



**KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

Ignas Kizelevičius

**VAISTINĖS KLIENTŲ LOJALUMO PROGNOZAVIMO
MODELIS**

Baigiamasis magistro projektas

Vadovai

Doc. dr. Kristina Šutienė

Doc. dr. Aistė Dovalienė

KAUNAS, 2017

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS**

**VAISTINĖS KLIENTŲ LOJALUMO PROGNOZAVIMO
MODELIS**

Baigiamasis magistro projektas
Didžiųjų verslo duomenų analitika (kodas 621G12002)

Vadovai

Doc. dr. Kristina Šutienė

Doc. dr. Aistė Dovalienė

Recenzantai

Doc. dr. Agnė Gadeikienė

Lekt. dr. Evaldas Vaičiukynas

Projektą atliko

Ignas Kizelevičius

KAUNAS, 2017



KAUNO TECHNOLOGIJOS UNIVERSITETAS

Matematikos ir gamtos mokslų

(Fakultetas)

Ignas Kizelevičius

(Studento vardas, pavardė)

Didžiųjų verslo duomenų analitika, 621G12002

(Studijų programos pavadinimas, kodas)

„Vaistinės klientų lojalumo prognozavimo modelis“

AKADEMINIO SAŽINGUMO DEKLARACIJA

20 17 m. birželio 2 d.
Kaunas

Patvirtinu, kad mano, **Ignas Kizelevičiaus**, baigiamasis projektas tema „Vaistinės klientų lojalumo prognozavimo modelis“ 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

ĮŽANGA.....	7
1. LITERATŪROS APŽVALGA.....	9
1.1. Klientų lojalumo samprata.....	9
1.2. Klientų elgseną apibūdinantys kintamieji ir jų sudarymas.....	12
1.3. Klientų lojalumo prognozavimo modeliai ir jų taikymas.....	16
1.4. Literatūros šaltinių analizės apibendrinimas	20
LITERATŪROS SĄRAŠAS	21

Kizelevičius, Ignas. Vaistinės klientų lojalumo prognozavimo modelis. Magistro baigiamasis projektas / vadovai doc. dr. Kristina Šutienė, doc. dr. Aistė Dovalienė. Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas.

Mokslo kryptis ir sritis: Matematika(01 P), fiziniai mokslai

Reikšminiai žodžiai: *Santykių su klientais valdymas, lojalumo prognozavimas, mašininio mokymo algoritmai.*

Kaunas, 2017. 23 p.

SANTRAUKA

Vienas iš pagrindinių įmonės, siekiančios sėkmingai valdyti santykius su klientais, tikslų yra lojalių klientų išsaugojimas (angl. *customer retention*). Pasak mokslinės literatūros, lojalūs klientai yra įmonės turtas, generuojantis didžiąją dalį pajamų. Taip pat, lojalių klientų išsaugojimas reikalauja gerokai mažesnių finansinių resursų, nei naujų klientų pritraukimas. Šiame darbe yra nagrinėjamas vaistinių tinklo klientų lojalumo atvejis. Šio darbo tikslas yra sudaryti modelį, skirtą vaistinės klientų lojalumui prognozuoti. Šiuo atveju, nėra tiksliai apibrėžta, kada klientas pasitraukia iš lojalių klientų rato. Šio darbo pirmame skyriuje atlikta mokslinių tyrimų, klientų lojalumo tematika, apžvalga. Antrame skyriuje suformuota klientų lojalumo tyrimo metodologija. Apibendrinus naujausiose mokslinėse publikacijose siūlomus lojalumo apibrėžimus, šiame darbe pasiūlytas apibendrintas nelojalių ir iš dalies nelojalių klientų nustatymo būdas, kuris leidžia įvertinti visą kliento santykių su įmone laikotarpį ir nustatyti pasitraukimą iš lojalių klientų rato pagal kliento individualią elgseną. Taip pat, sudarytas elgseną apibūdinančių kintamųjų rinkinys, susidedantis iš kliento demografinių, agreguotų transakcinių, išleistų pinigų kiekio laiko eilutės ir jos diskrečiosios bangelių transformacijos koeficientų. Trečiojoje dalyje, naudojantis suformuota metodologija, atlikta vaistinės vadovybės suteikto duomenų rinkinio analizė. Pritaikyti logistinės regresijos, atsitiktinių miškų ir atraminių vektorių mašinių algoritmai vaistinės klientų lojalumui prognozuoti. Atliktas prognozavimo eksperimentų stratifikuotas kryžminis validavimas ir palyginti modelių prognozės rezultatai, gauti naudojant nesubalansuotą ir subalansuotą mokymo imtis, skirtingus nepriklausomų kintamųjų rinkinius ir skirtingus mašininio mokymo algoritmus. Galiausiai, įvertinti prognozavimo tikslumo rezultatai ir išrinktas tiksliausias vaistinės klientų lojalumo prognozavimo modelis.

Kizelevičius, Ignas. Churn Prediction of Pharmacy Customers: Master's thesis in Business big data analytics / supervisors dr. assoc. prof. Kristina Štutenė, dr. assoc. prof. Aistė Dovalienė. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: Mathematics (01 P), physical sciences

Key words: *Customer relationship management, churn prediction, machine learning.*

Kaunas, 2017. 23 p.

SUMMARY

Customer relationship management is a comprehensive strategy to build, manage and strengthen long lasting relationships between business and its customers. One of the main tasks of this strategy is customer retention. The importance of this task is obvious. Firstly, loyal customers are considered to have the highest lifetime value because they generate the biggest portion of company's income. Secondly, acquiring new customers is much more expensive than retaining the loyal ones. Finally, it takes significant amount of time and financial resources to raise new customers' loyalty and lifetime value. In this paper we consider a non-contractual customer relationship setting of a retail pharmacy. In such case, it is a complex task to define whether the customer has churned or not. In section 1 we perform a literature review to get acquainted with previous works regarding the customer loyalty and customer churn prediction. In section 2 we construct a customer churn analysis methodology starting with a raw customer transactional database. Generalizing customer churn definitions from the literature, we propose improved definitions of complete and partial churn, considering the behavioural characteristics of an individual customer. We also construct a set of behavioural characteristics including demographic, aggregated transactional, time series and time series discrete wavelet transformation coefficients. In section 3 we apply the proposed methodology on a retail pharmacy customer transactional database. We adapt logistic regression, support vector machines and random forest machine learning algorithms to build churn prediction models. We perform a stratified cross validation to compare prediction models of original and down sampled training sets, different sets of independent variables and different machine learning algorithms. Finally, we evaluate the predictive accuracy of performed prediction experiments and select the most efficient pharmacy customer churn prediction model.

IŽANGA

Pastaruoju metu, sparčiai augant informacinėms technologijoms, augo ir verslo galimybės rinkti, kaupti ir analizuoti didelius duomenų kiekius. Galimybė analizuoti kiekvieno kliento transakcinių duomenų istoriją pastūmėjo verslą labiau orientuoti į klientą, o ne produktą.

Santykių su klientais valdymas yra visapusiška strategija kurti, valdyti ir stiprinti lojalius ir ilgai besitęsiančius santykius su klientais. Ši strategija yra visuotinai pripažinta ir plačiai taikoma telekomunikacijų, bankinėje ir draudimo, didmeninės bei mažmeninės prekybos srityse. Vienas iš pagrindinių strategijos tikslų yra lojalių klientų išsaugojimas (angl. *customer retention*). Pasak apžvelgtos mokslinės literatūros, lojalūs klientai yra įmonės turtas, generuojantis didžiąją dalį pajamų. Taip pat, lojalių klientų išsaugojimas reikalauja gerokai mažesnių finansinių resursų, nei naujų klientų pritraukimas. Pagaliau, netgi užtikrinus naujų klientų pritraukimą, jų integravimas į įmonės lojalių klientų ratą yra brangus procesas tiek laiko, tiek finansiniu atžvilgiu. Todėl, vis daugiau įmonių siekia sudaryti sprendimų, kurie užtikrintų lojalių klientų išsaugojimą, modelį.

Kai įmonė ir klientą sieja sutartimi apibrėžti santykiai, šios sutarties nutraukimas yra tapatus pasitraukimui iš lojalių klientų rato. Priešingu atveju, nėra aišku, kada klientas perėjo iš lojalių klientų rato į nelojalių klientų ratą. Taip pat, kai santykiai nėra apibrėžti paslaugų teikimo sutartimi, kliento pasitraukimas iš lojalių klientų rato gali įvykti ne tik dėl visiško neaktyvumo, bet ir dėl žymaus aktyvumo sumažėjimo.

Klientų lojalumo prognozavimas (angl. *churn prediction*) – tai įrankis atpažinti klientų elgsenos požymius ir iš anksto įvertinti, kurie klientai yra labiau linkę pasitraukti iš lojalių klientų rato. Klientų lojalumo prognozavimas gali būti atliekamas naudojant įvairius metodus. Atlikta mokslinės literatūros apžvalga rodo, kad mašininio mokymo algoritmai tiksliau prognozuoja klientų lojalumą už matematinius tikimybinus modelius.

Šiame darbe nagrinėsime mažmeninės prekybos vaistinių tinklo atvejį. Darbo tikslas – sudaryti modelį, skirtą vaistinės klientų lojalumui prognozuoti. Šiam tikslui pasiekti iškelti šie uždaviniai:

- Atlikti mokslinės literatūros, skirtos klientų lojalumo prognozavimui ir tyrimui, apžvalgą.
- Sudaryti vaistinės klientų lojalumo tyrimo schemą ir ja naudojantis atlikti vaistinės vadovybės suteikto duomenų rinkinio tyrybą.
- Sudaryti taisykles klientų lojalumui apibrėžti.
- Parinkti metodiką kontekstinių išskirčių nustatymui.
- Sudaryti rinkinį charakteristikų, skirtų klientų elgsenai apibūdinti, ir įvertinti jų svarbą prognozuojant klientų lojalumą.

- Sudaryti rinkinį prognozavimo modelių ir atlikti prognozavimo rezultatų palyginamąją analizę.

Šis darbas padalintas į tris pagrindines dalis. Pirmojoje dalyje atliksime mokslinių straipsnių, vartotojų lojalumo tyrimų tematika, apžvalgą. Antrojoje dalyje, remiantis atlikta mokslinių tyrimų apžvalga, suformuosime metodologiją vaistinės klientų duomenų paruošimui, išskirčių atpažinimui, klientų lojalumo apibrėžimui ir prognozavimo modelio sudarymui. Trečiojoje dalyje, remiantis suformuota metodologija, atliksime duomenų rinkinio, gauto iš vaistinės vadovybės, analizę, bei iš suformuotų prognozavimo modelių išrinksime geriausią.

Šio tyrimo pagrindu skaitomas pranešimas 58-oje Lietuvos matematikų draugijos konferencijoje.

1. LITERATŪROS APŽVALGA

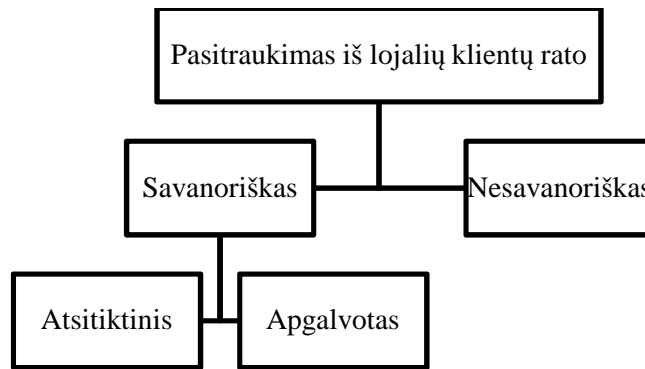
Per kelis pastaruosius dešimtmečius paskelbta daug mokslinių straipsnių vartotojų lojalumo tyrimų tematika. Susidomėjimas šia sritimi itin išaugo pastaruosiu metu, populiarėjant didžiųjų duomenų analitikos mokslui. Šiame darbe keliamas tikslas – sukurti vaistinės klientų lojalumo prognozavimo modelį. Šioje darbo dalyje apžvelgsime kitų autorių atliktus mokslinius tyrimus klientų lojalumo tyrimo tematika.

1.1. Klientų lojalumo samprata

Klientų lojalumas literatūroje apibrėžiamas įvairiais būdais. Vyrauja dviejų tipų lojalumas: požiūrio lojalumas ir elgsenos lojalumas [1, 2]. Požiūrio lojalumas reiškia, kad klientas naudojami įmonės paslaugomis, jomis yra patenkintas ir rekomenduoja draugams bei pažįstamiems žmonėms, kitaip sakant, yra įmonės „advokatas“. Elgsenos lojalumas – nuolatinė pakartotinė pirkimo elgsena, kuomet pirkimo dažnumas ir kiekis yra pastovus.

Norint nagrinėti požiūrio lojalumą, reikėtų įvertinti individualių klientų patirtį naudojantis vaistinės teikiamomis paslaugomis bei parduodamais produktais, ištirti jų nuomonę įvairiomis apklausomis. Vartotojų požiūriui nustatyti gali būti naudojama informacija apie kliento atsiliepimus, skambučius į klientų aptarnavimo centrą, socialinių tinklų duomenys. Duomenys apie kliento interneto naršymo istoriją taip pat gali būti panaudoti požiūriui nustatyti ir kliento poreikiams suprasti [3, 9]. Šiame darbe apsiribosime tik klientų elgsenos lojalumo tyrimais.

Kraljevičius ir Gotovacas [6] suskirstė klientų pasitraukimą į savanorišką ir nesavanorišką. Jo sudaryta lojalumo schema pavaizduota paveiksle 1.1. Nesavanoriškas pasitraukimas įvyksta, kai įmonė nusprendžia nutraukti paslaugų teikimo sutartį su klientu. Priešingai, savanoriškas pasitraukimas įvyksta, kai klientas savo iniciatyva nutraukia paslaugų teikimo sutartį. Atsitiktinis pasitraukimas – tai savanoriško pasitraukimo atšaka, nes nei įmonė, nei klientas, iš anksto neplanavo nutraukti paslaugų teikimo sutarties. Nutraukimas galėjo įvykti dėl tam tikrų pokyčių kliento privačiame gyvenime, pvz., pasikeitė gyvenamoji vieta, pasikeitė darbovietė ar finansinė situacija. Kita savanoriško pasitraukimo iš lojalių klientų rato atšaka – apgalvotas pasitraukimas, nes įvyksta, kai klientas pasitraukia dėl technologinių (klientas ieško naujesnių ar labiau pažengusių technologijų), ekonominių (jautrumas kainai), aptarnavimo (paslaugų kokybė), socialinių ar patogumo priežasčių. Apgalvotas pasitraukimas yra pagrindinė problema, kurią bandoma išspręsti klientų lojalumo prognozavimu.



1.1 paveikslas. Vartotojų lojalumo schema [6]

Kraljevičius ir Gotovacas [6] išskyrė telekomunikacijų bendrovės klientus į *pre-paid* (sumoka už paslaugas prieš jomis naudodamiesi) ir *post-paid* (sumoka už paslaugas po fakto). *Post-paid* klientai moka už paslaugas kas kiekvieną sutartą laiko periodą, dažniausiai mėnesį, todėl jų pasitraukimo iš lojalių klientų rato momentas sutampa su kliento atsijungimu nuo telekomunikacijų tinklo. Priešingai, *pre-paid* vartotojai gali atsijungti nuo tinklo gerokai vėliau nei įvyksta tikrasis pasitraukimas iš lojalių klientų rato. Dėl to yra sudėtingiau nustatyti *pre-paid* tipo klientų lojalumo praradimo faktą. Šiame darbe nagrinėjami vaistinės klientai taip pat yra *pre-paid* tipo vartotojai, nes už prekes sumoka prieš jas vartodami.

Apibrėžti, kas yra lojalus klientas, kai jo su įmone nesieja paslaugų teikimo sutartimi apibrėžtas ryšys, yra sudėtinga užduotis. Kai klientą ir įmonę sieja paslaugų teikimo sutartis, jos nutraukimas tuo pačiu reiškia ir kliento pasitraukimą iš lojalių klientų rato. Apibrėžiant elgsenos lojalumą, kai kliento ir įmonės nesieja paslaugų teikimo sutartis, didžiausia problema yra nustatyti, kada klientas palieka lojalių klientų ratą. Šiame skyriuje apžvelgsime, kaip tai atliko kiti mokslinių darbų autoriai.

Liu, Kjao ir Ksu [5] išskyrė papildomą pasitraukusių iš lojalių klientų rato vartotojų grupę – vertingus klientus. Tai yra klientai, kurie ypatingai svarbūs įmonei dėl jų išskirtinio elgesio, dažnų ir pelningų sandorių. Autorius sudarė klasifikavimo modelį, kuris prognozuoja tris lojalumo klases: įprastų pasitraukusių klientų, ypač svarbių pasitraukusių klientų ir klientų išlikusių lojaliais.

Tamadonis, Stakovičius ir Evingas [4] nagrinėjo klientų pasitraukimą iš lojalių klientų rato ribotame laiko periode, t.y. kaip laikiną kliento būseną, neatsižvelgiant į galimybę klientui sugrįžti į lojalių klientų ratą. Autoriai padalino klientų elgsenos istoriją į du laikotarpius: kalibravimo ir prognozavimo. Lygindami kliento aktyvumą tarp laikotarpių jie nustatė, kurie klientai yra aktyvūs, o kurie – neaktyvūs. Klientas buvo priskiriamas lojalių klientų ratui, jei atliko bent po vieną sandorį su įmone abiem laikotarpiais. Priešingai, jei klientas atliko bent vieną sandorį kalibravimo laikotarpiu, tačiau neatliko nei vieno sandorio prognozavimo laikotarpiu, jis buvo priskiriamas pasitraukusių iš lojalių klientų rato klientų aibei.

Mažmeninės prekybos verslo srityje vartotojai dažniausiai nėra susieti su įmone jokia paslaugų teikimo sritimi. Todėl jie gali laisvai rinktis paslaugų teikėją ir lengvai pereiti iš vienos įmonės lojalių klientų ratą į kitos. Šiuo atveju, tikslinga įvesti papildomą lojalumo apibrėžimą – dalinai lojalius klientus. Tai klientai, kurie vis dar naudojami įmonės paslaugomis, tačiau jų elgsena, paslaugų naudojimosi intensyvumas, yra pakitęs į blogąją, iš paslaugas teikiančios įmonės perspektyvos, pusę.

Bukinksas ir Poelis [3] pasiūlė dalinai nelojalių klientų apibrėžimą. Šie autoriai kiekvienu laiko momentu T apibrėžė lojalių ir nelojalių klientų aibes. Klientas patenka į lojalių klientų aibę, jeigu 5 mėnesius prieš laiko momentą T buvo tenkinamos šios sąlygos:

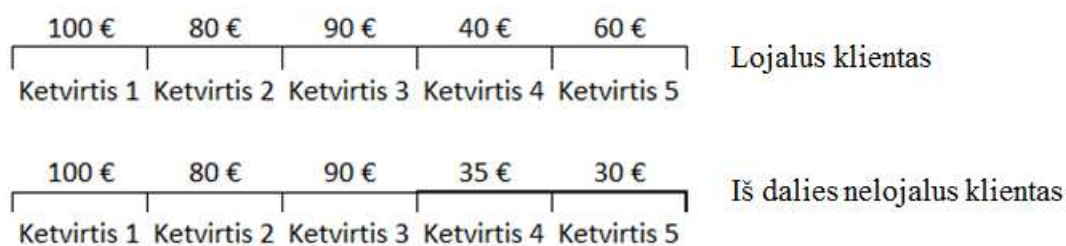
- a) sandorių su įmone dažnumas (sandorių kiekis per 5 mėnesius) yra didesnis už vidutinį;
- b) laikotarpio tarp gretimų sandorių variacijos koeficientas (CV – santykis tarp vidutinio sandorių kiekio per mėnesį ir sandorių per mėnesį kiekio standartinio nuokrypio) yra mažesnis už vidutinį.

Taigi, lojalūs klientai yra klientai, kurie atlieka sandorius dažnai ir jų paslaugų naudojimosi elgesys yra pastovus. Visi likę klientai yra laikomi nelojaliais. Bukinksas ir Poelis [3] tarė, kad klientas tapo dalinai nelojalus laiko momentu T , jei per 5 ateinančius mėnesius nuo momento T perėjo iš lojalių klientų aibės į nelojalių klientų aibę.

Clemente-Ciscaras [7] išplėtė Bukinkso ir Poelio [3] apibrėžimą. Jie sudarė vartotojų, kurių su įmone nesieja jokie sutartimi apibrėžti ryšiai, lojalumo apibrėžimo metodologiją. Autoriai suformavo vartotojų, dalinai pasitraukusių iš lojalių klientų rato, apibrėžimą. Kiekvienu kliento elgsenos stebėjimo momentu T , autoriai apibrėžė stebėjimo ($T - k$) ir pokyčių raidos ($T + k$) periodus. Autoriai nustatė, kad jų duomenų rinkiniui labiausiai tinka parametro k reikšmė $k = 5$. Jie pasiūlė rinkinį charakteristikų ribai, kada klientas pereina iš lojalių klientų rato į nelojalių klientų ratą, nustatyti. Atlikta analizė parodė, kad šios ribos, tarp lojalių ir nelojalių klientų, nustatymui geriausiai tinka vidutinė išleista pinigų suma ir apsipirkimų skaičiaus variacijos koeficientas. Jei kliento charakteristikos stebėjimo ir pokyčių raidos laikotarpiais yra virš nustatytų ribų, klientas buvo laikomas lojaliu. Tačiau, jei stebėjimo laikotarpiu yra virš ribos, o pokyčių raidos laikotarpiu – žemiau nustatytos ribos, klientas buvo laikomas pasitraukęs iš lojalių klientų rato į dalinai lojalių klientų ratą.

Migelis, Poelis, Camanho ir Cunha [8] publikacijoje pasiūlė dalinio klientų pasitraukimo iš lojalių klientų rato apibrėžimą. Autoriai susiejo dalinį pasitraukimą su klientų išleistos pinigų sumos mažėjimu. Jie sugrupavo klientų išlaidas į ketvirčius (trijų mėnesių laikotarpius). Klientas buvo laikomas iš dalies pasitraukusiu iš lojalių klientų rato, jei kitu ir visais likusiais aktyvumo istorijos ketvirčiais nieko nepirko arba pirkiniams išleido 40% mažiau nei ankstesniais

ketvirčiais. Lojalus ir iš dalies nelojalus klientų išleistų pinigų pavyzdys pavaizduotas 1.2 paveiksle.



1.2 paveikslas. Lojalus ir iš dalies nelojalus klientų išleistos pinigų per mėnesį pavyzdys [8]

Panašų lojalių ir dalinai nelojalių klientų atskyrimą suformulavo ir Migelis, Camanho, Cunha [10]. Jie pasitraukusiais klientais laikė tuos vartotojus, kurie pradėdant stebimu laikotarpiu visais ateinančiais kliento istorijos ketvirčiais išleido 30% mažiau pinigų.

1.2. Klientų elgseną apibūdinantys kintamieji ir jų sudarymas

Vienas iš svarbiausių uždavinių, sudarant klientų lojalumo prognozavimo modelį, yra nepriklausomų kintamųjų parinkimas. Nepriklausomi kintamieji turi atvaizduoti kiek įmanoma detalesnę informaciją apie klientą ir jo elgseną. Šiame skyriuje apžvelgsime, kaip yra formuojami ir parenkami nepriklausomi kintamieji mokslinėje literatūroje.

Klientų duomenų bazė dažniausiai būna sudaryta iš dviejų pagrindinių duomenų lentelių.

Pirmoji lentelė – statistiniai duomenys apie klientą. Juos sudaro geografiniai, demografiniai, gyvenimo būdo ir sociografiniai duomenys apie klientą. Demografiniai duomenys – tai kliento amžius, lytis, šeimos narių skaičius ir panašūs rodikliai. Duomenys apie kliento gyvenimo būdą suteikia informaciją apie kliento įpročius, kultūrinius pomėgius (pvz., prenumeruojami žurnalai), laisvalaikio hobius. Geografiniai duomenys – gyvenamoji vieta, darbo vietos adresas ir kiti duomenys susiję su geografine kliento padėtimi [9].

Antrojoje lentelėje – transakciniai duomenys apie kliento santykius su įmone. Tai gali būti informacija apie kliento atliktus sandorius, kliento apie įmonę palikti atsiliepimai, internetinio tinklalapio naršymo duomenys. Iš transakcinių duomenų, daugelyje apžvelgtų mokslinės literatūros publikacijų, yra formuojami RFM metodo kintamieji [3, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. RFM analizė yra plačiai paplitęs metodas identifikuoti klientų elgseną. Metodo pavadinimas susideda iš trijų elgsenos dedamųjų [13]:

- R – naujumas (angl. *recency*), apibrėžia laiko periodą nuo paskutinio kliento sandorio su įmone;
- F – dažnumas (angl. *frequency*), apibrėžia įvykdytų sandorių ar nupirktų produktų kiekį per nustatytą laiko periodą;

- c) M – piniginis kriterijus (angl. *monetary*), apibrėžia kliento elgseną pinigine išraiška. Tai gali būti paskutiniu pirkimu išleista pinigų suma, visu kliento aktyvumo periodu išleista pinigų suma ar vidutinė išleista pinigų suma per laikotarpį.

Remiantis RFM modelio metodika, visi klientai yra suskirstomi į penkias grupes pagal naujumo, dažnumo ir piniginio kriterijaus kintamuosius. Toks skirstymas leidžia kiekvieną klientą įvertinti trijų skaitmenų indeksu. Kuo šis indeksas mažesnis, tuo didesnė tikimybė, kad klientas pasitrauks iš lojalių klientų rato. Atsiradus sudėtingiems mašininio mokymo algoritmams, RFM idėja buvo panaudota formuojant prognozavimo modelių nepriklausomus kintamuosius. Buvo nustatyta, kad RFM metodu suformuoti nepriklausomi kintamieji yra vieni svarbiausių kintamųjų lojalumo prognozavimo modeliuose [18].

Straipsnio [8] autorius atliko vienos iš Europoje esančių mažmeninės prekybos įmonės klientų lojalumo analizę. Lojalumo prognozavimo modelyje autorius naudojo pagal RFM modelį suformuotus nepriklausomus kintamuosius. Naujumą autorius apibrėžė, kaip dienų skaičių nuo paskutinio pirkimo iki analizuojamo laikotarpio pabaigos. Dažnumą – kaip vidutinį kliento atliktų užsakymų skaičių per ketvirtį metų. Piniginio kriterijaus dedamoji buvo apibrėžta, kaip vidutinė kliento išleista pinigų suma per ketvirtį metų.

Priklausomai nuo verslo srities ir analizės uždavinių tikslinga kiekvienam RFM kintamajam suteikti svorinį koeficientą, kuris pabrėžtų kintamojo svarbumą. Straipsnio [13] autoriaus analizė buvo orientuota į juridinius banko klientus. Analizės tikslas buvo nustatyti klientų pelningumą pagal gildiją, t.y. kokios rūšies paslaugomis ar produktais prekiauja juridinis asmuo. Šiuo atveju, didžiausias svoris buvo suteiktas pelno (*monetary*) kintamajam, vidutinis – dažnumo (*frequency*), ir mažiausias – naujumo (*recency*) kintamajam.

Vienas iš RFM modelio naujumo dedamosios kintamųjų yra laikas tarp gretimų kliento ir įmonės sandorių. Šis kintamasis literatūroje vadinamas *inter-purchase time* (IPT). Vyrauja hipotezė, kad klientai, kurie dažniau lankosi parduotuvėje, t.y. praeina mažiau laiko tarp gretimų apsilankymų, yra mažiau linkę pasitraukti iš lojalių klientų rato [10].

Klientų lojalumo prognozei naudojant mašininio mokymo algoritmus svarbu apsibrėžti kuo įvairesnių kintamųjų. Mokslinėje literatūroje randama daug RFM kintamųjų sudarymo variacijų. Bukinksas ir Poelis [3] atliko vienos didžiausių pasaulyje įmonių prekiaujančios greito vartojimo produktais klientų analizę. Duomenys apie klientus įmonėje buvo renkami naudojantis lojalumo kortelėmis. Autorius, atlikęs literatūros apie nepriklausomus kintamuosius, naudojamus lojalumo prognozavime, analizę, suformavo 61 kintamųjų rinkinį, kurį įdiegė į prognozavimo modelį. Dalis modelio kintamųjų buvo RFM metodu suformuoti kintamieji:

- naujumas – praėjusių nuo paskutinio pirkimo dienų skaičius, vidutinis dienų skaičius tarp gretimų pirkimų (IPT), IPT standartinis nuokrypis, IPT variacijos koeficientas (santykis tarp IPT standartinio nuokrypio ir vidutinio IPT);
- dažnumas – apsilankymų parduotuvėje skaičius per visą kliento istoriją (skaičiuojami tik apsilankymai, kurių metu buvo nupirkta(os) prekė(s)), apsilankymų skaičius per mėnesį, apsilankymų skaičius per paskutinį mėnesį, apsilankymų skaičius per paskutiniąją savaitę;
- piniginis kriterijus – bendras kliento išleistų pinigų kiekis, per mėnesį išleistų pinigų kiekis, apsipirkimų, kurių metu buvo išleista daugiau nei vidutiniškai skaičius.

Literatūroje galima rasti įžvalgų apie klientų lojalumo priklausomybę nuo jų aktyvumo laikotarpio ilgio. Šis kintamasis vadinamas santykių su klientais trukmė (angl. *length of relationship (LoR)*) [5, 14, 16, 18]. Liu, Kjao ir Ksu [5] nagrinėjo Kinijos telekomunikacijų bendrovės klientų duomenis. Juos sudarė 5 metų laikotarpio klientų skambučių ir išlaidų duomenys. Pastebėta, kad klientai, kurie įmonės paslaugomis naudojami 5 metus, gerokai mažiau linkę pasitraukti iš lojalių klientų rato. Priešingai, klientai, kurie naudojami paslaugomis vos kelis mėnesius yra labiau linkę nutraukti sutartį. Ju, Guo ir Hujangas [19] bei Hadijis, Sifas, Dračenas, Turau, Kerstingas ir Bukadžas [20] atliko kompiuterinių žaidimų žaidėjų lojalumo analizę. Prognozavimo modelyje buvo panaudoti demografiniai duomenys, žaidėjų prisijungimo duomenys, transakciniai mokėjimų už virtualias prekes duomenys ir internetinio žaidimų puslapių naršymo duomenys. Abiejų publikacijų autorių pastebėta, kad šioje srityje įprastas atvejis, jog žaidėjai pasitraukia iš lojalių klientų rato po keleto dienų ar netgi po kelių sužaistų žaidimo partijų.

Modeliuojant klientų lojalumą su dideliu kiekiu kintamųjų yra rizika pabloginti prognozės rezultatus dėl per didelio kiekio kintamųjų. Šiai problemai spręsti naudojami įvairūs kintamųjų atrinkimo būdai. Kintamųjų atrinkimo tikslas – prognozės modelyje naudoti tik kintamuosius, kurie suteikia reikšmingą kiekį informacijos apie priklausomą kintamąjį. Keli populiariausi kintamųjų atrinkimo būdai: *Chi-kvadrato* metodas [12], *forward* ir *backward* atrinkimo metodai [21]. Abasimeris, Setakas ir Soruras [12] viso suformavo 171 nepriklausomą kintamąjį, tačiau po kintamųjų atrinkimo prognozės modelyje naudojo tik 20 svarbiausių. Rungè, Gao, Garcinas ir Faltingsas [21] analizavo mobiliųjų žaidimų klientų lojalumą. Iš viso autoriaus nagrinėjamuose duomenyse buvo virš kelių šimtų kintamųjų. Naudojantis *forward feature selection* metodu buvo pašalinti nereikšmingi kintamieji. Likusius reikšmingus kintamuosius sudarė sužaistų partijų bei prisijungimų laiko eilučių duomenys, žaidėjo tikslumas, žaidėjo lygis, išsiustų pakvietimų skaičius, virtualių dienų praleistų žaidime skaičius, paskutinio pirkimo metu išleista pinigų suma, dienų skaičius nuo paskutinio pirkimo.

Modeliuojant klientų lojalumo prognozę mašininio mokymo algoritmais, gaunami ne tik priklausomo kintamojo prognozės rezultatai, bet ir nepriklausomų kintamųjų svarbos rodikliai. Klientų lojalumo prognozavimo modeliuose, svarbiausi kintamieji, pasak mokslinės literatūros, yra RFM [3, 9, 14, 15, 18, 20] ir LoR [9, 14, 15]. Mažiau svarbus, tačiau reikšmingas kintamasis – IPT [3, 9, 14, 20].

Praktikoje įprasta analizuoti transakcinius duomenis, prieš tai juos transformavus į statiškus, pvz. pirkimų skaičius per laikotarpį, išleista pinigų suma per laikotarpį. Taip prarandamas didelis kiekis informacijos apie kliento elgseną. Vienas iš paprasčiausių būdų pagerinti situaciją yra agreguoti duomenis ne pagal laiko periodą (pvz. pardavimų skaičius per mėnesį), bet per visą kliento aktyvumo istoriją [11]. Šiek tiek efektyvesnis būdas atvaizduoti transakcijų dinamiškumą – padalinti kliento aktyvumo laikotarpį į kelis periodus ir, apskaičiavus tolydžių kintamųjų reikšmių skirtingais periodais santykius, juos naudoti kaip papildomus nepriklausomus kintamuosius prognozės modelyje [6]. Daug skaičiavimo resursų reikalaujantis, tačiau pats efektyviausias būdas yra naudoti mėnesines ar retesnes tolydžių kintamųjų laiko eilutes. Chenas, Fanas ir Sunas [22] savo pasiūlytame lojalumo prognozavimo modelyje naudojo ne tik statinius ar agreguotus per visą kliento laikotarpį duomenis, tačiau ir dinaminis duomenis, pvz., kiekvieną mėnesį kliento išleistą pinigų sumą. Naudojant modelį su dinaminiais kintamaisiais, prognozė buvo tikslesnė už modelius, kuriuose buvo naudojami tik statiški kintamieji. Chenas ir Fanas [23] naudojo demografinių kintamųjų rinkinį ir RFM kintamuosius. Autoriai palygino prognozės rezultatus, kai modeliai mokomi tik demografiniais duomenimis arba tik laiko eilučių kintamaisiais ir bendrą modelį, kuris buvo mokomas demografiniais ir laiko eilučių kintamaisiais. Pastarasis modelis parodė geriausias prognozės rezultatus.

Dažnai literatūroje sutinkamas nepriklausomų kintamųjų formavimo būdas – kintamieji suformuoti atsižvelgiant į kategorinius duomenis. Kai kintamasis turi nedidelį kiekį unikalių kategorijų, tikslinga apskaičiuoti RFM reikšmes kiekvienai kategorijai atskirai [3, 4, 10]. Jei kategorijų yra ypač daug, tokie kintamieji vadinami didelio kardinalumo kintamaisiais. Moeyersomsas ir Martensas [24] susitelkė į didelio kardinalumo duomenų panaudojimą klientų duomenų analizei patobulinti. Didelio kardinalumo duomenys retai kada naudojami analizei, nes sukelia sunkumų. Kardinalūs kintamieji, tai kintamieji, kurie yra kategorijų arba vardų skalės ir turi daug skirtingų reikšmių. Tokie kintamieji, kaip kliento gyvenamasis miestas, pašto kodas, pavardė, turi labai daug skirtingų reikšmių ir yra laikomi didelio kardinalumo kintamaisiais. Kiekvienai reikšmei įvedant pseudokintamąjį, jų skaičius išaugtų iki didelio kiekio kintamųjų matricos ir norint panaudoti tokį transformuotą duomenų rinkinį prognozavimo modelyje reikėtų milžiniškų skaičiavimo resursų. Moeyersomsas ir Martensas [24] pasiūlė kardinalių kintamųjų apdorojimo metodus:

- semantinis grupavimas – didelio kardinalumo kintamojo reikšmės yra sugrupuojamos pagal panašumą ir transformuojamos į nedidelio kardinalumo kintamąjį. Toks grupavimas reikalauja žinių apie reikšmių panašumą. Atliekant grupavimą yra prarandama dalis informacijos;
- transformacija į tolydų kintamąjį. Autoriai pateikė tris metrikas: WOE (*Weight of evidence*), *Supervised Ratio* (SR) ir *Perlich Ratio* (PR). Šios metrikos apibūdina santykį tarp lojalių ir nelojalių klientų kiekvienai didelio kardinalumo kintamojo reikšmei.

Eksperimentai parodė, kad didelio kardinalumo kintamųjų įtraukimas į modelį gerokai pagerina prognozės tikslumą. Tiksliausi prognozės rezultatai gauti naudojant WOE transformaciją.

Rečiau literatūroje sutinkami, tačiau reikšmingi, lojalumo prognozavimo nepriklausomi kintamieji [3]:

- prekės ženklas (angl. *brand*) – išleista pinigų suma pagal prekės ženklą (vietinis prekės ženklas, prekybos centro prekės ženklas, mažesnio biudžeto prekės ženklas);
- laikas (angl. *timing*) – vidutinis apsipirkimo momento laikas, vidutinio apsipirkimo momento standartinis nuokrypis, paskutiniojo apsilankymo laiko momentas;
- apmokėjimo tipas (angl. *mode of payment*) – procentine išraiška išreikšta pinigų suma, išleista pagal apmokėjimo tipą (grynieji pinigai, čekis, kuponas, debetinė mokėjimo kortelė, kreditinė mokėjimo kortelė), pinigų kiekis išleistas pagal rūšį (grynieji, čekis, mokėjimo kortelė), gražintos į parduotuvę užstatinės taros kiekis;
- skatinimas (angl. *promotions*) – apsilankymų, kurių metu buvo panaudotas kuponas, skaičius, kiek kartų klientas apsilankė parduotuvėje po paskutiniojo apsilankymo, kurio metu buvo panaudotas kuponas, vidutinė panaudotų kuponų vertė (per vieną apsipirkimą), uždirbtų lojalumo taškų skaičius.

1.3. Klientų lojalumo prognozavimo modeliai ir jų taikymas

Vartotojų lojalumo prognozavimas šiais laikais yra laikomas binarine klasifikavimo problema. Šiai problemai spręsti yra suformuota daug įvairių algoritmų, pradedant paprasčiausiu RFM modeliu ir baigiant sudėtingais mašininio mokymo algoritmų ansambliais. Šiame skyriuje apžvelgsime, kokie metodai yra plačiausiai paplitę mokslinėje literatūroje, susijusioje su klientų lojalumo prognozavimu.

Metodo pasirinkimas priklauso nuo daug faktorių. Skirtingiems duomenų rinkiniams gali tikti visiškai skirtingi klasifikavimo metodai. Pasak [25] straipsnio autoriaus, 14-oje iš 18-os apžvelgtų literatūros šaltinių, kuriuose nagrinėjama klientų lojalumo prognozavimo tema,

autoriai kaip klasifikavimo etaloną laiko logistinės regresijos modelį. Autoriai šį modelį naudoja rezultatų palyginimui su jų pasiūlytais klientų lojalumo prognozavimo sprendimais.

Logistinė regresija yra vienas iš gerai žinomų klasifikavimo metodų prognozuoti binarinį priklausomą kintamąjį. Šis metodas yra itin populiarius dėl kelių priežasčių [14]:

- a) metodas yra lengvai suprantamas ir interpretuojamas;
- b) klasifikavimo rezultatas yra tikimybinis skirstinys;
- c) reikalingi sąlyginai nedideli skaičiavimo resursai.

Tačiau, norint pasiekti aukštus rezultatus, prieš modeliuojant klientų lojalumą logistine regresija, duomenis reikia apdoroti, t.y. atlikti duomenų transformacijas, grupavimą pagal kategorijas, pašalinti išskirtis, suvienodinti balansą tarp prognozuojamo kintamojo klasių. Gordinis ir Veglio [25] sudarė 9 (sprendimų medžiai, neuroniniai tinklai, naivusis Bajeso klasifikatorius, Bajeso neuroniniai tinklai, atsitiktiniai miškai, stochastinis gradientinis boosting, radialinio branduolio atraminių vektorių mašinos, CART sprendimų medžiai) klientų lojalumo modelius su neapdorotais duomenimis. Taip pat autorius pasiūlė duomenų apdorojimo būdą, skirtą pagerinti prognozės tikslumą, ir jį išbandė sudarydamas logistinės regresijos modelį. Atlikus palyginamąją prognozės rezultatų analizę, nustatyta, kad logistinė regresija su specialiai apdorotu duomenų rinkiniu gali pasiekti tokį patį ar net geresnį tikslumą kaip ir sudėtingesni mašininio mokymo algoritmai. Pažymėtina, kad Gordinio ir Veglio [25] atliktame tyrime tiksliausiai prognozę su neapdorotu duomenų rinkiniu atliko atsitiktinių miškų algoritmas. Migelis, Poelas, Camanho ir Cumha [8] atlikta analizė parodė, kad logistinė regresija tiksliau prognozuoja klientų lojalumą už RFM modelį.

Dažniausiai literatūroje sutinkami klientų lojalumo prognozavimo metodai: logistinė regresija, dirbtiniai neuroniniai tinklai, sprendimų medžiai, atraminių vektorių mašinos (SVM), atsitiktiniai miškai. Daugelyje atliktų mokslinių studijų šie modeliai lyginami tarpusavyje ir išrenkamas tiksliausiai prognozę atlikęs klasifikatorius. Straipsniuose [6] ir [20] tiksliausiu klasifikatoriumi išrinktas sprendimų medžių modelis. Straipsniuose [10], [14], [24], [26] – SVM modelis. Autorių paminėti SVM modelio trūkumai – itin didelių skaičiavimo resursų (kompleksiškumas – $O(n^3)$) poreikis, daug laiko reikalaujanti optimaliausių įvesties parametrų paieška bei branduolio funkcijos parinkimas. Neuroniniai tinklai tiksliausiai atliko prognozę Rungės, Gao, Garcinio ir Faltingso [21] aprašytame tyrime. Bukinksas ir Poelis [3] sudarė logistinės regresijos, neuroninių tinklų ir atsitiktinių medžių prognozavimo modelius. Tiksliausi rezultatai gauti atsitiktinių miškų modeliu. Pasak autoriaus, atsitiktiniai miškai patrauklūs dėl keleto savybių:

- a) aukštas prognozės tikslumas;
- b) stabilūs ir vientisi prognozės rezultatai, nėra poreikio atlikti kryžminį validavimą;

- c) nereikalingas įvesties parametrų parinkimas;
- d) lengva išmatuoti nepriklausomų kintamųjų svarbą;
- e) palyginti trumpas skaičiavimo laikas (300 kartų ilgesnis nei logistinė regresija, tačiau 300 greitesnis už neuroninius tinklus).

Apžvelgtoje literatūroje galima rasti patobulintų ankščiau aprašytų prognozavimo metodų variantų. Gordinis ir Veglio [14] sudarė modelius mažmeninės prekybos tinklo klientų lojalumui prognozuoti. Autoriai pasiūlė kintamųjų atrinkimo metodiką SVM modeliui atlikti remiantis AUC (angl. *area under curve* – plotas esantis po ROC (angl. *receiver operator characteristic*) kreive) kriterijumi. Ištirti pasiūlytos metodikos pranašumui buvo sudaryti neuroninių tinklų ir logistinės regresijos prognozavimo modeliai. Rezultatų analizė parodė, kad nauja metodologija pasiteisino, buvo pasiektas 3,54 % didesnis tikslumas už SVM su paprastu kintamųjų atrinkimu, 4,57% didesnis tikslumas už neuroninių tinklų modelį ir net 5,87% didesnis tikslumas už logistinės regresijos modelį. Chenas ir Fanas [23] klientų lojalumo prognozavimui naudojo SVM, keleto branduolių SVM ir patobulintą, keleto bendradarbiaujančių (angl. *collaborative*) SVM modelį. Geriausi prognozės rezultatai buvo gauti patobulintu C-MK-SVM modeliu. Ju, Guo ir Hujangas [19] sudarė elektroninės prekybos įmonės klientų lojalumo prognozavimo modelius. Autoriai suformavo patobulintą SVM modelį. Patobulintu modeliu atlikta prognozė palyginta su kitų mašininio mokymo algoritmų rezultatais (neuroniniai tinklai, sprendimų medžiai, SVM su tiesiniu, spinduliniu ir polinominiu branduoliais). Patobulintu modeliu gauta tikslesnė už kitų mašininio mokymo algoritmų prognozė.

Chenas [22] sudarė klientų lojalumo prognozavimo modelius naudodamas dinامينius kintamuosius, t.y. kiekvienas kintamasis yra išskaidytas į laiko eilutę. Autoriaus tikslas buvo sukurti metodiką ir apjungti dinامينius kintamuosius su statiniais. Tam jis išplėtojo SVM metodą ir pasiūlė naują – hierarchinį keleto branduolių atraminių vektorių mašinų modelį. Pasiūlytas modelis yra sudarytas iš dviejų klasifikatorių: vienas klasifikuoja klientus į lojalus ir nelojalus naudodamas demografinius duomenis, kitas – dinامينius pirkimo transakcijų duomenis. Modelio mokymo metu yra atliekamas nepriklausomų kintamųjų atrinkimas. Dinaminiam duomenims taip pat atliekamas mokymo laikotarpio parinkimas. Chenas [22] suformulavo tris modelio sudarymo struktūras:

- standartinė modelio struktūra – naudojami demografiniai duomenys bei į statiškus duomenis transformuoti transakciniai duomenys;
- modelio struktūra naudojant kintamųjų atrinkimą – kiekvienas transakcinių duomenų kintamasis yra išskaidomas į T kintamųjų (T – laiko periodų skaičius) ir naudojamas vienas bendras klasifikatorius demografiniams ir transakciniams duomenims;

- modelių ansamblio struktūra – sudaromi atskiri klasifikavimo modeliai demografiniams ir transakciniams duomenims. Šių modelių ansamblis naudojamas kaip galutinis klasifikatorius.

Chenas [22] atliko prognozės modelio sudarymo struktūros palyginamąją analizę naudojant tris duomenų rinkinius – mažmeninės prekybos įmonės duomenis, didmeninės prekybos įmonės duomenis ir telekomunikacijų bendrovės duomenis. Duomenų rinkiniai buvo išanalizuoti naudojant tris anksčiau aprašytus modelio sudarymo būdus bei įprastus klasifikatorius, tokius kaip atsitiktiniai miškai, logistinė regresija, sprendimų medžiai. Modelių prognozės rezultatai vertinami pagal jautrumo ir specifiškumo statistikas, AUC ir lift kriterijus. Lyginant tris SVM modelių sudarymo metodikas, tiksliausi prognozės rezultatai gauti naudojant hierarchinį keleto branduolių atraminių vektorių mašinų metodą. Lyginant H-MK-SVM modeliu gautus rezultatus su klasikiniiais klasifikatoriais nustatyta, kad H-MK-SVM tiksliau atlieka prognozę. Taip pat Chenas [22] atliko jautrumo poimčių sudarymui analizę. Gauti rezultatai parodė, kad pasiūlyti atraminių vektorių mašinų algoritmai yra mažiau jautrūs poimčių sudarymui negu klasikiniai algoritmai.

Mokslinėje literatūroje pastebėta tendencija klasifikatoriams pritaikyti *Boosting* algoritimą. *Boosting* – metodika, kai klasifikavimas atliekamas keletu iteracijų, kiekvienoje kitoj iteracijoje atsižvelgiant į prieš tai buvusioje padarytas klasifikavimo klaidas. Taip klasifikatorius stengiasi išvengti klaidų ir gaunami tikslesni prognozės rezultatai. Vafeiadis, Diamantaras, Sarigianidis ir Chatzisavas [11] atliko įvairių prognozavimo metodų palyginamąją analizę. Sudaryti modeliai buvo skirti telekomunikacijų bendrovės klientų lojalumui prognozuoti. Autoriai sudarė prognozavimo modelius naudodami dirbtinius neuroninius tinklus, sprendimų medžius, atraminių vektorių mašinas, naivųjį Bajeso klasifikatorių ir logistinę regresiją. Papildomai, kiekvienam iš klasifikatorių buvo pritaikyta *Boosting* metodologija ir patikrinta, ar tai prideda papildomo prognozavimo tikslumo. Autoriai nustatė, kad prognozę tiksliausiai atlieka SVM klasifikatorius su *Boosting* metodologija. Autoriai taip pat sudarė klientų lojalumo prognozavimo modelį duomenims, kurie pasižymi dideliu disbalansu tarp lojalių ir nelojalių klientų. Prognozavimo modelis buvo sudarytas naudojant autorių patobulintą *Real Adaboost boosting* metodą. Prognozės rezultatai buvo palyginti su sprendimų medžių modeliu, kuriam pritaikyta *bagging* metodika. Prognozės rezultatai parodė, kad tikslesnis modelis gautas naudojant patobulintą *Real Adaboost* modelį su *Boosting* metodika. Tamadonis, Stakovičius ir Evingas [4] sudarė prognozavimo modelius naudodami atraminių vektorių mašinas, *Boosting* klasifikatorių ansamblį, Pareto tikimybinį modelį, RFM, logistinės regresijos modelį. Geriausi rezultatai pasiekti naudojant *Boosting* metodologiją, prasčiausi – tikimybinį Pareto modelį. Nustatyta, kad Pareto modelis atlieka prognozę tiksliau nei mašininio mokymo algoritmai tik tais atvejais, kai

duomenų rinkinys yra labai mažas, o logistinė regresija – kai santykis tarp lojalių ir nelojalių klientų yra labai didelis. Gladis, Basensas ir Kruksas [17] sudarė 5 prognozavimo modelius: logistinės regresijos, neuroninių tinklų, dviejų tipų sprendimų medžių modelius ir *AdaCost boosting* prognozavimo modelį. Tiksliausia prognozė gauta *cost – sensitive* sprendimų medžių modeliu ir *AdaCost boosting* modeliu.

Didelio potencialo, tačiau ir didelių skaičiavimo resursų, reikalaujanti metodika – keleto klasifikatorių sujungimas į klasifikatorių ansamblį. Autorių Keramačio, Jafari-Marandžio, Alianejadžio, Mozafario ir Abasio [26] tikslas buvo išgauti maksimalų prognozavimo tikslumą naudojantis sprendimų medžiais, dirbtiniais neuroniniais tinklais, K artimiausių kaimynų ir atraminių vektorių mašinų metodais. Autoriai atliko kiekvieno metodo geriausių įvesties parametrų paiešką. Naudojant geriausių įvesties parametrų rinkinį autoriai sujungė 4 prognozavimo modelius į ansamblį ir pakartoto prognozę naudodami kombinuotą klasifikatorių. Kombinuotas klasifikatorius parodė geriausius rezultatus. Kusementas ir De Bokas [18] klientų lojalumo prognozei naudojo *CART* sprendimų medžius, apibendrintą adityvųjį modelį, atsitiktinių miškų modelį ir ansamblio klasifikatorių *GAMens*. Geriausi rezultatai gauti atsitiktinių miškų klasifikatoriumi.

1.4.Literatūros šaltinių analizės apibendrinimas

Šiame skyriuje buvo apžvelgti moksliniai tyrimai klientų lojalumo tematika. Nepaisant įvairių atliktų tyrimų įvairovės, neaptikta tyrimų, susijusių su vaistinės klientų lojalumo atveju.

Priklausomai nuo tiriamų duomenų konteksto, klientų lojalumas gali būti apibrėžiamas įvairiais būdais. Šiame darbe bus nagrinėjamas klientų lojalumo atvejis, kai klientus su įmone sieja sutartimi neapibrėžti santykiai. Remiantis moksliniais tyrimais, bus pasiūlyti nelojalių ir iš dalies nelojalių klientų apibrėžimai.

Klientų elgsenos lojalumui įvertinti gali būti naudojami įvairūs demografiniai rodikliai bei transakciniai duomenys. Šiame darbe suformuosime požymių, skirtų vaistinės klientų lojalumui įvertinti, rinkinį. Šį kintamųjų rinkinį panaudosime klientų lojalumo prognozavimo modeliams apmokyti.

Klientų lojalumo prognozavimas gali būti atliekamas naudojant rinkodaros metodus (pvz. RFM), tikimybinis modelius (pvz. Pareto) arba mašininio mokymo algoritmus (pvz. atsitiktiniai miškai, atraminių vektorių mašinos). Prognozavimo etalonu daugelyje tyrimų laikomas logistinės regresijos modelis. Jis tikslumu pranoksta ne tik RFM, tačiau ir tikimybinis prognozavimo modelius. Taip pat, daugelyje tyrimų buvo naudojami įvairūs mašininio mokymo algoritmai. Šiame darbe sudarysime logistinės regresijos, atsitiktinių miškų ir atraminių vektorių mašinų klientų lojalumo prognozavimo modelius ir juos palyginsime.

LITERATŪROS SĄRAŠAS

1. KUMAR, V., & WERNER, R. (2016 m. Lapkritis). Creating Enduring Customer Value. *Journal of Marketing*, 80(6), 36-68. doi:<http://dx.doi.org/10.1509/jm.15.0414>
2. BANDYOPADHYAY, S., & MARTELL, M. (2007). Does attitudinal loyalty influence behavioral loyalty? A theoretical and empirical study. *Journal of Retailing and Consumer Services*, 14(1), 35-44. doi:<https://doi.org/10.1016/j.jretconser.2006.03.002>
3. BUCKINX, W., & VAN DEN POEL, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), 252–268. doi:<https://doi.org/10.1016/j.ejor.2003.12.010>
4. TAMADDONI, A., Stakhovych, S., & EWING, M. (2016). Comparing Churn Prediction Techniques and Assessing Their Performance: A Contingent Perspective. *Journal of Service Research*, 19(2), 123-141. doi:<https://doi.org/10.1177/1094670515616376>
5. LIU, M., QIAO, X., & XU, W. (2011). Three Categories Customer Churn Prediction Based on the Adjusted Real Adaboost. *Communications in Statistics - Simulation and Computation*, 40(10), 1548-1562. doi:<http://dx.doi.org/10.1080/03610918.2011.589732>
6. KRALJEVIC, G., & GOTOVAC, S. (2010). Modeling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services. *Automatika*, 51(3), 275-283. doi:<http://dx.doi.org/10.1080/00051144.2010.11828381>
7. CLEMENTE-CISCAR, M., SAN MATIAS, S., & GINER-BOSCH, V. (2014). A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings. *European Journal of Operational Research*, 239(1), 276-285. doi:<https://doi.org/10.1016/j.ejor.2014.04.029>
8. MIGUEIS, V., VAN DEN POEL, D., CAMANHO, A., & CUNHA, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, 39(12), 11250–11256. doi:<https://doi.org/10.1016/j.eswa.2012.03.073>
9. BOSE, I., & Chen, X. (2009). Quantitative models for direct marketing: A review from systems perspective. 195(1), 1-16. doi:<https://doi.org/10.1016/j.ejor.2008.04.006>
10. MIGUEIS, V., CAMANHO, A., & CUNHA, J. (2013). Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines. *Expert Systems with Applications*, 40(1), 6225–6232. doi:<http://dx.doi.org/10.1016/j.eswa.2013.05.069>
11. VAFEIADIS, T., DIAMANTARAS, K., SARIGIANNIDIS, G., & CHATZISAVVAS, K. (2015). A comparison of machine learning techniques for customer churn prediction. 55(1), 1-9. doi:<http://dx.doi.org/10.1016/j.simpat.2015.03.003>
12. ABBASIMEHR, H., SETAK, M., & SOROOR, J. (2013). A framework for identification of high-value customers by including social network based variables for churn prediction using neuro-fuzzy techniques. *International Journal of Production Research*, 54(4), 1279-1294. doi:<http://doi.org/10.1080/00207543.2012.707342>
13. KHOBZI, H., AKHONDZADEH-NOUGHABI, E., & MINAEI-BIDGOLI, B. (2014). A New Application of RFM Clustering for Guild Segmentation to Mine the Pattern of Using Banks' e-Payment Services. *Journal of Global Marketing*, 27(3), 178-190. doi:<http://dx.doi.org/10.1080/08911762.2013.878428>

14. GORDINI, N., & VEGLIO, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62(1), 100-107. doi:<http://dx.doi.org/10.1016/j.indmarman.2016.08.003>
15. BALLINGS, M., & VAN DEN POEL, D. (2012). Customer event history for churn prediction: How long is long enough? *Expert Systems with Applications*, 39(1), 13517–13522. doi:<http://dx.doi.org/10.1016/j.eswa.2012.07.006>
16. EKINCI, Y., ULENGIN, F., URAY, N., & ULENGIN, B. (2014). Analysis of customer lifetime value and marketing expenditure decisions through a Markovian-based model. *European Journal of Operational Research*, 237(1), 278-288. doi:<http://dx.doi.org/10.1016/j.ejor.2014.01.014>
17. GLADY, N., BAESENS, B., & CROUX, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402-411. doi:<http://doi.org/10.1016/j.ejor.2008.06.027>
18. COUSSEMENT, K., & DE BOCK, K. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, 66(1), 1629-1636. doi:<http://dx.doi.org/10.1016/j.jbusres.2012.12.008>
19. YU, X., GUO, S., GUO, J., & HUANG, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 38(1), 1425-1430. doi:<http://dx.doi.org/10.1016/j.eswa.2010.07.049>
20. HADIJI, F., SIFA, R., DRACHEN, A., THURAU, C., KERSTING, K., & BAUCKHAGE, C. (2014). Predicting Player Churn In the Wild. *Computational Intelligence and Games*. Dortmund, Germany: IEEE. doi:<http://dx.doi.org/10.1109/CIG.2014.6932876>
21. RUNGE, J., GAO, P., GARCIN, F., & FALTINGS, B. (2014). Churn prediction for high-value players in casual social games. *Computational Intelligence and Games (CIG)*. Dortmund: IEEE. doi:dx.doi.org/10.1109/CIG.2014.6932875
22. CHEN, Z., FAN, Z., & SUN, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 223(2), 461-472. doi:<https://doi.org/10.1016/j.ejor.2012.06.040>
23. CHEN, Z., & FAN, Z. (2012). Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM approach. *Knowledge-Based Systems*, 35, 111-119. doi:<https://doi.org/10.1016/j.knosys.2012.04.023>
24. MOEYERSOMS, J., & MARTENS, D. (2015). Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. *Decision Support Systems*, 72, 72-81. doi:<https://doi.org/10.1016/j.dss.2015.02.007>
25. COUSSEMENT, K., LESSMANN, S., & VERSTRAETEN, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, 95, 27-36. doi:<https://doi.org/10.1016/j.dss.2016.11.007>
26. KERAMATI, A., JAFARI-MARANDI, R., ALIANNEJADI, M., AHMADIAN, I., MOZAFFARI, M., & ABBASI, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24, 994-1012. doi:<https://doi.org/10.1016/j.asoc.2014.08.041>

27. MACQUEEN, J. (1967). Some methods for classification and analysis of multivariate observations. Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, p. 281-297. California. Nuskaityta iš <http://projecteuclid.org/euclid.bsm/1200512992>
28. CHAOVALIT, P., GANGOPADHYAY, A., KARABATIS, G., & CHEN, Z. (2011). Discrete wavelet transform-based time series analysis and mining. ACM Computing Surveys, 43(2). doi:<http://dx.doi.org/10.1145/1883612.1883613>
29. WITTEN, I., FRANK, E., & HALL, M. (2011). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
30. HSU, C., CHANG, C., & LIN, C. (2003). A Practical Guide to Support Vector Classi. Taipei, Taiwan.
31. WALLACH, D., MAKOWSKI, D., JONES, J., & BRUN, F. (2014). Chapter 3 – The R Programming Language and Software. Esantis Working with Dynamic Crop Models (Second Edition) (p. 71-117). Academic Press. doi:<https://doi.org/10.1016/B978-0-12-397008-4.00003-4>
32. RAHM, E., & DO, H. (2000). Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bulletin, 23.
33. MULLER, H., & FREYTAG, J. (2003). Problems, Methods, and Challenges in Comprehensive Data Cleansing. Nuskaityta iš http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/techreports/2003-hub_ib_164-mueller.pdf