informaciją apie vartotojus, kurie naudojami konkurentų siūlomomis paslaugomis.
- Tinklaraščiai (angl. blogs). Pasinaudojant tinklaraščiais galima sužinoti vartotojų nuomones apie konkrečias prekes ar įmonės prekinį ženklą.
- Socialiniai tinklai. Kiekvieną dieną daugybė žmonių viešai rodo savo gyvenimą socialiniuose tinkluose, tokia informacija, apdorojant ją tam tikrais metodais gali būti labai naudinga aiškinantis vartotojų elgseną ar jos motyvus.

Iš tiesų vis daugiau didžiųjų duomenų ir jų analitikos priemonių tampa vis lengviau prieinamos ne tik patiems mažiausiems verslams, bet ir pradedantiesiems. Tai puikios priemonės bendrovėms pagerinti marketingo sprendimus ir pasiekti vartotojus naujais ir inovatyviais metodais. Didieji duomenys naudojant juos tinkamame kontekste reiškia naujus santykius su vartotojais, naujas strategijas, naujas galimybes, geresnius sprendimus ir didelį konkurencinį pranašumą. tinkluose

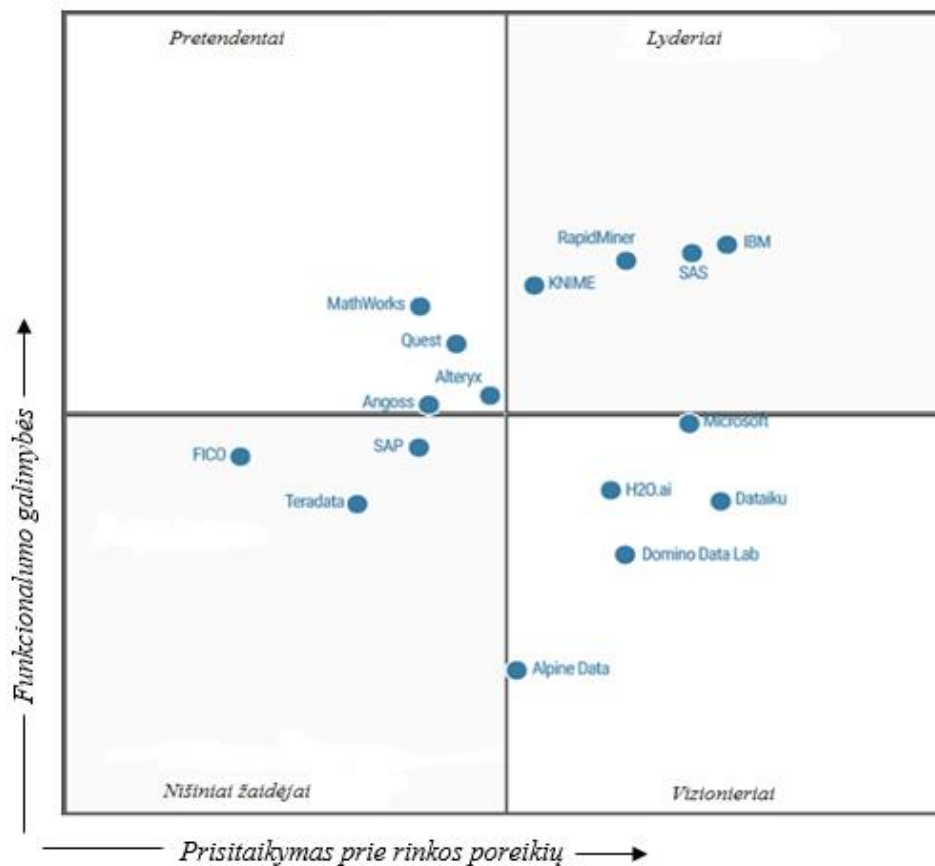
1.4. Duomenų tyrybos sistemų apžvalga

Norint atlikti vartotojų elgsenos tyrimus itin svarbu tinkamai ir teisingai apdoroti didžiuosius duomenis, todėl reikalinga išmanyti duomenų tyrybos sistemas, programinę įrangą, jos panaudojimo galimybes, privalumus bei trūkumus. Todėl šiame skyriuje bus apžvelgiamos pagrindinės duomenų tyrybos sistemos, remiantis Gartner atlikta analize [18]. Gartner yra JAV tyrimų ir konsultavimo įmonė, teikianti informacinių technologijų įžvalgas IT ir kitiems verslo lyderiams visame pasaulyje. 2017 m. Gartner paskelbtoje ataskaitoje buvo vertinamos 16 geriausių duomenų mokslo platformų. Nors atviro kodo platformos kaip Python ir R vaidina svarbų vaidmenį duomenų mokslo srityje į Gartner'io tyrimų metodologiją jos nebuvo įtrauktos, kadangi šioje ataskaitoje vertinami tik komercinių pardavėjų duomenų tyrybos įrankiai. Gartner'io verslo analitikos ir analitikos platformų magiškajame kvadrate pateikta 16 geriausių duomenų mokslo įmonių, įvertinus jas pagal duomenų vizualizacijos ir savitarnos analitikos įrankius.

Kaip matome pagal 1.4.1 pav. lyderių pozicijas užima IBM, SAS, RapidMiner, KNIME, pretendentų – MathWorks, Quest (anksčiau Dell), Alteryx, Angoss, vizionierių - Microsoft, H2O.ai, Dataiku, Domino Data Lab, Alpine Data, o nišinių žaidėjų - FICO, SAP, Teradata. Svarbu paminėti, jog net mažiausiai balų surinkusios įmonės yra tarp 16 geriausių iš 100 esančių duomenų mokslo įmonių visame pasaulyje.

Toliau pateikiamas 2016 - 2017 m. magiškujų kvadratų rezultatų palyginimus (žr. 1.4.2 pav.). Žalios linijos nurodo, kad įmonės pozicija žymiai pagerėjo lyginant su 2016 m. Tuo tarpu raudona linija pavaizduota prastesnė įmonės padėtis lyginant su praėjusiais metais. Žaliu apskritimu

pažymėtos naujai Gartner'io magiškajame kvadrate atsiradusios įmonės, o raudonai - jau išnykusios.



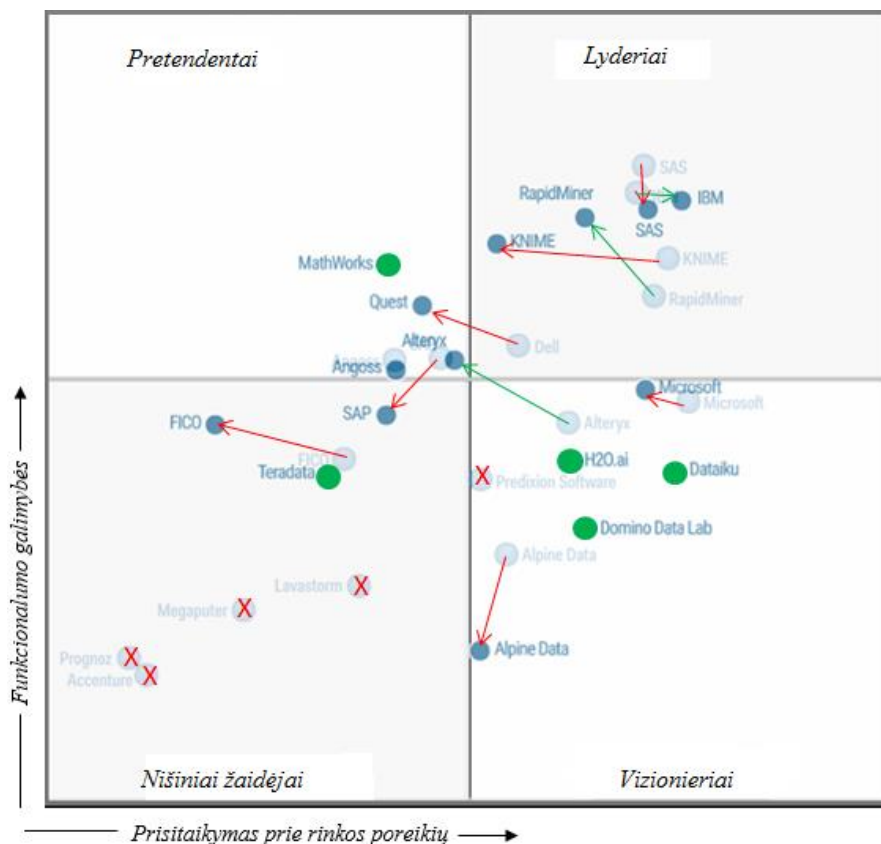
1.4.1 pav. Duomenų mokslo platformų Gartner'io magiškasis kvadratas, 2017m. (Piatetsky, 2017)

Kaip matome, lyderiaujančias pozicijas 2017 m. užima tos pačios 4 kompanijos - IBM, SAS, RapidMiner ir KNIME kaip ir 2014-2016 m. laikotarpiu. Dell Software, kuri 2016 m. buvo parduota ir pervadinta į Quest, persikėlė iš lyderių į pretendentų ketvirtį. IBM sugebėjo šiek tiek labiau prisitaikyti prie rinkos poreikių, o SAS prarado šiek tiek funkcionalumo. Tuo tarpu RapidMiner tapo funkcionalesnė, tačiau kartu su KNIME ne taip sparčiai gebėjo prisitaikyti prie augančių rinkos poreikių.

Lyderių ketvirtis. Pagal 1.4.2 pav. galime pastebėti, jog *IBM* vėl lyderis, dėl savo SPSS Modeler ir SPSS Statistics įrankių. Gartner nevertino naujos IBM Data Science Experience platformos, tačiau tai prisidėjo prie IBM teigiamų pokyčių palei prisitaikymo prie rinkos poreikių ašies. IBM privalumai apima didžiulę klientų bazę ir toliau plėtojamas duomenų tyrybos ir „machine learning“ galimybės. Svarbu paminėti, jog IBM platus duomenų tyrybos įrankių pasirinkimas (SPSS, IBM Watson DSX, Cognos) sukuria tam tikrą painiavą rinkoje, kuri ypač atsiliepia SPSS linijai.

SAS turi daug programinės įrangos produktų skirtų duomenų analizės mokslui. Gartner įvertino SAS Enterprise Miner (EM) ir SAS Visual Analytics Suite (VAS) paketus. Šiuo metu

SAS daugiau dėmesio skiria interaktyviam modeliavimui su VAS, bet ir toliau palaiko savo tradicinę SAS bazę. SAS ir toliau išlaiko tvirtą poziciją tarp lyderių, bet maišatis tarp kelių SAS produktų ir sąlyginai aukštos kainos sumažino SAS funkcionalumo galimybes.



1.4.2 pav. Duomenų mokslo platformų Gartner'io magišksasis kvadratas, 2016-2017m. (Piatetsky, 2017)

KNIME siūlo atviro kodo KNIME Analytics platformą su stipriu funkcionalumu pažengusiems duomenų analizės specialistams. Šis įrankis itin stiprus keliose srityse, ypač gamybos ir gamtos moksluose. KNIME prarado savo pozicijas prisitaiko prie rinkos poreikių dimensijoje dėl silpnos reklamos ir inovacijų lyginant su kitomis lyderėmis.

RapidMiner siūlo GUI pagrįstą duomenų mokslo platformą, tinkančią tiek pradedantiesiems, tiek pažengusiems. Be to, ji taip pat suteikia atviro kodo prieigą. RapidMiner yra prieinama tiek nemokama versija, tiek komercinio leidimo (su papildomu funkcionalumu dideliems duomenų kiekiam bei galimybe prijungti daugiau duomenų šaltinių). RapidMiner yra lyderių ketvirtyje dėl savo rinkos padėties ir gerai išvystyto produkto.

Pretendentų ketvirtis. MATHWORKS ir jo gerai žinomas duomenų tyrybos įrankis MATLAB - itin populiarus tarp inžinierių dėl savo didelės tyrybos įrankių įvairovės.

Quest, kuri buvo parduota privataus kapitalo įmonei, dabar parduoda Statistica platformą. Quest yra pretendentų pozicijoje dėl trukumų susijusių su debesų kompiuterija.

Tuo tarpu, Alteryx siūlo paprastai naudojamą duomenų mokslo platformą, su duomenų paruošimo savitarna ir pažangia analitika. Svarbu paminėti, jog buvo prijungtos modeliavimo ir optimizavimo galimybės. Palyginti su 2016 m, ši įmonė persikėlė iš vizionierių į pretendentes dėl išaugusio vartotojų skaičiaus.

Angoss suteikia puikius prognozavimo optimizavimo ir analizės įrankius. Angoss išliko beveik toje pačioje pozicijoje kaip ir 2016 m.

Vizionierių ketvirtis. Microsoft vertinimas buvo pagrįstas Azure Machine Learning platforma ir dalimi Microsoft Cortana Cortana Intelligence Suite, kurie siūlo stiprią debesų kompiuterija grįstą duomenų analizės platformą. Microsoft išliko vizionierių ketvirtyje ir 2017m.

H2O.ai siūlo atviro kodo duomenų mokslo platformą su greito vykdymo apmokymo (angl. deep learning) ir kitais „machine learning“ metodais.

Dataiku buvo įtrauktas į vizionierių ketvirtį dėl savo novatoriško pobūdžio, atvirumo, suderinamumo funkcijų ir tinkamumo įvairaus lygio specialistams.

Domino Data Lab pagrindinis dėmesys skiriamas bendradarbiavimui ir plačiam atviro kodo technologijų palaikymui.

Tuo tarpu Alpine Data siūlo bendradarbiavimo aplinką, kuri įgalina verslo lygio analitikos specialistus analizuoti duomenis. Palyginti su 2016 Alpine Data liko vizionierių ketvirtyje, tačiau sumažėjo jos gebėjimas vykdyti dėl nuolatinės kovos įgyti savo rinkos dalį.

Nišinių žaidėjų ketvirtis. SAP sumažėjęs gebėjimas vykdyti perkėlė jį iš konkurentų į nišinių žaidėjų ketvirtį, kadangi trūksta Spark integracijos, paramos atviro kodo bei Python platformai.

Fico Decision Management Suite siūlo daugybę analitinių įrankių. Ji išliko nišinių žaidėjų ketvirtyje.

Teradata siūlo 3 sluoksnių Aster Analytics platformą - analizės sistemų, analizės funkcijų ir Aster AppCenter, skirtą analizei ir prisijungimui prie išorinių įrankių. Dėl sąlyginai mažo susidomėjimo - tai nišinis žaidėjas.

2. TYRYMO METODAI

2.1. Aprašomoji statistika

Kiekviena duomenų analizė pirmiausia prasideda nuo susistemintos duomenų informacijos, tai yra aprašomosios statistikos. Aprašomoji statistika, naudojama aprašyti pagrindines analizuojamų duomenų charakteristikas bei pavaizduoti jas grafikuose.

Šiame darbe kiekybiniais duomenims aprašyti buvo naudotos imties pločio, sumos, vidurkio, mažiausios bei didžiausios reikšmės, standartinio nuokrypio, asimetrijos ir eksceso koeficientų charakteristikos. Toliau pateikiamos jų formulės ir paaiškinimai [20].

Imties plotis (r) - tai skirtumas tarp didžiausios ir mažiausios imties reikšmės. Šis dydis apskaičiuojamas pagal formulę (1):

$$r = x_{max} - x_{min}, \quad (1)$$

čia x_{max} – didžiausia imties reikšmė;

x_{min} – mažiausia imties reikšmė.

Suma – bendras visų imties elementų kiekis, apskaičiuojamas pagal (2) formulę.

$$\sum_{i=1}^n x_i, \quad (2)$$

Imties vidurkis (\bar{x}) - visų imties elementų suma, padalinta iš visų imties elementų kiekio. Toliau pateikta imties vidurkio formulė (3).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3)$$

čia n – imties elementų kiekis;

x_i – i -oji kintamojo reikšmė.

Standartinis nuokrypis (s_w) - tai dažniausiai praktikoje taikomas sklaidos matas, kuris nusako atsitiktinio dydžio įgyjamų reikšmių sklaidą apie vidurkį (4) [20].

$$s_w = \sqrt{\frac{1}{d} \sum_{i=1}^n w_i (x_i - \bar{x}_w)^2}, \quad (4)$$

čia x_i – i -oji kintamojo reikšmė;

\bar{x}_w - imties vidurkis;

w_i - svoris susietas su i -ąja kintamojo reikšme;

d – svorių suma, apskaičiuojama (5):

$$d = \sum_{i=1}^n w_i, \quad (5)$$

Imties asimetrijos koeficientas (A_s), skirtas apibūdinti duomenų pasiskirstymo asimetriškumą. Jei asimetrijos koeficientas arti 0 – skirstinys simetriškas vidurkio atžvilgiu. Jei asimetrijos koeficientas teigiamas – fiksuojama dešinioji skirstinio asimetrija, jei neigiamas – kairioji [21].

$$A_s = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n w_i^2 \left(\frac{x_i - \bar{x}_w}{s_w} \right)^3, \quad (6)$$

čia n – imties elementų kiekis;

x_i – i -oji kintamojo reikšmė;

\bar{x}_w – imties vidurkis;

s_w – imties standartinis nuokrypis;

w_i – svoris susietas su i -ąją kintamojo reikšme.

Imties eksceso koeficientas skirtas apibūdinti empirinio skirstinio tankio smailumą ($E_k > 0$) arba lėkštumą ($E_k < 0$). Normaliojo skirstinio atveju eksceso koeficiento reikšmė lygi 0 [20].

$$E_k = \frac{1}{n} \sum_{i=1}^n w_i^2 \left(\frac{x_i - \bar{x}_w}{s_w} \right)^4 - 3, \quad (7)$$

čia n – imties elementų kiekis;

x_i – i -oji kintamojo reikšmė;

\bar{x}_w – imties vidurkis;

s_w – imties standartinis nuokrypis;

w_i – svoris susietas su i -ąją kintamojo reikšme

2.2. Suderinamumo hipotezių tikrinimo metodai

Norint palyginti kelis empirinius ar empirinį su teoriniu skirstinius yra tikrinamos suderinamumo hipotezės. Egzistuoja daug skirtingų būdų ir statistinių kriterijų, kaip patikrinti skirtingas suderinamumo hipotezes (Chi kvadrato (χ^2), Kolmogorov-Smirnov D, Anderson-Darling, Cramer V statistiniai kriterijai ir kt.). Suderinamumo hipotezės susideda iš nulinės ir alternatyviosios hipotezių [21]:

Nulinė hipotezė (H_0): imtys yra pasiskirsčiusios pagal normalųjį skirstinį.

Alternatyvioji hipotezė (H_a): imtys nėra pasiskirsčiusios pagal normalųjį skirstinį.

Taip pat svarbu nusistatyti reikšmingumo lygmenį (α), tai yra tikimybę pagrįstai atmesti nulinę hipotezę (H_0), kai ji teisinga. Dažniausiai reikšmingumo lygmeniu pasirenkami 5% (0,05) arba 1% (0,01). Jei p reikšmė $\geq \alpha$ reikšmingumo lygmenį, tai nulinė hipotezė (H_0) apie imties duomenų pasiskirstymą pagal normalųjį skirstinį yra priimama. Jei p reikšmė $\leq \alpha$, tai alternatyviają hipotezę (H_a) priimame, o nulinę (H_0) – atmetame [21].

Tikrinant suderinamumo hipotezes dažniausiai naudojamas Kolmogorov-Smirnov statistinis D kriterijus, kuris įvertina didžiausią atstumą tarp empirinio ir teorinių skirstinių. Jo formulė pateikiama toliau [21]:

$$D = \sqrt{n} \max |F_n(x) - F(x)|, \quad (8)$$

čia n – imties elementų kiekis;

$F_n(x)$ - empirinio skirstinio funkcija;

$F(x)$ - teorinio skirstinio funkcija.

2.3. Koreliacinė analizė

Siekiant nustatyti tiriamų kintamųjų priklausomumą ir tarpusavio ryšio stiprumą, reikalinga atlikti koreliacinę analizę. Kintamųjų tarpusavio priklausomybę išmatuoti padeda koreliacijos koeficientas. Jei analizuojami kintamieji yra pasiskirstę pagal normalųjį skirstinį koreliacinei analizei atlikti skaičiuojamas Pearson'o koreliacijos koeficientas. Jei kintamieji nėra pasiskirstę pagal normalųjį skirstinį – taikomas Spearman'o koreliacijos koeficientas [21].

Pearson'o koreliacinis matas yra skirtas įvertinti tarp kintamųjų esantį tiesinio ryšio stiprumą. Jis gali būti naudojamas tik intervaliniams ir santykiniais kintamiesiems. Šis koeficientas apskaičiuojamas pagal (9) formulę [22]:

$$r_{xy} = \frac{\sum_i ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}, \quad (9)$$

čia \bar{x} - x imties vidurkis;

\bar{y} - y imties vidurkis;

Spearman'o ranginės koreliacijos koeficientas, skirtas kintamųjų išmatuotų santykių, intervalų ir tvarkos skalėse ryšio stiprumui įvertinti [22].

$$\theta = \frac{\sum_i ((R_i - \bar{R})(S_i - \bar{S}))}{\sqrt{\sum_i (R_i - \bar{R})^2 \sum_i (S_i - \bar{S})^2}}, \quad (10)$$

čia R_i – stebimų x_i reikšmių rangas;

S_i – stebimų y_i reikšmių rangas;

\bar{R} – R_i imties vidurkis; \bar{S} – S_i imties vidurkis.

Tiek Pearson'o, tiek Spearman'o koreliacijos koeficientai gali įgyti reikšmes nuo -1 iki 1. Kuo ryšys tarp kintamųjų stipresnis, tuo koreliacijos koeficiento reikšmė ($|r|$) artimesnė 1. Toliau pateikiama koreliacijos koeficiento reikšmių vertinimo lentelė (žr. 2.3.1 lent.) [23].

2.3.1 lent. koreliacijos koeficiento reikšmių vertinimas

Nėra ryšio	Labai silpna	Silpna	Vidutinė	Stipri	Labai stipri
0	0 - 0,19	0,19 – 0,39	0,40 – 0,69	0,70 – 0,89	0,9 - 1

Hipotezės apie koreliacijos koeficientų reikšmingumą patikrinimui, naudojama Stjudento charakteristika (t), ji apskaičiuojama pagal toliau pateiktą formulę [22]:

$$t = (n-2)^{\frac{1}{2}} \left(\frac{r^2}{1-r^2} \right)^{\frac{1}{2}}, \quad (11)$$

čia n – imties elementų kiekis;

r – koreliacijos koeficientas.

Jei apskaičiuota Stjudento reikšmė (t) $< \alpha$ reikšmingumo lygmenį, tai ryšį tarp kintamųjų laikome statistiškai reikšmingu. Jei $t > \alpha$, tai koreliacija nėra statistiškai reikšminga.

2.4. Dispersinė analizė

Dispersinė analizė – tai tyrimas skirtas ištirti atsitiktinių dydžių skirstinių priklausomybes nuo konkrečių faktorių. Faktoriai gali būti tiek kiekybiniai, tiek kokybiniai. Vienas dispersinės analizės tikslų – ištirti ar priklausomo kintamojo Y , išmatuoto skirtingose populiacijose, vidurkiai skiriasi. Dispersinė analizė gali būti vienfaktorinė (*ANOVA*), kai tiriamas vieno nepriklausomojo kintamojo poveikis, dvifaktorinė ir daugiafaktorinė (*MANOVA*), kai tiriamas atitinkamai dviejų ar daugiau nepriklausomų kintamųjų poveikis priklausomam kintamajam [21].

Jei dispersinės analizės tiesinio modelio prielaidos (kintamųjų pasiskirstymas pagal normalųjį skirstinį, dispersijų lygybė, kintamųjų tarpusavio nepriklausomumas) nėra tenkinamos, taikoma ranginė vienfaktorinė dispersinė analizė. Svarbu patikrinti vienfaktorinės dispersinės analizės hipotezes [21]:

H_0 : visi tiriamų kintamųjų vidurkiai yra lygūs.

H_a : tiriamų kintamųjų vidurkiai nėra lygūs.

Šių hipotezių tikrinimui naudojama Kruskal-Wallis'o statistika (C), kuri apskaičiuojama pagal toliau pateiktą formulę [24]:

$$C = \left(\frac{\sum_{i=1}^r (T_i - E_0[T_i])^2}{n_i} \right) / S^2, \quad (12)$$

čia T_i - i -ojo lygmens rangų suma;

$E_0[T_i]$ - tikėtina i -ojo lygmens rangų suma;

n - imties stebinių skaičius;

S^2 - imties rangų dispersija.

i -ojo lygmens rangų suma (T_i) apskaičiuojama pagal šią formulę [24]:

$$T_i = \sum_{j=1}^n c_{ij} a(R_j), \quad (13)$$

čia $a(R_j)$ - j -ojo stebinio rangas;

c_{ij} - indikatorius, parodantis ar j -asis stebinys yra veikiamas i -ojo lygmens.

Tikėtina i -ojo lygmens rangų suma kai teisinga nulinė hipotezė (H_0), apskaičiuojama pagal (14) formulę [24]:

$$E_0[T_i] = \frac{n_i}{n} \sum_{j=1}^n a(R_j), \quad (14)$$

Imties rangų dispersija randama pagal (15) formulę [24]:

$$S^2 = \frac{1}{(n-1)} \sum_{j=1}^n (a(R_j) - \bar{a})^2, \quad (15)$$

čia \bar{a} - imties rangų vidurkis, apskaičiuojamas pagal žemiau pateiktą (16) formulę [24]:

$$\bar{a} = \frac{1}{n} \sum_{j=1}^n a(R_j), \quad (16)$$

2.5. Požymių priklausomumo lentelių tyrimas

Požymių priklausomumo lentelių tyrimas yra skirtas įvertinti priklausomybę tarp vardinių ir rangų skalės kintamųjų. Priklausomybės laipsniui tarp kintamųjų įvertinti dažniausiai iš neparametrinių kriterijų taikomas Chi-kvadrato (χ^2) kriterijus, kuris naudojamas hipotezėms apie kintamojo skirstinį populiacijoje tikrinti. Šis kriterijus parodo, ar skirtumas tarp empirinio ir teorinio skirstinių yra reikšmingas, tai reiškia, jog tikrinama, ar turimas empirinis skirstinys yra suderintas su teoriniu modeliu. 2*2 požymių priklausomumo lentelėse reikia vadovautis Chi-kvadratu su Yates'o pataisa. Taigi tikrinama požymių tarpusavio nepriklausomumo hipotezė [21]:

H_0 : požymiai tarpusavyje nėra statistiškai reikšmingai priklausomi.

H_a : požymiai tarpusavyje yra statistiškai reikšmingai priklausomi.

Nulinė (H_0) kintamųjų požymių nepriklausomumo hipotezė atmetama ir alternatyvioji hipotezė (H_a) priimama kai Chi kvadrato p reikšmė mažesnė už α reikšmingumo lygmenį. Jei Chi kvadrato p reikšmė $> \alpha$ reikšmingumo lygmenį, nulinę hipotezę (H_0) priimame, o alternatyviają (H_a) atmetame [21].

Iš tiesų, priklausomybės laipsniui tarp kintamųjų įvertinti yra didelė charakteristikų įvairovė. Jei duomenys aprašomi 2×2 dažnių lentelė ir kai nors vienas tikėtinas stebėjimų skaičius mažiau penkių, papildomai skaičiuojamas tikslus Fisher'io kriterijus. Matuojamiems pagal intervalų skalę kintamiesiems yra skaičiuojamas Pearson'o koreliacijos koeficientas. Kuomet stebimi kategoriniai kintamieji matuojami pagal vardinę skalę, naudojami Phi, Kontingencijos ir Kramerio V ryšio stiprumo matai. Kuo šie matai didesni, tuo požymių priklausomybė yra stipresnė [21].

Phi (φ) koeficientas skaičiuojamas Chi kvadrato (χ^2) pagrindu eliminuojant imties dydžio įtaką ir taikomas binariniams kintamiesiems, tai yra tuomet, kai duomenys aprašomi 2×2 kontingencijos lentelėmis. Iš tiesų Phi koeficiento reikšmė priklauso nuo lentelės dydžio ir gali viršyti 1 [21].

$$\varphi = \sqrt{\chi^2/n}, \quad (17)$$

Kontingencijos koeficientas (C) yra ryšio matas $I \times J$ dažnių lentelėms. Tai Phi koeficiento modifikacija, skirta didesnėms kontingencijos lentelėms. Įprastai šis koeficientas taikomas 5×5 ir didesnėms požymių priklausomumo lentelėms. Kontingencijos koeficiento reikšmė priklauso nuo eilučių ir stulpelių skaičiaus lentelėje ir neviršija 1 [21].

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}}, \quad 0 \leq C \leq \sqrt{\frac{\min(I,J)-1}{\min(I,J)}}, \quad (18)$$

Kramerio V koeficientas (V) yra dažniausiai naudojamas vardų skalėje išmatuotų kintamųjų ryšio matas, kuris taipogi apskaičiuojamas χ^2 pagrindu. Šio koeficiento dydis nepriklauso nuo lentelės dydžio, tuomet kai eilučių skaičius lygus stulpelių skaičiui. Keturlaukių, tai yra 2×2 , lentelių atveju, Kramerio V ir Phi koeficientai sutampa [21].

$$V = \frac{\varphi}{\sqrt{\min(I-1, J-1)}} \quad 0 \leq V \leq 1, \quad (19)$$

2.6. Logistinė regresija

Regresijos modelis yra skirtas vieno kintamojo priklausomybei nuo kitų kintamųjų aprašyti, tai yra norint išsiaiškinti nuo ko priklauso tiriamas reiškinys. Logistinė regresija nuo tiesinės regresijos skiriasi tuo, jog priklausomas kintamasis, kurio elgesys ir yra modeliuojamas, yra kategorinis arba ranginis, tačiau ne intervalinis [25].

Kintamasis Y vadinamas priklausomu (arba regresuojamu) kintamuoju, kintamieji X , Z , W vadinami nepriklausomais (aiškinamaisiais) kintamaisiais arba regresoriais. Matematinis modelis sudaromas ne pačiam priklausomam kintamajam, o jo tikimybių santykio logaritmui [25]:

$$Z(x) = \ln \frac{P(Y=1)}{P(Y=0)}, \quad (20)$$

čia $Z(x) = c + \beta_1 x_1 + \dots + \beta_k x_k$.

Jeigu koeficientas prie kurio nors kintamojo yra teigiamas, tai šiam kintamajam didėjant, $Y=1$ įvykio tikimybė didėja. Jeigu koeficientas neigiamas, tai atitinkamam kintamajam didėjant, tikimybė Y įgyti vieneto reikšmę mažėja. Koeficiento β_k eksponentė dar vadinama galimybių santykiu (angl. odds ratio). Galimybių santykis parodo, kaip keičiasi $Y=1$ galimybė, kai x_k padidėja vienetu, o kiti x nekinta [25].

Sudaryto modelio tinkamumą duomenims parodo [26]:

- Didžiausio tikėtimumo santykio Chi-kvadrato statistika. Šis kriterijus parodo ar sudarytame modelyje yra bent vienas reikalingas regresorius. Jeigu statistikos p reikšmė $\geq \alpha$ reikšmingumo lygmenį tuomet regresijos modelis nėra tinkamas. Jeigu $p < \alpha$, tai sudarytas modelis duomenims tinka.
- Voldo kriterijus regresoriams, padedantis nuspręsti ar kintamąjį reikia pašalinti iš sudaryto modelio. Jeigu regresoriui Voldo testo p reikšmė $< \alpha$, tai regresorius yra statistiškai reikšmingas ir modelyje paliekamas. Jeigu p reikšmė $\geq 0,05$, tuomet regresorius yra statistiškai nereikšmingas ir dažniausiai iš modelio pašalinamas.

$$W = \frac{\hat{\theta} - \theta_0}{S_e(\hat{\theta})}, \quad (21)$$

čia $\hat{\theta}$ – didžiausias tikėtimumo santykis

S_e – standartinis $\hat{\theta}$ nuokrypis.

- Determinacijos koeficientai, parodo bendrąjį modelio tikimą duomenims. Tai svarbiausia modelio tikimo duomenims charakteristika. Determinacijos koeficientas lygina skirtumus tarp Y reikšmių, kai atsižvelgiama į regresijos modelį, su skirtumais tarp Y reikšmių, kai į modelį neatsižvelgiama. Šis koeficientas gali įgyti reikšmes nuo 0 iki 1. Kuo koeficiento reikšmė didesnė, tuo sudarytas modelis geriau aprašo duomenims [27].

2.7. Faktorinė analizė

Faktorinė analizė – taikoma dideliame kintamųjų skaičiui sumažinti, juos pakeičiant tiesiogiai nestebimais faktoriais. Šis metodas plačiai taikomas šiuolaikinėje psichologijoje, socialiniuose bei humanitariniuose moksluose [21].

Dažnai išskirtus faktorius yra sunku tinkamai interpretuoti, nes vienas kintamasis yra reikšmingai vienodai susijęs su keletu faktorių. Šios problemos sprendimo būdas – faktorių

sukimas, kuris reiškia faktorių svorių matricos transformavimą į lengviau interpretuojamą pavidalą. Faktorių ašių sukimas gali būti ortogonalus (išlaikant statų kampą) ir neortogonalus (stataus kampo neišlaikant). Status kampas tarp faktorių ašių yra būtinas jei norima, kad faktoriai būtų nekoreliuoti tarpusavyje. Praktikoje dažniausiai naudojami toliau iššvardyti ortogonalieji sukimai [21]:

1. Varimax. Minimizuojamas su kiekvienu faktoriumi reikšmingai susijusių kintamųjų skaičius.
2. Quartimax. Minimizuojamas kiekvieną kintamąjį aprašančių faktorių skaičius.
3. Equamax. Kombinuotas Varimax ir Quartimax metodas. Minimizuojamas kiekvieną kintamąjį aprašančių faktorių skaičius, bei su kiekvienu faktoriumi reikšmingai susijusių kintamųjų skaičius.

Faktorinės analizės metodus galime suskirstyti į dvi grupes[21]:

- Pagrindinių faktorių, minimaliųjų liekanų, didžiausio tikėtimumo metodai - kai tam tikra stebėtų kintamųjų dispersijos dalis gali būti visiškai paaiškinta išskirtais faktoriais, o likusi dispersijos dalis lieka nepaaiškinta dėl kintamųjų specifiškumo.
- Pagrindinių komponenčių metodas - kai išskirti faktoriai paaiškina visą stebimų kintamųjų dispersiją.

Toliau analizuojamas pagrindinių komponenčių metodas. Pagrindinę komponentę galime išreikšti tokia stebimų kintamųjų tiesine išraiška [21]:

$$C_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{in}X_n, \quad (22)$$

čia C_i - išskirta i -oji komponentė;

a_{ij} - i -osios komponentės svoris j -ojo kintamojo atžvilgiu (kintamojo j ir komponentės i koreliacijos koeficientas);

X_j - kintamojo j reikšmės.

Atliekant pagrindinių komponenčių analizę reikia rasti svorius a_{ij} ir nustatyti, kiek pradinių kintamųjų dispersijos paaiškina kiekviena komponentė C_i .

Taigi, pirmiausia išskiriame pirmąją komponentę C_1 , kuri paaiškina didžiausią dalį pradinių kintamųjų X_j ($j = 1, \dots, n$) dispersijos ir apskaičiuojame jos svorių vektorių $b_1 = (b_{11}, b_{12}, \dots, b_{1n})^T$ atitinkančią tikrinę reikšmę λ_1 (angl. eigenvalue), parodančią paaiškinamos dispersijos dalį ($DC_1 = \lambda_1$). Tuomet ieškome antrosios komponentės C_2 , kuri paaiškina didžiausią dalį likusios dispersijos ir yra nekoreliuota su pirmąja, $cov(C_1, C_2) = 0$ bei randame tikrinę reikšmę λ_2 . Tokį procesą tęsiame tol, kol išskiriamos visos komponentės C_3, \dots, C_k ir visa kintamųjų X_j ($j = 1, \dots, n$) dispersija yra paaiškinama komponentėmis, t.y. $\lambda_1 + \lambda_2 + \dots + \lambda_k = DX_1 + DX_2 + \dots + DX_n$. Įprastai dažniausiai kelios pirmosios komponentės paaiškina beveik visą pradinių kintamųjų dispersiją [21].

2.8. Diskriminantinė analizė

Diskriminantinė analizė – daugiamačės statistikos analizės metodas, priskiriantis objektą vienai iš kelių iš anksto apibrėžtų grupių, vadinamų klasėmis. Diskriminantinės analizės tikslas yra kuo patikimiau nustatyti nežinomo objekto klasę, remiantis to objekto kintamaisiais, kurie vadinami požymiais. Šie požymiai turi būti išmatuoti intervalinėje arba santykių skalėse. Diskriminantinės analizės modelis yra sudaromas, išanalizavus ir panaudojus objektų, kurių klasės jau žinomos, požymius. Vėliau modelis taikomas nežinomo objekto klasei prognozuoti [26].

Tarkime, kad matuojame kiekvieno objekto p intervalinių (vadinamų diskriminavimo) kintamųjų (X_1, X_2, \dots, X_p) . Žinome, kad objektų populiaciją sudaro g grupių. Imties duomenis sudaro stebėjimai $(x_{ijk}), i = 1, \dots, p, j = 1, \dots, g, k = 1, \dots, n_j$,

čia (x_{ijk}) yra i – tojo kintamojo k – tasis stebėjimas j – oje grupėje,

n_j - stebėjimų skaičius j – oje grupėje,

$n = n_1 + n_2 + \dots + n_g$ - imties didumas.

Toliau pateikiamos diskriminantinės analizės naudojimo prielaidos [26]:

1. Grupių skaičius g yra baigtinis.
2. Grupės yra nepriklausomos ir neturi bendrų objektų, t.y. objektas vienu metu negali priklausyti kelioms grupėms.
3. Diskriminavimo kintamieji $X_i \approx N(\mu, \sigma^2)$ ir matuojami pagal intervalų skalę.
4. Nė vienas diskriminavimo kintamasis negali būti kitų kintamųjų tiesinė daugdara.
5. Diskriminavimo kintamųjų kovariacijų matricos grupėse lygios.

Kovariacijų matricų lygybei tikrinti naudojamas Box'o M kriterijus. Nulinė hipotezė teigia, kad grupių kovariacijų matricos nesiskiria. Box'o kriterijus patvirtina, kad duomenys diskriminantinei analizei tinka, kai p – reikšmė $\geq \alpha$. Apie duomenų tinkamumą diskriminantinei analizei negalima spręsti vien tik pagal Box'o kriterijų. Jei duomenų aibė neturi daug išskirčių, tai diskriminantinė analizė stabili net tada, kai kovariacijų matricos nehomogeniškos. Jei stebėjimų labai daug, tai net maži kovariacijų matricų skirtumai taikant Box'o kriterijų pripažįstami statistiškai reikšmingi [26].

Objektai priskiriami klasėms panaudojant sudarytą modelį. Kadangi objektų klasės yra žinomos, randami teisingų ir klaidingų priskyrimų dažniai bei įvertinama klaidingo klasifikavimo tikimybė [26].

3. TYRIMO REZULTATAI

3.1. Darbe naudojami duomenys ir statistinės analizės modeliai

Darbe buvo naudojami telekomunikacijų bendrovės duomenys. Telekomunikacijų bendrovės abonentai buvo kviečiami dalyvauti akcijoje – nemokamai užsisakyti 3 mėn. žurnalo „Verslo klasė“ prenumeratą. Apie vykdomą akciją vartotojai buvo informuojami vienu iš komunikacijos kanalų, tai yra SMS žinute, el. paštu arba balso pranešimu. Taipogi buvo fiksuojama vartotojo reakcijos trukmė, tai yra per kiek laiko vartotojas peržiūrėjo gautą pranešimą nuo jo gavimo. Toliau pateiktoje lentelėje (žr. 3.1.1 lent.) pavaizduotas duomenų rinkinio fragmentas, kuriame matomi visi analizuojami kintamieji bei jų paaiškinimai.

Remiantis SOR modelio logika buvo tiriama bendrovės vartotojų elgsena. Komunikacijos kanalas veikia kaip stimulus, tai yra sukelia organizmui priėmimo arba atmetimo reakciją. Organizmo reakcija į kiekvieną komunikacijos kanalą yra skirtinga, todėl aktualu išsiaiškinti kiekvieno komunikacijos kanalo įtaką galutiniam vartotojo apsisprendimui, tai yra SMS žinutė, el. paštas ar balso pranešimas vartotojus veikia labiausiai. Teigiama reakcija fiksuojama kai vartotojas sutinka dalyvauti akcijoje, tai yra užsisako žurnalo prenumeratą, o neigiama, kai šios paslaugos atsisako. Taigi, organizmo reakcija buvo analizuota statistinės duomenų analizės modeliais.

3.1.1 lentelė. Kintamųjų aprašas

Kintamasis	Aprašas
AKCIJA	Akcija (T/N)
DUOMENYS	Priimtų bei išsiųstų duomenų kiekis, kb
ETERIS	Išėinančių balso skambučių bendra trukmė, s
ID	Vartotojo identifikatorius
KANALAS	Komunikacijos kanalas (A - SMS, B - el. paštas, C - balso pranešimas)
LYTIS	Vartotojo atstovo lytis
NUOLAIDA	Bendra nuolaidų suma per metus
SKAMBUČIŲ SKAIČIUS	Išėinančių balso skambučių skaičius
SMS SKAIČIUS	Išėinančių SMS skaičius
TRUKMĖ	Reakcijos trukmė į pranešimą, min.

Toliau pateikiami kintamųjų paaiškinimai:

Akcija – Vartotojas sutiko užsisakyti prenumeratą – T, nesutiko – N.

Duomenys – Vartotojo priimtų bei išsiųstų duomenų kiekis, kb.

Eteris – Bendra vartotojo išėinančių ir įėinančių skambučių trukmė, sekundėmis.

ID – Vartotojo identifikacijos numeris.

Kanalas – Komunikacijos kanalas, kuriuo buvo siūloma užsisakyti paslaugą.

Lytis – Vartotojo atstovo lytis (vyras arba moteris).

Nuolaida – Bendra vartotojo nuolaidų suma per metus, litais.

Skambučių skaičius – Vartotojo išeinančių balso skambučių skaičius ankstesnį mėnesį.

SMS skaičius – Vartotojų išeinančių SMS žinučių skaičius ankstesnį mėnesį.

Trukmė – Vartotojo reakcijos trukmė į gautą pasiūlymą užsisakyti paslaugą, minutėmis.

3.2. Aprašomoji statistika

Siekiant išsiaiškinti ir įvertinti duomenų sklaidos charakteristikas buvo vertinta aprašomoji statistika, tai yra imties plotis, suma, vidurkis, mažiausia ir didžiausia reikšmės, standartinis nuokrypis, asimetrijos bei eksceso koeficientai. Tam buvo pasinaudota programos SAS procedūra PROC MEANS. Visos tolidžiųjų kintamųjų skaitinės charakteristikos pateiktos 3.2.1 ir 3.2.2 lentelėse.

3.2.1 lentelė. Kintamųjų skaitinės charakteristikos

Kintamasis	Imties plotis	Suma	Vidurkis	Mažiausia reikšmė	Didžiausia reikšmė
DUOMENYS	5029462.81	700712353024	671402.93	0	5029462.81
ETERIS	120296.00	24538568466	23512.17	0	120296.00
NUOLAIDA	34.00	73109098.47	70.0510883	54.00	88.00
SKAMBUČIŲ SKAIČIUS	4571.00	178574778	171.1053453	0	4571.00
SMS SKAIČIUS	14959.00	1160361160	1111.83	0	14959.00
TRUKMĖ	54.00	20069893.00	19.2304087	1.00	55.00

3.2.2 lentelė. Kintamųjų skaitinės charakteristikos

Kintamasis	Standartinis nuokrypis	Asimetrijos koeficientas	Eksceso koeficientas
DUOMENYS	425059.27	1.1505200	2.3965226
ETERIS	16049.38	0.8388691	0.4057499
NUOLAIDA	5.8950090	0.1998630	-0.3140196
SKAMBUČIU SKAIČIUS	260.4993201	3.9298748	22.6394348
SMS SKAIČIUS	1173.18	1.6419418	4.0044479
TRUKMĖ	11.6128771	0.4429950	-0.1249599

Kaip matome vartotojo išsiųstų bei priimtų duomenų kiekis (DUOMENYS) stipriai varijuoja, kadangi yra vartotojų, kurie išviso nesinaudoja šia paslauga ir išsiųstų/priimtų duomenų kiekis

lygus 0, kai tuo tarpu kitų vartotojų išsiųstų/priimtų duomenų kiekis siekia net 5029462.81 kb kas sudaro 5 GB žr. (3.2.1 lent.). Tokią sklaidą atskleidžia ir nemažas standartinis nuokrypis, parodantis sklaidą apie vidurkį (425059,27). Vidutiniškai šios komunikacijų bendrovės paslaugų vartotojas duomenų išsiunčia/priima 671402.93 kb, kas sudaro apie 0,67 GB.

Taip pat galime pastebėti, jog bendra vartotojo išeinančių ir įeinančių skambučių trukmė, sekundėmis (ETERIS) taipogi stipriai varijuoja nuo 0 iki 120296 sekundžių, kas sudaro beveik 2005 minutes. Vidutiniškai vartotojas eteriye praleidžia 23512,17 sekundžių, tai yra 391,87 minutės.

Iš lentelės matome, jog didžiausia vartotojui suteikta nuolaida sudaro 88 Lt, o mažiausia 54 Lt. Vidutinė nuolaidos suma sudaro 70 Lt. Vartotojo reakcijos trukmė į gautą pasiūlymą užsisakyti paslaugą, minutėmis varijuoja nuo 1 min. iki 55 min. Vidutiniškai į gautą pasiūlymą vartotojas sureagoja per 19,23 min. Nulinės kai kurių vartotojų duomenų, eterio, skambučių ir SMS skaičiaus reikšmės rodo, kad šios paslaugos kai kuriems vartotojams nėra aktualios.

Spendžiant apie asimetrijos koeficientą (žr. 3.2.2 lent.) matome, jog visų analizuojamų kintamųjų asimetrija yra teigiama, kas reiškia, jog nuo vidurkio į dešinę pusę reikšmių sklaida yra didesnė. Tuo tarpu neigiami nuolaidos ir trukmės eksceso koeficientai rodo, kad šie dažnių skirstiniai šiek tiek lėkštesni nei normaliojo skirstinio. Pagal lentelę matome, jog didžiausią eksceso koeficiento reikšmę sudaro skambučių skaičiaus kintamasis, kas rodo apie labiau buką dažnių skirstinį palyginus su normaliuoju skirstiniu.

3.3. Suderinamumo hipotezių tikrinimas

Tolimesnėje dalyje buvo atliekamas suderinamumo hipotezių tikrinimas pasinaudojant SAS programos PROC UNIVARIATE procedūra. Buvo tikrinama ar imtys yra pasiskirsčiusios pagal normalųjį skirstinį, tai yra:

H_0 : imtys yra pasiskirsčiusios pagal normalųjį skirstinį.

H_a : imtys nėra pasiskirsčiusios pagal normalųjį skirstinį.

Atlikus suderinamumo hipotezių tikrinimą (žr. 3.3.1 lent.) matome, kad visų kintamųjų pasiskirstymo p reikšmė $< 0,01$ ($p < 0.05$), todėl hipotezę apie duomenų pasiskirstymą pagal normalųjį skirstinį atmetame kai reikšmingumo lygmeniu laikome 0.05.

3.3.1 lentelė. Suderinamumo hipotezių tikrinimas

Rodiklis	Kolmogorov-Smirnov statistika	Kolmogorov-Smirnov p reikšmė
SKAMBUČIŲ SKAIČIUS	0.255643	<0.0100
SMS SKAIČIUS	0.262789	<0.0100
TRUKMĖ	0.069373	<0.0100
DUOMENYS	0.068549	<0.0100
ETERIS	0.076365	<0.0100

NUOLAIDA	0.027023	<0.0100
Kanalas	0.222998	<0.0100
Akcija	0.384322	<0.0100
Lytis	0.341346	<0.0100

3.4. Koreliacinė analizė

Tolimesnėje tyrimo eigoje siekiant nustatyti kintamųjų (duomenys, eteris, nuolaida, skambučių skaičius, SMS skaičius, trukmė) tarpusavio ryšį buvo atlikta koreliacinė analizė. Šiai analizei atlikti buvo naudota PROC CORR procedūra. Kadangi hipotezių tikrinimo metu hipotezė, jog kintamieji yra pasiskirstę pagal normalųjį skirstinį buvo atmesta, visiems šiems kintamiesiems analizuoti buvo taikytas Spearman'o koreliacijos koeficientas.

Pagal žemiau pateiktą koreliacinę matricą (žr. 3.4.1 lent.) galime pastebėti, jog stipriausias teigiamas ryšys fiksuojamas tarp skambučių skaičiaus ir išsiųstų SMS skaičiaus, jų koreliacijos koeficientas lygus 0,64187 kas rodo apie vidutinį ryšio stiprumą. Teigiamas vidutinio stiprumo ryšys egzistuoja tarp eterio (bendra vartotojo išeinančių ir įeinančių skambučių trukmė, sekundėmis) ir duomenų bei eterio ir skambučių skaičiaus, jų Spearman'o koreliacijos koeficientai atitinkamai lygūs 0,62012 ir 0,60679. Taipogi pastebime, jog duomenų kiekio kintamasis silpnai koreliuoja su skambučių skaičiumi (Spearman'o koreliacijos koeficientas lygus 0,36906) bei SMS skaičius silpnai koreliuoja su eteriu (koreliacijos koeficientas lygus 0,2203). Tarp likusių kintamųjų fiksuojamas tik labai silpnas ryšys. Kadangi visų kintamųjų p reikšmė mažesnė už 0,001, koreliaciją galime laikyti statistiškai reikšminga.

3.4.1 lentelė. Spearman'o koreliacijos koeficientai

	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
DUOMENYS	1.00000	0.62012	-0.01785	0.36906	0.14342	-0.03878
		<.0001	<.0001	<.0001	<.0001	<.0001
ETERIS	0.62012	1.00000	-0.02566	0.60679	0.22031	-0.05897
	<.0001		<.0001	<.0001	<.0001	<.0001
NUOLAIDA	-0.01785	-0.02566	1.00000	-0.04083	-0.08456	-0.17505
	<.0001	<.0001		<.0001	<.0001	<.0001
SKAMBUČIŲ SKAIČIUS	0.36906	0.60679	-0.04083	1.00000	0.64187	-0.07155
	<.0001	<.0001	<.0001		<.0001	<.0001
SMS SKAIČIUS	0.14342	0.22031	-0.08456	0.64187	1.00000	-0.07341
	<.0001	<.0001	<.0001	<.0001		<.0001
TRUKMĖ	-0.03878	-0.05897	-0.17505	-0.07155	-0.07341	1.00000
	<.0001	<.0001	<.0001	<.0001	<.0001	

Tolimesnėje darbo eigoje buvo apskaičiuoti Spearman'o koreliacijos koeficientai atskirai moterims (žr. 3.4.2 lent.). Kaip matome stipriausias teigiamas ryšys (koreliacijos koeficientas lygus 0,64212) fiksuojamas tarp išsiųstų SMS ir skambučių skaičiaus. Šiek tiek silpnesnė koreliacija matoma tarp įeinančių skambučių trukmės (eterio) ir išsiųstų/priimtų duomenų kiekio bei tarp eterio ir skambučių skaičiaus. Silpnas ryšys fiksuojamas tarp skambučių skaičiaus ir duomenų.

3.4.2 lentelė. Koreliacijos koeficientai, apskaičiuoti atskirai moterims.

	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
DUOMENYS	1.00000	0.63455 <.0001	-0.01605 <.0001	0.33555 <.0001	0.13862 <.0001	-0.01723 <.0001
ETERIS	0.63455 <.0001	1.00000	-0.02428 <.0001	0.54354 <.0001	0.20793 <.0001	-0.02399 <.0001
NUOLAIDA	-0.01605 <.0001	-0.02428 <.0001	1.00000	-0.07065 <.0001	-0.11844 <.0001	-0.11092 <.0001
SKAMBUČIŲ SKAIČIUS	0.33555 <.0001	0.54354 <.0001	-0.07065 <.0001	1.00000	0.64212 <.0001	-0.03623 <.0001
SMS SKAIČIUS	0.13862 <.0001	0.20793 <.0001	-0.11844 <.0001	0.64212 <.0001	1.00000	-0.06340 <.0001
TRUKMĖ	-0.01723 <.0001	-0.02399 <.0001	-0.11092 <.0001	-0.03623 <.0001	-0.06340 <.0001	1.00000

Atskirai vyrams apskaičiuota koreliacija (žr. 3.4.3 lent.) parodė labai panašius rezultatus kaip ir koreliacija atlikta atskirai moterims. Visi kintamieji susiję panašaus stiprumo ryšiu, todėl galime teigti, jog lytis didelės reikšmės koreliacinei analizei neturi.

3.4.3 lentelė. Koreliacijos koeficientai, apskaičiuoti atskirai vyrams.

	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
DUOMENYS	1.00000	0.60487 <.0001	-0.01122 <.0001	0.41878 <.0001	0.15571 <.0001	-0.06383 <.0001
ETERIS	0.60487 <.0001	1.00000	-0.01575 <.0001	0.69376 <.0001	0.24211 <.0001	-0.09970 <.0001
NUOLAIDA	-0.01122 <.0001	-0.01575 <.0001	1.00000	-0.04151 <.0001	-0.06387 <.0001	-0.21775 <.0001

SKAMBUČIŲ SKAIČIUS	0.41878 <.0001	0.69376 <.0001	-0.04151 <.0001	1.00000	0.62691 <.0001	-0.10005 <.0001
SMS SKAIČIUS	0.15571 <.0001	0.24211 <.0001	-0.06387 <.0001	0.62691 <.0001	1.00000	-0.07850 <.0001
TRUKMĖ	-0.06383 <.0001	-0.09970 <.0001	-0.21775 <.0001	-0.10005 <.0001	-0.07850 <.0001	1.00000

Taipogi buvo apskaičiuota koreliacija atskirai dalyvaujantiems akcijoje ir nedalyvaujantiems (žr. 3.4.4 ir 3.4.5 lenteles). Ši analizė vėl gi reikšmingų skirtumų neparodė. Visi kintamieji koreliuoja tarpusavyje labai panašiai tiek dalyvaujančių, tiek nedalyvaujančių akcijoje.

3.4.4 lentelė. Koreliacijos koeficientai, apskaičiuoti užsisakiusiems paslaugą.

	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
DUOMENYS	1.00000	0.64305 <.0001	-0.13670 <.0001	0.35689 <.0001	0.14531 <.0001	-0.05395 <.0001
ETERIS	0.64305 <.0001	1.00000	-0.21355 <.0001	0.56349 <.0001	0.22332 <.0001	-0.08120 <.0001
NUOLAIDA	-0.13670 <.0001	-0.21355 <.0001	1.00000	-0.10877 <.0001	-0.08816 <.0001	0.02726 <.0001
SKAMBUČIŲ SKAIČIUS	0.35689 <.0001	0.56349 <.0001	-0.10877 <.0001	1.00000	0.65990 <.0001	-0.00577 0.0001
SMS SKAIČIUS	0.14531 <.0001	0.22332 <.0001	-0.08816 <.0001	0.65990 <.0001	1.00000	0.01304 <.0001
TRUKMĖ	-0.05395 <.0001	-0.08120 <.0001	0.02726 <.0001	-0.00577 0.0001	0.01304 <.0001	1.00000

3.4.5 lentelė. Koreliacijos koeficientai, apskaičiuoti paslaugos neužsisakiusiems.

	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
DUOMENYS	1.00000	0.58430 <.0001	0.00271 0.0341	0.36106 <.0001	0.16820 <.0001	0.00063 0.6253
ETERIS	0.58430 <.0001	1.00000	0.00743 <.0001	0.62342 <.0001	0.26303 <.0001	0.00219 0.0877
NUOLAIDA	0.00271 0.0341	0.00743 <.0001	1.00000	-0.03268 <.0001	-0.07844 <.0001	-0.26387 <.0001

SKAMBUČIŲ SKAIČIUS	0.36106 <.0001	0.62342 <.0001	-0.03268 <.0001	1.00000	0.66289 <.0001	-0.09120 <.0001
SMS SKAIČIUS	0.16820 <.0001	0.26303 <.0001	-0.07844 <.0001	0.66289 <.0001	1.00000	-0.11374 <.0001
TRUKMĖ	0.00063 0.6253	0.00219 0.0877	-0.26387 <.0001	-0.09120 <.0001	-0.11374 <.0001	1.00000

3.5. Dispersinė analizė

Pasinaudojant dispersine analize buvo tiriama faktorių (akcija, lytis, kanalas) įtaka kitiems rodikliams. Dėl neparametrinių kriterijų, šiai analizei atlikti buvo pasinaudota PROC NPAR1WAY procedūra.

Pagal gautus Kruskal-Wallis'o kriterijus, visų tirtų faktorių p reikšmės mažesnės už reikšmingumo kriterijų 0,005 ($p < 0,0001$), tai reiškia, kad visi tiriami faktoriai yra statistiškai reikšmingi. Be to, pasinaudojant nuokrypių kvadratų suma buvo apskaičiuotas įtakos procentinis dydis sklaidai, kuris parodo kiekvieno tirto faktoriaus įtaką kintamiesiems (žr. 3.5.1 lent.).

3.5.1 lentelė. Faktorių akcija, lytis, kanalas dispersinės analizės lentelė

Veiksniai	Analitės					
	Duomenys	Eteris	Nuolaida	Skambučių skaičius	SMS skaičius	Trukmė
Faktoriaus Akcija įtakos procentas sklaidai	1,57%	4,74%	10,22%	0,12%	2,38%	2,59%
Faktoriaus Lytis įtakos procentas sklaidai	0,05%	0,16%	1,75%	0,52%	0,02%	0,24%
Faktoriaus Kanalas įtakos procentas sklaidai	8,19%	24,81%	1,21%	9,83%	71,55%	2,84%

Taigi, pagal 3.5.1 lentelės duomenis matome, kad duomenims bei eteriui didžiausią poveikį daro komunikacijos kanalas, šių faktorių įtaka duomenų sklaidai atitinkamai sudaro 8,19% ir 24,81%. Duomenų ir eterio kintamiesiems mažiausią įtaką daro vartotojo atstovo lytis. Didžiausią, tai yra 10,22% įtaką nuolaidai daro akcija, kai tuo tarpu lyties ir kanalo įtaka nuolaidai nesudaro nei 2%. Skambučių skaičiui didžiausią reikšmę turi komunikacijos kanalas, kuris daro šiam kintamajam 9,83% įtaką. SMS skaičių kaip ir skambučių skaičių vėl gi labiausiai veikia komunikacijos kanalas, darydamas kintamiesiems net 71,55% ir 9,83% įtaką sklaidai. Trukmės rodikliui didžiausią poveikį turi taipogi kanalas (2,84%), toliau seka akcija (2,59%) bei lytis (0,24%).

Kadangi atmetėme hipotezę apie duomenų pasiskirstymą pagal normalųjį skirstinį, tam, kad išsiaiškintume, kurio faktoriaus rangų vidurkis daro didesnę įtaką kintamiesiems, galime naudotis Wilcoxon'o rangų sumos lentele.

Kadangi duomenys turi du faktorius, darančius jam įtaką, juos ir apžvelgsime. Kaip matome pagal 3.5.2 lentelę vartotojai priėmę pasiūlymą dalyvauti akcijoje daro didesnę įtaką duomenims, nes jų rangų vidurkis yra didesnis.

3.5.2 lentelė. Wilcoxon'o rangų sumos lentelė (duomenys)

Akcija	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Ne	608800	2.95666E11	3.17689E11	151738278	485653.098
Taip	434854	2.48942E11	2.26919E11	151738278	572472.029

Pagal žemiau esančią 3.5.3 lentelę galime pastebėti, jog didžiausią įtaką duomenims turi trečiasis komunikacijos kanalas, tai yra balso pranešimas, mažesnę įtaką turi pirmasis - SMS žinutė, o mažiausią įtaką daro el. paštu gautas pranešimas.

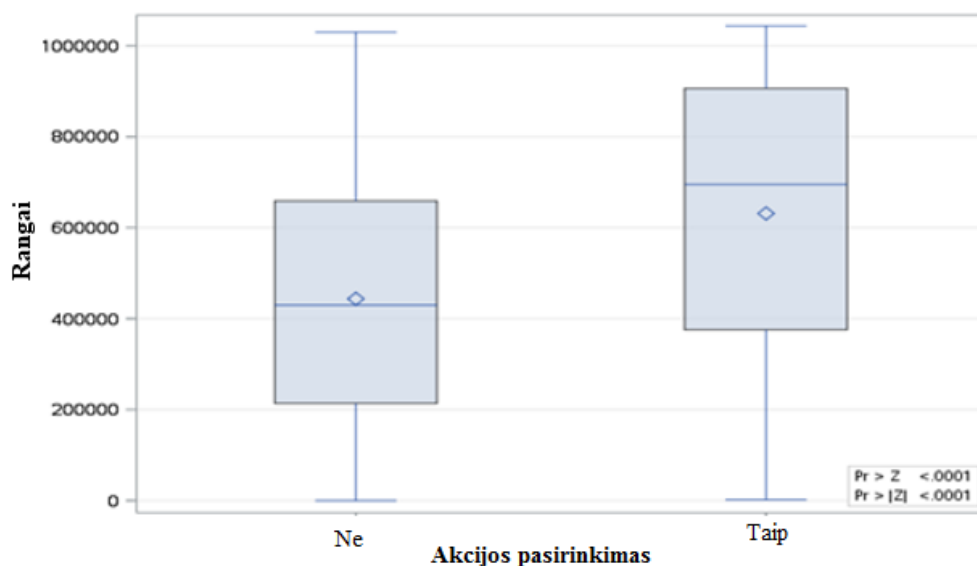
3.5.3 lentelė. Wilcoxon'o rangų sumos lentelė (duomenys)

Komunikacijos kanalas	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Balso pranešimas	347883	2.2714E11	1.81535E11	145089458	652919.524
El. paštas	347886	1.48149E11	1.81536E11	145089771	425854.006
SMS	347885	1.69319E11	1.81536E11	145089667	486709.999

Kintamasis nuolaida turi tris faktorius darančius jam įtaką. Pagal 3.5.4 lent. matome, jog didžiausią įtaką nuolaidai turi akcijoje dalyvavę vartotojai. Jų rangų sumos vidurkis žymiai didesnis. Visa tai parodo ir vidurkių pasiskirstymo grafikas (žr. 3.5.1 pav.), kuriame matomi aiškūs rangų vidurkių skirtumai, parodantys, kad faktorius tikrai veikia.

3.5.4 lentelė. Wilcoxon'o rangų sumos lentelė (nuolaida)

Akcija	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Ne	608800	2.70068E11	3.17689E11	151738874	443607.299
Taip	434854	2.74539E11	2.26919E11	151738874	631336.573



3.5.1 pav. Wilcoxon'o vidurkių pasiskirstymo grafikas pagal nuolaidą.

3.5.5 lent. matome lyties įtaką nuolaidai. Galime teigti, jog moterys turi šiek tiek mažesnę įtaką nuolaidai nei vyrai, kadangi jų rangų sumos vidurkis mažesnis.

3.5.5 lentelė. Wilcoxon'o rangų sumos lentelė (nuolaida)

Lytis	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Moteris	521830	2.50864E11	2.72305E11	153891399	480738.165
Vyras	521824	2.93744E11	2.72302E11	153891399	562917.308

3.5.6 lent. pateikta komunikacijos kanalo įtaka suteikiamai nuolaidai. Antrasis komunikacijos kanalas - el. paštas turi didžiausią, šiek tiek mažesnę – balso pranešimas, o mažiausią įtaką turi SMS pranešimas.

3.5.6 lentelė. Wilcoxon'o rangų sumos lentelė (nuolaida)

Komunikacijos kanalas	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Balso	347883	1.78437E11	1.81535E11	145090028	512923.173
El. paštas	347886	1.97135E11	1.81536E11	145090341	566665.106
SMS	347885	1.69035E11	1.81536E11	145090237	485894.041

3.5.7 lent. matome akcijos įtaką vartotojo išeinančių balso skambučių skaičiui. Vartotojai sutikę dalyvauti akcijoje taipogi daro didesnę poveikį kintamajam nei dalyvauti atsisakę.

3.5.7 lentelė. Wilcoxon'o rangų sumos lentelė (skambučių skaičius)

Akcija	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Ne	608800	3.03046E11	3.17689E11	151736801	497776.426
Taip	434854	2.41561E11	2.26919E11	151736801	555499.248

3.5.8 lent. pateikta lyties įtaka skambučių skaičiui. Vėl gi vyrų įtaka vartotojo skambučių skaičiui – didesnė.

3.5.8 lentelė. Wilcoxon'o rangų sumos lentelė (skambučių skaičius)

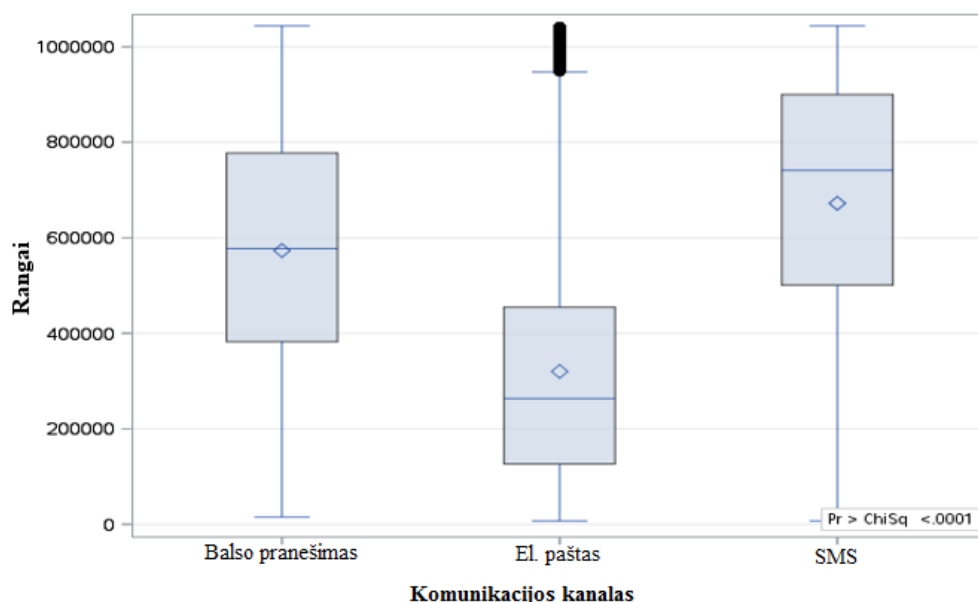
Lytis	Imties tūris	Rangų suma		Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
0	521830	2.53259E11	2.72305E11	153889296	153889296	485328.627
1	521824	2.91348E11	2.72302E11	153889296	153889296	558326.792

3.5.9 lent. matome komunikacijos kanalo įtaką skambučių skaičiui. Šiuo atveju, didžiausią įtaką daro pirmasis komunikacijos kanalas - SMS žinutė, toliau seka trečiasis komunikacijos kanalas – balso pranešimas ir mažiausią poveikį turi antrasis, tai yra el. paštas. 3.5.2 pav. matome

3.5.9 lentelė. Wilcoxon'o rangų sumos lentelė (skambučių skaičius)

Komunikacijos kanalas	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Balso	347883	1.99453E11	1.81535E11	145088046	573332.866
El. paštas	347886	1.11349E11	1.81536E11	145088359	320073.182
SMS	347885	2.33806E11	1.81536E11	145088255	672077.328

Visų trijų komunikacijos kanalų rangų vidurkiai išties skiriasi, tai atsispindi ir 3.5.2 pav.



3.5.2 pav. Wilcoxon'o vidurkių pasiskirstymo grafikas pagal skambučių skaičių.

Kitų faktorių Wilcoxon'o rangų sumos lentelės pateiktos Priede 2.

3.6. Požymių priklausomumo lentelių tyrimas

Priklausomybei tarp kategorinių kintamųjų ištirti buvo atliekamas požymių priklausomumo lentelių tyrimas. Šiai analizei atlikti buvo naudota procedūra PROC FREQ.

Buvo tikrinama požymių tarpusavio nepriklausomumo hipotezė, tai yra:

H_0 : požymiai tarpusavyje nėra statistiškai reikšmingai priklausomi.

H_a : požymiai tarpusavyje yra statistiškai reikšmingai priklausomi.

Atlikus hipotezių tikrinimą (žr. 3.6.1 lent.) matome, kad lyties ir komunikacijos kanalo kintamųjų Chi-kvadrato kriterijaus p reikšmė lygi 1, todėl H_0 priimame ir galime teigti, jog šie kintamieji nėra statistiškai reikšmingai priklausomi, be to tarp jų neegzistuoja joks ryšys kadangi Phi, kontingencijos ir Kramerio V koeficientai lygūs 0.

3.6.1 lentelė. Kintamųjų ryšio matų lentelė

Chi-kvadratas	1.0000
Phi koeficientas	0.0000
Kontingencijos koeficientas	0.0000
Kramerio V koeficientas	0.0000

3.6.2 lentelė. Požymių priklausomumo lentelė (lytis ir kanalas)

	Lytis	Kanalas			
		El. paštas	SMS	Balso pranešimas	Iš viso
Dažnis	Vyras	173942	173945	173943	521830
Visumos procentas		16.67	16.67	16.67	50.00
Eilutės procentas		33.33	33.33	33.33	
Stulpelio procentas		50.00	50.00	50.00	
Dažnis	Moteris	173941	173941	173942	521824
Visumos procentas		16.67	16.67	16.67	50.00
Eilutės procentas		33.33	33.33	33.33	
Stulpelio procentas		50.00	50.00	50.00	
	Iš viso	347883	347886	347885	1043654
		33.33	33.33	33.33	100.00

Patikrinus akcijos ir komunikacijos kanalo požymių priklausomumą, galime atmesti H_0 , kadangi p reikšmė $<0,001$. Tai reiškia, jog tiriami akcijos ir kanalo kintamieji yra statistiškai reikšmingai priklausomi ir tarp jų egzistuoja vidutinio stiprumo ryšys (žr. 3.6.3 lent.)

3.6.3 lentelė. Akcijos ir komunikacijos kanalo kintamųjų ryšio matų lentelė

Chi-kvadratas	<.0001
Phi koeficientas	0.47809
Kontingencijos koeficientas	0.43133
Kramerio V koeficientas	0.47809

Pagal 3.6.4 lentelę matome, kad į 14,29% akcijoje nedalyvaujančių vartotojų, buvo kreipiamasi balso pranešimu, tuo tarpu į 42,86% nedalyvaujančių el. paštu ir SMS žinute. Tarp visų su akcijos pasiūlymų sutikusių į 60% buvo kreipiamasi balso pranešimu, o į 20% el. paštu arba SMS pranešimu. Be to, svarbu paminėti, jog tarp visų vartotojų į kuriuos buvo kreiptasi el. paštu arba SMS pranešimu 75% atsisakė pasiūlymo dalyvauti akcijoje, o 25% sutiko. Į tuos vartotojus į kuriuos buvo kreipiamasi balso žinute dalyvauti akcijoje atsisakė tik 25%, o 75% pasiūlymą dalyvauti akcijoje priėmė.

3.6.4 lentelė. Požymių priklausomumo lentelė (akcija ir kanalas)

	Akcija	Kanalas			
		Balso pranešimas	El. paštas	SMS	Iš viso
Dažnis	Ne	86972	260915	260913	608800
Visumos procentas		8.33	25.00	25.00	58.33
Eilutės procentas		14.29	42.86	42.86	
Stulpelio procentas		25.00	75.00	75.00	
Dažnis	Taip	260911	86971	86972	434854
Visumos procentas		25.00	8.33	8.33	41.67
Eilutės procentas		60.00	20.00	20.00	
Stulpelio procentas		75.00	25.00	25.00	
	Iš viso	347883	347886	347885	1043654
		33.33	33.33	33.33	100.00

Toliau buvo tiriama priklausomybė tarp lyties ir akcijos kintamųjų. Mažas pataisyto Chi-kvadrato kriterijaus reikšmingumas 3.6.5 lentelėje rodo, kad ryšys tarp akcijos ir lyties kintamųjų yra ($p < 0,001$), todėl nulinę hipotezę apie požymių nepriklausomumą atmetame. Tarp šių kintamųjų fiksuojamas silpnas ryšys.

3.6.5 lentelė. Akcijos ir lyties kintamųjų ryšio matų lentelė

Pataisytas Chi-kvadratas	<.0001
Phi koeficientas	0.3043
Kontingencijos koeficientas	0.2911
Kramerio V koeficientas	0.3043

3.6.6 lentelė. Požymių priklausomumo lentelė

	Akcija	Lytis		
		Moteris	Vyras	Iš viso
Dažnis	Ne	382676	226124	608800
Visumos procentas		36.67	21.67	58.33
Eilutės procentas		62.86	37.14	
Stulpelio		73.33	43.33	
Dažnis	Taip	139154	295700	434854
Visumos procentas		13.33	28.33	41.67
Eilutės procentas		32.00	68.00	
Stulpelio		26.67	56.67	
	Iš viso	521830	521824	1043654
		50.00	50.00	100.00

Pagal 3.6.6 lentelės duomenis matome, kad iš visų atsisakiusių dalyvauti akcijoje 62,86% buvo moterys, o 37,14% vyrai. Iš visų pasiūlymą priėmusių, 32% buvo moterų ir 68% vyrų. Taipogi, iš visų moterų pasiūlymą dalyvauti akcijoje priėmė tik 26,67%, tuo tarpu iš visų vyrų pasiūlymą priėmė 56,67%.

3.7. Logistinė regresija

Apjungus anksčiau minėtus kintamuosius į logistinės regresijos modelį, tiriama kaip akcijos pasirinkimas (akcija) priklauso nuo komunikacijos kanalo (kanalas), vartotojo atstovo lyties (lytis), bendros vartotojo išeinančių ir įeinančių skambučių trukmės (eteris), vartotojo priimtų bei išsiųstų duomenų kiekio (duomenys), suteiktos nuolaidos (nuolaida), išeinančių skambučių skaičiaus (skambučių skaičius), išeinančių SMS skaičiaus (SMS skaičius) bei reakcijos trukmės į pranešimą (trukmė). Šiam žingsniui atlikti buvo pasinaudota PROC LOGISTIC procedūra.

Tikriname hipotezę:

H_0 : visi $\beta_m=0$, tai yra Y nepriklauso nuo X;

H_a : ne visi $\beta_m=0$, tai yra Y priklauso nuo X.

Pagal apačioje pateiktą 3.7.1 lentelę, matome, kad p reikšmė yra $<0,001$, o tai mažiau už reikšmingumo lygmenį α , kuris lygus 0,05. Tai rodo pakankamai gerą modelio tikimą duomenims, todėl nulinę hipotezę atmetame.

3.7.1 lentelė. Regresijos modelio statistika

Testai	Chi-kvadratas	<i>p</i> reikšmė
Tikėtinumo santykio	516647.096	<.0001
Score	419944.151	<.0001
Voldo	252707.714	<.0001

Vis dėl to, pagal Voldo kriterijaus *p* reikšmę (žr. 3.7.2 lent.), kintamieji duomenys (*p* reikšmė 0.8670), eteris (*p* reikšmė 0.3673), skambučių skaičius (*p* reikšmė 0.1817), SMS skaičius (*p* reikšmė 0.7656), turėtų būti pašalinti iš regresijos modelio, nes šių kintamųjų *p* reikšmės viršija α reikšmingumo lygmenį 0,05.

3.7.2 lentelė. Regresijos modelis su visais kintamaisiais

Kriterijai		Estimate	Standartinis nuokrypis	Voldo Chi-kvadratas	<i>p</i> reikšmė
Intercept		12.6479	0.0407	96672.4876	<.0001
kanalas	1	0.8719	0.00904	9294.1319	<.0001
kanalas	2	1.2179	0.00626	37847.8832	<.0001
lytis	0	0.9087	0.00294	95736.1868	<.0001
kanalas*lytis	1 0	-0.1036	0.00410	638.7539	<.0001
kanalas*lytis	2 0	0.0308	0.00411	55.9187	<.0001
DUOMENYS		1.257E-9	7.508E-9	0.0280	0.8670
ETERIS		-2.14E-7	2.369E-7	0.8129	0.3673
NUOLAIDA		-0.1767	0.000556	101079.670	<.0001
SKAMBUČIŲ SKAIČIUS		0.000023	0.000017	1.7837	0.1817
SMS SKAIČIUS		-1.94E-6	6.515E-6	0.0889	0.7656
TRUKMĖ		0.0111	0.000241	2112.1095	<.0001

Svarbu paminėti, kad sudarytas modelis paaiškina 39,05% priklausomojo kintamojo akcija reikšmės (žr. 3.7.3). Be to, Nagelkerkės determinacijos koeficientas vadinamas lygus 0,5256, kas rodo neblogą modelio tinkamumą duomenims.

3.7.3 lentelė. Determinacijos koeficientai

R-kvadratas	0.3905	Nagelkerkės determinacijos koeficientas	0.5256
-------------	--------	---	--------

Toliau, pašalinus statistiškai nereikšmingus regresorius, sudarome naują regresijos modelį (žr. 3.7.4 lent.). Šio regresijos modelio *p* reikšmės taipogi neviršija α 0,05 lygmens, kas rodo pakankamai gerą modelio tikimą duomenims.

3.7.4 lentelė. Regresijos modelio statistika

Kriterijai	Chi-kvadratas	<i>p</i> reikšmė
Tikėtinumo santykio	516644.336	<.0001
Score	419942.313	<.0001
Voldo	252706.999	<.0001

Kaip matome, visų kintamųjų p reikšmės mažesnės už α reikšmingumo lygmenį (žr. 3.7.5 lent.). Be to, pašalinus šiuos kintamuosius, modelio R kvadratas išliko beveik toks pats 0,3904. Tai reiškia, jog sudarytas modelis paaiškina 39,04% priklausomojo kintamojo (Akcija) reikšmės (žr. 3.7.6 lent.). Nagelkerkės determinacijos koeficientas nepakito.

3.7.5 lentelė. Regresijos modelis.

Kriterijai			Įvertis	Standartinis nuokrypis	Voldo Chi-kvadratas	p reikšmė
laisvasis narys			12.6457	0.0401	99486.5797	<.0001
kanalas_N	1		0.8719	0.00408	45778.6773	<.0001
kanalas_N	2		1.2187	0.00448	73899.8674	<.0001
lytis_N	0		0.9082	0.00292	96889.8423	<.0001
kanalas_N*lytis_N	1	0	-0.1031	0.00409	636.4611	<.0001
kanalas_N*lytis_N	2	0	0.0313	0.00410	58.2553	<.0001
NUOLAIDA			-0.1767	0.000556	101082.001	<.0001
TRUKME			0.0111	0.000241	2111.9899	<.0001

3.7.6 lentelė. Determinacijos koeficientai

R-kvadratas	0.3904	Nagelkerkės determinacijos koeficientas	0.5256
-------------	--------	---	--------

3.8. Faktorinė analizė

Toliau buvo atliekama faktorinė analizė, pasinaudojant pagrindinių komponentių metodu. Šiai užduočiai atlikti buvo naudojama PROC FACTOR procedūra. Pagrindinių komponentių analizei buvo naudojami duomenų, eterio, nuolaidos, skambučių skaičiaus, SMS skaičiaus ir trukmės kintamieji.

Pagal apačioje pateiktą tikrinių reikšmių lentelę (žr. 3.8.1 lent.) matome, jog gavome tris pagrindines komponentes, kadangi tik jų tikrinės reikšmės yra didesnės už 1. Pirmoje komponentėje sukaupta 34,43% visos informacijos. Na, o visos trys komponentės sudaro net 74,30 % visos duomenų sklaidos. Netenkama 25,7% informacijos, tačiau kintamųjų skaičius sumažintas dvigubai.

3.8.1 lentelė. Tikrinės reikšmės

	Tikrinė reikšmė	Skirtumas	Proporcija	Kaupiamasis dydis
1	2.06600274	0.82588226	0.3443	0.3443
2	1.24012048	0.08845276	0.2067	0.5510
3	1.15166773	0.34951838	0.1919	0.7430
4	0.80214935	0.34828456	0.1337	0.8767
5	0.45386478	0.16766986	0.0756	0.9523
6	0.28619492		0.0477	1.0000

Remiantis 3.8.2 lentele matome, jog pirmajai komponentei (apimties) priskirti duomenų bei eterio kintamieji. Antrąją komponentę (intensyvumo) sudaro vartotojo išeinančių balso skambučių

ir SMS skaičius. Trečiajai komponentei (lojalumo) priklauso nuolaida bei reakcijos trukmė į gautą pranešimą.

3.8.2 lentelė. Faktorių pasiskirstymas, Varimax metodu

	Apimties faktorius		Intensyvumo faktorius		Lojalumo faktorius	
DUOMENYS	88	*	2		0	
ETERIS	87	*	20		3	
NUOLAIDA	-3		-10		77	*
SKAMBUČIŲ	32		85	*	2	
SMS SKAIČIUS	-4		92	*	-3	
TRUKMĖ	-5		-9		-77	*

3.8.3 lentelėje pateikiamos visų komponentių reikšmės

3.8.3 lentelė. Standartizuotos koeficientų reikšmės

	Apimties faktorius	Intensyvumo faktorius	Lojalumo faktorius
DUOMENYS	0.31256	0.36334	0.35115
ETERIS	0.36747	0.29006	0.25710
NUOLAIDA	-0.03246	0.43560	-0.48394
SKAMBUČIŲ SKAIČIUS	0.39666	-0.23911	-0.21033
SMS SKAIČIUS	0.29874	-0.44849	-0.35022
TRUKMĖ	-0.06050	-0.37703	0.52708

Toliau, pagal 3.8.3 lent. buvo sudarytos pagrindinių trijų komponentių išraiškos, kurios yra pradinių kintamųjų tiesinė kombinacija:

$$Z_1 = 0.31256X_1 + 0.36747X_2 - 0.03246X_3 + 0.39666 X_4 + 0.29874X_5 - 0.06050X_6;$$

$$Z_2 = 0.36334X_1 + 0.29006X_2 + 0.43560X_3 - 0.23911X_4 - 0.44849X_5 - 0.37703X_6;$$

$$Z_3 = 0.35115X_1 + 0.25710X_2 - 0.48394X_3 - 0.21033X_4 - 0.35022X_5 + 0.52708 X_6.$$

3.9. Diskriminantinė analizė

Naudojantis gautais faktorinės analizės rezultatais buvo atlikta diskriminantinė analizė. Pirmiausia pasinaudojant PROC SURVEYSELECT procedūra buvo atskirtos apmokymo ir validumo imtys. Apmokymo imčiai naudota 30%, o validumo - 70% pradinių duomenų. Pačiai diskriminantinei analizei atlikti buvo naudota PROC DISCRIM procedūra. Taip pat buvo nustatytas artimiausių kaimynų skaičiaus parametras, $k=100$.

3.9.1 lentelėje pateikiama stebinių skaičiaus, kintamųjų ir klasių informacijos suvestinė. Kaip matome, šios apmokymo imties dydis sudarė 313 097 eilutes. Trys faktoriai buvo priskirti dviem klasėms, tai yra priėmusiems pasiūlymą dalyvauti akcijoje ir atsakiusiems.

3.9.1 lentelė. Stebinių, kintamųjų ir klasių informacija

Imties dydis	313097	Laisvės laipsnių skaičius iš viso	313096
Kintamieji	3	Laisvės laipsnių skaičius klasių viduje	313095
Klasės	2	Laisvės laipsnių skaičius tarp klasių	1

Žemiau esančioje 3.9.2 lentelėje pateikiami kiekvienos klasės dažnio, svorio, proporcijos bei tikimybės rodikliai. Kaip matome tikimybė faktoriams būti priskirtiems kiekvienai klasei sudaro 0,5.

3.9.2 lentelė. Klasių informacija

Akcija	Kintamojo vardas	Dažnis	Svoris	Proporcija	Svoriai
0	0	182635	182635	0.583318	0.500000
1	1	130462	130462	0.416682	0.500000

Toliau (žr. 3.9.3 lent.) pateikiama mokymo imties klaidų matrica, parodanti teisingą ir klaidingą priskyrimo į klases skaičių. Matome, jog 125 874 atsisakiusiųjų dalyvauti akcijoje kintamieji arba beveik 69% jų buvo suklasifikuoti teisingai su 31% klaidingu priskyrimu. Priėmusių pasiūlymą klasifikavimo rezultatai geresni. Teisingai suklasifikuota beveik 73%, klaidingai 27%. Bendras klaidingo priskyrimo į klases santykis lygus 29% (žr. 3.9.4 lent.).

3.9.3 lentelė. Klaidingo priskyrimo klasėms santykių matrica

		Modelio įvertintos reikšmės			
		Klasifikuota iš Akcijos	Ne	Taip	Iš viso
Stebėtos reikšmės	Ne		125874	56761	182635
			68.92	31.08	100.00
	Taip		35408	95054	130462
			27.14	72.86	100.00
Iš viso			161282	151815	313097
			51.51	48.49	100.00

3.9.4 lentelė. Klaidingo priskyrimo klasėms santykis

Klaidingo priskyrimo santykis Akcijai			
	Ne	Taip	Iš viso
Santykis	0.3108	0.2714	0.2911
Svoriai	0.5000	0.5000	

Taipogi buvo apskaičiuota validavimo imties klaidų matrica (žr. 3.9.5 lent.). Ši matrica padeda geriau įvertinti klasifikavimo tikslumą. Kaip matome pagal žemiau pateiktą lentelę, 69% akcijoje nedalyvaujančiųjų buvo teisingai priskirti klasėms kai tuo tarpu 31% klaidingai. Akcijoje dalyvaujantys 73% atveju klasifikuojami teisingai, o 27% neteisingai. Abiejų klasifikacijų

rezultatai beveik nesiskiria, bendras klaidingo priskyrimo į klases santykis lygus 29% (žr. 3.9.6 lent.).

3.9.5 lentelė. Klaidingo priskyrimo klasėms santykių matrica

		Modelio įvertintos reikšmės		
Klasifikuota iš Akcijos		Ne	Taip	Iš viso
Stebėtos reikšmės	Ne	294243	131922	426165
		69.04	30.96	100.00
	Taip	82429	221963	304392
		27.08	72.92	100.00
	Iš viso	376672	353885	730557
		51.56	48.44	100.00

3.9.6 lentelė. Klaidingo priskyrimo klasėms santykis

Klaidingo priskyrimo santykis Akcijos klasei			
	Ne	Taip	Iš viso
Santykis	0.3096	0.2708	0.2902
Svoriai	0.5000	0.5000	

IŠVADOS

1. Atlikus literatūros analizę, išsiaiškinta, kad dauguma mokslininkų vartotojų elgseną apibūdina kaip vartotojų prekių ar paslaugų rinkimosi, pirkimo ir naudojimosi procesą siekiant patenkinti savo poreikius. Norint atlikti vartotojų elgsenos tyrimus ir tinkamai apdoroti didžiuosius duomenis svarbu pasirinkti funkcionalias ir prie rinkos poreikių prisitaikančias duomenų mokslo platformas - IBM, SAS, RapidMiner, KNIME.
2. Vartotojų elgsenos tyrimui atlikti buvo atrinkti aprašomosios statistikos, suderinamumo hipotezių tikrinimo, koreliacinės, dispersinės analizių, požymių priklausomumo lentelių tyrimo, logistinė regresijos, faktorinės ir diskriminantinės analizių metodai.
3. Atlikus telekomunikacijų bendrovės vartotojų elgsenos tyrimą gauti rezultatai:
 - Koreliacinė analizė parodė, jog kuo didesnis vartotojo išeinančių balso skambučių skaičius, tuo tikėtiniau, kad ir išsiųstų SMS skaičius bus didesnis. Taipogi kuo didesnė bendra vartotojo įeinančių ir išeinančių skambučių trukmė, tuo tikėtiniau, kad vartotojo išeinančių balso skambučių skaičius bei priimtų ir išsiųstų duomenų kiekis bus didesnis.
 - Pasinaudojus dispersine analize buvo išsiaiškinta, kad vartotojų sprendimui priimti akcijos pasiūlymą didžiausią statistiškai reikšmingą įtaką daro ankščiau gautų nuolaidų suma. Tuo tarpu komunikacijos kanalo parinkimui didžiausią statistiškai reikšmingą įtaką daro duomenų, eterio, trukmės, skambučių ir SMS skaičiaus kintamieji.
 - Atlikus požymių priklausomumo lentelių tyrimą paaiškėjo, kad tarp vartotojo atstovo lyties ir komunikacijos kanalo statistiškai reikšmingas ryšys neegzistuoja. Taip pat išsiaiškinta, kad į tuos vartotojus į kuriuos buvo kreipiamasi balso žinute dalyvauti akcijoje sutiko net 75%, tuo tarpu į tuos vartotojus į kuriuos buvo kreiptasi el. paštu arba SMS pranešimu, sutiko tik 25%. Taipogi, vyrai yra labiau linkę dalyvauti akcijoje kadangi iš visų moterų prenumeratos pasiūlymą priėmė 26,67%, tuo tarpu iš visų vyrų pasiūlymą priėmė 56,67%.
 - Pasinaudojus logistine regresija paaiškėjo, kad akcijos pasirinkimas priklauso nuo komunikacijos kanalo, lyties, nuolaidos bei vartotojo reakcijos trukmės į gautą pasiūlymą užsisakyti paslaugą.
 - Atlikus pagrindinių komponentių analizę apimties komponentei priskirti duomenų bei eterio kintamieji, intensyvumo komponentei - skambučių ir SMS skaičius, lojalumo komponentei - nuolaida bei reakcijos trukmė į pranešimą. Šios analizės pagalba buvo sumažintas nagrinėjamų kintamųjų skaičius, kas leidžia sumažinti laiko resursus bei pagreitinti skaičiavimus.
 - Diskriminantinė analizė parodė, jog sudarytą modelį galima naudoti vartotojų elgsenos prognozavimui ateityje, kadangi sudaryto modelio tikslumas 73%.

LITERATŪRA

1. FAISON, J. and EDMUND, W. *The Neglected Variety Drive: A Useful Concept for Consumer Behavior* [interaktyvus]. 1977, *Journal of Consumer Research*, Vol. 4, No. 3, 172-175 [žiūrėta 2017-04-15]. Prieiga per: <https://academic.oup.com/jcr/article-abstract/4/3/172/1807368/The-Neglected-Variety-Drive-A-Useful-Concept-for?redirectedFrom=fulltext>
2. KOTLER, P., AMSTRONG, G. *Principles of Marketing* [interaktyvus]. 1994 Prentice Hall [žiūrėta 2017-04-15]. Prieiga per: https://books.google.lt/books/about/Principles_of_Marketing.html?id=Ecy_QgAACAAJ&redir_esc=y
3. SOLOMON, M. R. *Consumer Behaviour* [interaktyvus]. 1995, New Jersey: Prentice Hall [žiūrėta 2017-04-15]. Prieiga per: [https://books.google.lt/books?hl=lt&lr=&id=ajDiBAAAQBAJ&oi=fnd&pg=PP1&dq=Solomon,+M.+\(1995\)+%E2%80%9CConsumer+Behaviour%E2%80%9D&ots=ewYTJv-kMI&sig=pC3o61dSusehf-EV0Z76IP03w0o&redir_esc=y#v=onepage&q=Solomon%2C%20M.%20\(1995\)%20%E2%80%9CConsumer%20Behaviour%E2%80%9D&f=false](https://books.google.lt/books?hl=lt&lr=&id=ajDiBAAAQBAJ&oi=fnd&pg=PP1&dq=Solomon,+M.+(1995)+%E2%80%9CConsumer+Behaviour%E2%80%9D&ots=ewYTJv-kMI&sig=pC3o61dSusehf-EV0Z76IP03w0o&redir_esc=y#v=onepage&q=Solomon%2C%20M.%20(1995)%20%E2%80%9CConsumer%20Behaviour%E2%80%9D&f=false)
4. SCHIFFMAN, L., KANUK, L. *Consumer Behavior* [interaktyvus]. 2000, Pearson Education, 7th Edition [žiūrėta 2017-04-15]. Prieiga per: <http://www.zeljezno-polje.com/download/consumer-behavior-7th-edition.pdf>
5. BLACKWELL, R., MINIARD, P., ENGEL, J. *Consumer behavior* [interaktyvus]. 2006 Thomson/South-Western [žiūrėta 2017 04 20]. Prieiga per: https://books.google.lt/books/about/Consumer_Behavior.html?id=96TxAAAAMAAJ&redir_esc=y
6. STALLWORTH, P. *Consumer behaviour and marketing strategic* [interaktyvus]. 2008 [žiūrėta 2017 04 20]. Prieiga per: <http://research-methodology.net/consumer-buyer-behaviour-definition>
7. KOTLER, P. and KELLER, K.L. *Marketing Management* [interaktyvus]. 2011, London: Pearson Education [žiūrėta 2017-04-15]. Prieiga per: http://socioline.ru/files/5/283/kotler_keller_-_marketing_management_14th_edition.pdf
8. CHAND, S. *Consumer Behaviour: Meaning/Definition and Nature of Consumer Behaviour* [interaktyvus]. [žiūrėta 2017-04-15]. Prieiga per: <http://www.yourarticlelibrary.com/marketing/market-segmentation/consumer-behaviour-meaningdefinition-and-nature-of-consumer-behaviour/32301/>
9. EGAN, J. *Relationship Marketing: Exploring Relational Strategies in Marketing* [interaktyvus]. Ashford Colour Press Ltd, Gosport, Hampshire, 2011. [žiūrėta 2017-04-15]. ISBN: 978-0-273-73778-0. Prieiga per: <http://www.ssnpstudents.com/wp/wp-content/uploads/2015/02/Relationship-MaRketing.pdf>
10. EROGLU, S.A., MACHLEIT, K.A., DAVIS, L.M. *Empirical Testing of a Model of Online Store Atmospherics and Shopper Responses* [interaktyvus]. 2003, *Psychology & marketing*, 20(2), 139-150. New York, NY: Wiley, ISSN 0742-6046, ZDB-ID 2269338. [žiūrėta 2017-04-15]. Prieiga per: http://j.pelet.free.fr/publications/atmosphere/Empirical_Testing_of_a_Model_of_Online_Store_Atmospherics_and_Shopper_Responses.pdf
11. DAUGHERTY, B., EPP, S., HOSKINS, E., JOHNSON, C., MAILTON, J., McHUGH. L. *Service Marketing* [interaktyvus]. 2010. [žiūrėta 2017-04-15]. Prieiga per: <https://www.slideshare.net/Ncjohnson/final-servicescape-prez>
12. PENG, C., KIM, Y.G. *Application of the Stimuli-Organism-Response (S-O-R) Framework to Online Shopping Behavior* [interaktyvus]. 2014, *Journal of Internet Commerce*, Volume 13, Issue 3-4. Prieiga per: doi: <http://dx.doi.org/10.1080/15332861.2014.944437>
13. CHANG, Hyo-Jung, ECKMAN, Molly and YAN, Ruoh-Nan. *Application of the Stimulus-Organism-Response model to the retail environment: the role of hedonic motivation in impulse buying behavior* [interaktyvus]. 2011, *The International Review of Retail, Distribution and Consumer Research*, 21:3, 233-249. [žiūrėta 2017-04-15]. Prieiga per: https://www.researchgate.net/publication/239794427_Application_of_the_Stimulus-Organism-Response_model_to_the_retail_environment_the_role_of_hedonic_motivation_in_impulse_buying_behavior
14. SVILAR, M., CHAKRABORTY, A., KANIOURA, A. *Big Data Analytics in Marketing. From Hype to Real Help: Finding Valuable Consumer Insight in a Stream of Data* [interaktyvus]. *OR/MS Today* Oct., 40(5) 22, 2013. [žiūrėta 2017-04-15]. Prieiga per: <https://www.informs.org/ORMS-Today/Public-Articles/October-Volume-40-Number-5/Big-data-analytics-in-marketing>
15. STOICESCU, C. *Big Data, the perfect instrument to study today's consumer behavior* [interaktyvus]. 2015 [žiūrėta 2017-04-15]. Prieiga per: http://www.dbjournal.ro/archive/21/21_4.pdf
16. MARR, B. *How Big Data Analytics Can Improve Your Marketing* [interaktyvus]. 2016 [žiūrėta 2017-04-15]. Prieiga per: <http://data-informed.com/how-big-data-analytics-can-improve-your-marketing/>

17. VOJTOVIČ, S., NAVICKAS, V., GRUZAUSKAS, V. *Strategy of sustainable competitiveness methodology of real-time customers' segmentation for retail shops* [interaktyvus]. *Journal of Security and Sustainability Issues*, 2016, 5(4): 489–499. [žiūrėta 2017-05-15]. Prieiga per: doi: [http://dx.doi.org/10.9770/jssi.2016.5.4\(4\)](http://dx.doi.org/10.9770/jssi.2016.5.4(4))
18. SMITH, K. *How to Understand and Influence Consumer Behavior* [interaktyvus]. 2016 [žiūrėta 2017 04 20]. Prieiga per: <https://www.brandwatch.com/blog/how-understand-influence-consumer-behavior/>
19. PIATETSKY, G. *Gartner 2017 Magic Quadrant for Data Science Platforms: Gainers and Losers* [interaktyvus]. *Kdnuggets*, 2017. [žiūrėta 2017-04-15]. Prieiga per: <http://www.kdnuggets.com/2017/02/gartner-2017-mq-data-science-platforms-gainers-losers.html>
20. FISHER, R. A. *Statistical Methods for Research Workers* [interaktyvus]. 1973, Fourteenth Edition, New York: Hafner Publishing. [žiūrėta 2017 04 20]. Prieiga per: http://support.sas.com/documentation/cdl/en/procstat/63963/HTML/default/viewer.htm#procstat_univariate_section026.htm
21. JANILIONIS, V., MORKEVIČIUS, V., RAULECKAS, R. *Pavyzdinis metodologinis mokomasis studijų paketas. Statistinė kiekybinių duomenų analizė su SPSS IR STATA* [interaktyvus]. 2008 [žiūrėta 2017 04 20]. Prieiga per: http://www.lidata.eu/files/mokymai/stat/Statistika_galutine_redakcija.pdf
22. COX, N. R. *Estimation of the Correlation between a Continuous and a Discrete Variable* [interaktyvus]. 1974 *Biometrics*, 30, 171–178. [žiūrėta 2017 05 07]. Prieiga per: http://support.sas.com/documentation/cdl/en/procstat/66703/HTML/default/viewer.htm#procstat_corr_details03.htm
23. KASIUKEVIČIUS, V. IR DENAPIENĖ, G. *Statistikos taikymas mokslinių tyrimų analizėje* [interaktyvus]. 2008 [žiūrėta 2017 05 07]. Prieiga per: http://www.gerontologija.lt/files/edit_files/File/pdf/2008/nr_3/2008_176_180.pdf
24. SAS DOCUMENTATION [interaktyvus]. *One-Way ANOVA Tests* [žiūrėta 2017 04 20]. Prieiga per: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_npar1way_a0000000199.htm
25. ČEKANA VIČIUS, V., MURAU SKAS, G. *Statistika ir jos taikymai, II knyga*. 2002 ISBN 9955-491-16-7
26. ČEKANA VIČIUS, V., MURAU SKAS, G. *Statistika ir jos taikymai, I knyga*. 2000. ISBN 9986-546-93-1.
27. ČEKANA VIČIUS, V. *Taikomoji regresinė analizė socialiniuose tyrimuose* [interaktyvus]. Pavyzdinis metodologinis mokomasis studijų paketas [žiūrėta 2017 04 20]. Prieiga per: http://www.lidata.eu/index.php?file=files/mokymai/trast/trast.html&course_file=trast_2_1_4.html

PRIEDAS 1

Dispersinės analizės skaičiavimui naudotos lentelės

<i>Akcija</i>	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
Tarp sklaidos	2,9592E+15	1,27499E+13	3707663,806	84365107,71	34141221197	3649642,474
Viduje sklaidos	1,85603E+17	2,56E+14	32560458,29	70737818668	1,40228E+12	137096268,8
Viso	1,88562E+17	2,68827E+14	36268122,09	70822183776	1,43642E+12	140745911,3
Įtakos procentas sklaidai	1,57%	4,74%	10,22%	0,12%	2,38%	2,59%

<i>Lytis</i>	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
Tarp sklaidos	8,87265E+13	4,32752E+11	634006,3483	371435787,6	340415300	335940,4951
Viduje sklaidos	1,88E+17	2,68E+14	35634115,74	70450747988	1,43608E+12	140409970,8
Viso	1,89E+17	2,68827E+14	36268122,09	70822183776	1,43642E+12	140745911,3
Įtakos procentas sklaidai	0,05%	0,16%	1,75%	0,52%	0,02%	0,24%

<i>Kanalas</i>	DUOMENYS	ETERIS	NUOLAIDA	SKAMBUČIŲ SKAIČIUS	SMS SKAIČIUS	TRUKMĖ
Tarp sklaidos	1,54E+16	6,67023E+13	440140,3056	6962033857	1,02773E+12	4001865,771
Viduje sklaidos	1,73E+17	2,02E+14	35827981,79	63860149919	4,08688E+11	136744045,5
Viso	1,89E+17	2,68827E+14	36268122,09	70822183776	1,43642E+12	140745911,3
Įtakos procentas sklaidai	8,19%	24,81%	1,21%	9,83%	71,55%	2,84%

PRIEDAS 2

Wilcoxon'o rangų sumos lentelė (SMS skaičius)

Akcija	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Ne	608800	3.19484E11	3.17689E11	151734005	524776.008
Taip	434854	2.25124E11	2.26919E11	151734005	517699.559

Wilcoxon'o rangų sumos lentelė (SMS skaičius)

Lytis	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Moteris	521830	2.6536E11	2.72305E11	153886460	508517.711
Vyras	521824	2.79248E11	2.72302E11	153886460	535137.442

Wilcoxon'o rangų sumos lentelė (SMS skaičius)

Komunikacijos kanalas	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Balso	347883	1.7381E11	1.81535E11	145085372	499620.820
El. paštas	347886	7.81788E10	1.81536E11	145085685	224725.356
SMS	347885	2.92619E11	1.81536E11	145085581	841137.050

Wilcoxon'o rangų sumos lentelė (trukmė)

Akcija	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Ne	608800	3.38415E11	3.17689E11	151666048	555871.448
Taip	434854	2.06193E11	2.26919E11	151666048	474165.628

Wilcoxon'o rangų sumos lentelė (trukmė)

Lytis	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Moteris	521830	2.76106E11	2.72305E11	153817540	529111.344
Vyras	521824	2.68501E11	2.72302E11	153817540	514543.573

Wilcoxon'o rangų sumos lentelė (trukmė)

Komunikacijos kanalas	Imties tūris	Rangų suma	Tikėtina rangų suma	Standartinis nuokrypis	Rangų vidurkis
Balso	347883	1.66009E11	1.81535E11	145020394	477197.762
El. paštas	347886	2.00382E11	1.81536E11	145020706	575997.716
SMS	347885	1.78217E11	1.81536E11	145020602	512286.610

PRIEDAS 3

Programinis kodas

```
PROC SQL;
CREATE TABLE WORK.query AS
SELECT ID , KANALAS , LYTIS , NUOLAIDA , TRUKME , AKCIJA , ETERIS ,
SKAMBUCIU_SK , DUOMENYS , SMS_SK FROM _TEMPO.tele;
RUN;
QUIT;
PROC DATASETS NOLIST NODETAILS;
CONTENTS DATA=WORK.query OUT=WORK.details;
RUN;
PROC PRINT DATA=WORK.details;
RUN;

proc means data=work.query range sum mean min max std skew kurt;
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
run;

proc univariate data=work.query;
histogram duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
run;

data work.query2;
  set work.query;
  if lytis='M' then lytis_N=0;
  if lytis='V' then lytis_N=1;
  if kanalas='A' then kanalas_N=1;
  if kanalas='B' then kanalas_N=2;
  if kanalas='C' then kanalas_N=3;
  if akcija='N' then akcija_N=0;
  if akcija='T' then akcija_N=1;
  drop lytis kanalas akcija;
run;
/*Požymių priklausomumo lentelių tyrimas*/
proc freq data=work.query2 order=data;
  tables lytis_N*kanalas_N / chisq relrisk;
  exact pchi or;
run;

proc freq data=work.query2 order=data;
  tables akcija_N*kanalas_N / chisq relrisk;
  exact pchi or;
run;

proc freq data=work.query2 order=data;
  tables akcija_N*lytis_N / chisq relrisk;
  exact pchi or;
run;
/*Hipotezių tikrinimas*/
ods graphics on;
ods select Moments TestsForNormality ProbPlot;
proc univariate data=work.query2 normaltest;
  var kanalas_N;
  probplot kanalas_N / normal (mu=est sigma=est)
  square;
  label kanalas_N = 'Position Deviation';
  inset mean std / format=6.4;
run;
/*koreliacinė analizė (Spearmano koeficientas)*/
proc corr data=work.query2 spearman;
```

```

var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
WITH duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
run;

/*Surikiavom pagal lyti*/
proc sort data=work.query2;
by lytis_N;
run;

proc corr data=work.query2 spearman;
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
where lytis_N=1;
run;

/*Surikiavom pagal akcija*/
proc sort data=work.query2;
by akcija_N;
run;

proc corr data=work.query2 spearman;
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
where akcija_N=1;
run;

/*Dispersinė analizė*/
proc nparlway anova data=work.query2
plots=(wilcoxonboxplot medianplot);
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
class akcija_N;
run;

proc nparlway anova data=work.query2
plots=(wilcoxonboxplot medianplot);
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
class lytis_N;
run;

proc nparlway anova data=work.query2
plots=(wilcoxonboxplot medianplot);
var duomenys eteris nuolaida skambuciu_sk sms_sk trukme;
class kanalas_N;
run;
/* logistinė regresija */
proc logistic data=work.query2;
model akcija_N=kanalas_N lytis_N duomenys eteris nuolaida skambuciu_sk SMS_sk
trukme / rsq lackfit rsquare ctable;
run;

proc logistic data=work.query2;
class kanalas_N lytis_N;
model akcija_N= kanalas_N lytis_N duomenys eteris nuolaida skambuciu_sk
SMS_sk trukme /rsq rsquare;
run;

proc logistic data=work.query2;
class kanalas_N lytis_N;
model akcija_N= kanalas_N lytis_N nuolaida trukme / rsq rsquare;
run;

/*faktorine analize(Principal Component Analysis)*/
proc factor data=work.query2 out=factors1
simple
method=prin
priors=one

```

```

mineigen=1
nfact=4
rotate=varimax
round
flag=0.40;
var duomenys eteris nuolaida skambuciu_sk SMS_sk trukme;
run;

/*Imciu atskyrimas*/
proc surveyselect data=factors1 method=srs seed=43543 outall
  samprate=0.3 out=subsets;

* Values of Selected variable: 1 means selected for training set, 0 means
test set;
proc print data=subsets;
  title "Subsets Dataset";

data training;
  set subsets;
  if Selected=1;

proc print;

data test;
  set subsets;
  if Selected=0;

proc print;

/*Diskriminantinė analize*/
proc discrim data=training method=npair pool=yes
  k=100 testdata=test testout=plotp testoutd=plotd
  short noclassify crosslisterr;
  class Akcija_N;
  var Factor1 Factor2 Factor3;
run;

```