



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

# **Sukčiavimo rizikos aptikimas tarpusavio skolinimosi platformose**

Baigiamasis magistro studijų projektas

---

**Samanta Graužinienė**

Projekto autorė

Doc. dr. Tomas Ruzgas

Vadovas

Doc. dr. Aušrinė Lakštutienė

Vadovė

**Kaunas, 2023**

Graužinienė Samanta. Sukčiavimo rizikos aptikimas tarpusavio skolinimosi platformose. Magistro studijų baigiamasis projektas / vadovas doc. dr. Tomas Ruzgas, vadovė doc. dr. Aušrinė Lakštutienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Taikomoji matematika.

Reikšminiai žodžiai: sukčiavimo rizika, pagrindinių komponentų analizė, atsitiktiniai miškai, neuroniniai tinklai, atraminių vektorių mašinos.

Kaunas, 2023. 59 p.

## **Santrauka**

Ilgą laiką investavimas į žmonių paskolas buvo prieinamas tik didiesiems bankams. Pradėjus steigti ir sėkmingai plėtojantis tarpusavio skolinimosi platformoms ši galimybė tapo prieinama ir fiziniams asmenims. Plečiantis tarpusavio skolinimosi platformų rinkai, investuotojai susiduria su vis daugiau sukčiavimo atvejų. Šio tyrimo tikslas yra sudaryti sukčiavimo rizikos vertinimo modelį, pagrįstą tarpusavio skolinimosi platformos UAB „Bendras finansavimas“ duomenimis, siekiant iš anksto identifikuoti sukčiavimo atvejus ir užkirsti kelią jų įgyvendinimui.

Pirmoje darbo dalyje išanalizuoti mokslinėje literatūroje taikyti metodai sukčiavimo rizikai vertinti, išskirti reikšmingiausi nustatyti sukčiavimo riziką lemiantys veiksniai.

Antroje darbo dalyje pristatomi surinkti viešai prieinami duomenys apie UAB „Bendras finansavimas“ antrinėje rinkoje parduodamas paskolas. Apibrėžiami tyrime naudoti duomenų imties tvarkymo bei sukčiavimo rizikos vertinimo metodai: atsitiktiniai miškai, atraminių vektorių mašinos, dirbtiniai neuroniniai tinklai.

Trečioje darbo dalyje duomenys apdoroti pagrindinių komponentų analizės būdu. Sudaryti atsitiktinių miškų, atraminių vektorių mašinų bei dirbtinių neuroninių tinklų modeliai naudojant tris skirtingas duomenų imtis: pradinį duomenų rinkinį (neapdorotą pagrindinių komponentų analizės būdu); imtį su išskirtomis aštuoniomis pagrindinėmis komponentėmis; imtį su išskirtomis aštuoniomis pagrindinėmis komponentėmis bei kokybiniais tyrimo kintamaisiais.

Atlikus tyrimą nustatyta, kad atsitiktinių miškų modelis su aštuonių pagrindinių komponentų bei kokybinių kintamųjų duomenų failu tiksliausiai suklasifikuoja tyrimo objektą – sukčiavimo faktą. Reikšmingiausiais sukčiavimą lemiančiais veiksniais išskirti: bloga kredito istorija, įsipareigojimų ir pajamų santykis, paskolos suma, kliento nemokumo rizikos įvertis bei amžius.

Graužinienė Samanta. Fraud risk detection on peer-to-peer lending platforms. Master's Final Degree Project / Supervisor doc. dr. Tomas Ruzgas; doc. dr. Aušrinė Lakštutienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics.

Keywords: fraud risk, principal component analysis, random forests, artificial neural networks, support vector machines.

Kaunas, 2023. 59 p.

### **Summary**

For a long time, investing in people's credit was only possible for the big banks. With the establishment and successful development of peer-to-peer lending platforms, this opportunity became available to individuals. With the expansion of the peer-to-peer lending platform market, investors are facing more and more scams. The aim of this study is to build a fraud risk assessment model based on the data of the peer-to-peer lending platform UAB "Bendras finansavimas" in order to detect frauds in advance and prevent their implementation.

In the first part of the paper, the methods used in the scientific literature to assess the risk of fraud are analyzed and the main factors determining the risk of fraud are elaborated.

The second part of the paper presents the collected publicly available data on the loans sold by UAB "Bendras finansavimas" on the secondary market. The data sample management and fraud risk assessment methods used in the study are defined: Random Forests, Support Vector Machines, Artificial Neural Networks.

In the third part of the paper, the data are processed using principal component analysis. Models of Random Forests, Support Vector Machines and Artificial Neural Networks were created with three different data samples: Primary dataset (without implemented principal component analysis); a sample with eight extracted principal components; sample with eight principal components and qualitative research variables.

After conducting the research, it was found that the Random Forest model with the dataset of eight principal components and qualitative variables most accurately classified the object of study - the fact of fraud. The main factors determining fraud are: bad credit history, debt service-to-income ratio, the amount of the loan, the client's assessment of the risk of fraud and age.

## Turinys

<b>Lentelių sąrašas</b> .....	5
<b>Paveikslų sąrašas</b> .....	6
<b>Įvadas</b> .....	7
<b>1. Privačių asmenų tarpusavio skolinimosi sukčiavimo atvejų kredito rizikos vertinimo teoriniai aspektai</b> .....	9
1.1. Privačių asmenų tarpusavio skolinimosi platformos bei jų reguliavimas.....	9
1.2. Sukčiavimo riziką didinančių veiksnių analizė.....	14
1.3. Sukčiavimo rizikos vertinimo modelių analizė.....	18
<b>2. Sukčiavimo atvejų nustatymo tyrimo metodologija</b> .....	24
2.1. Tarpusavio skolinimosi platformos „SAVY“ analizė.....	24
2.2. Tyrimo metodologija .....	31
<b>3. Tyrimų rezultatai ir jų aptarimas</b> .....	37
3.1. Pagrindinių komponentų analizė.....	37
3.2. Mašininio mokymosi metodų taikymas .....	40
3.2.1. Duomenų imties paruošimas .....	40
3.2.2. Atsitiktinių miškų metodas.....	41
3.2.3. Atraminių vektorių mašinų metodas.....	42
3.2.4. Neuroniniai tinklai.....	43
3.2.5. Atsitiktiniai miškai. Priklausomojo kintamojo disbalanso šalinimas.....	47
3.3. Sukčiavimą lemiančių veiksnių svarba.....	49
3.4. Rezultatų palyginimas.....	50
<b>Išvados</b> .....	52
<b>Literatūros sąrašas</b> .....	53
<b>Informacijos šaltinių sąrašas</b> .....	57
<b>Priedai</b> .....	58

## Lentelių sąrašas

<b>1 lentelė.</b> TSPO reguliavimo lyderiaujančiose valstybėse apžvalga. Sudaryta remiantis: Huang ir Wang, (2021); Nemoto, Huang ir Storey (2019); Chiu (2016).....	11
<b>2 lentelė.</b> Sukčiavimo rizikos veiksniai.....	16
<b>3 lentelė.</b> Rizikos vertinimo modeliai .....	18
<b>4 lentelė.</b> Atsitiktinių miškų, neuroninių tinklų ir atraminių vektorių metodų palyginimas. Sudaryta remiantis: Wainberg, Alipanahi ir Frey (2016); Płoński (2019). .....	21
<b>5 lentelė.</b> Turimas UAB „SAVY“ duomenų rinkinys.....	24
<b>6 lentelė.</b> Pagrindinių komponentų analizės vertinimo rodikliai (sudaryta remiantis: Pūkėnas, 2009; Stankūnaitė, 2009; Shrestha, 2021). .....	32
<b>7 lentelė.</b> Modelių vertinimo kriterijai. Sudaryta remiantis: Pukėnas (2009) ir Varghese (2018).....	35
<b>8 lentelė.</b> Duomenų rinkinio tikimo pagrindinių komponentų analizei rodikliai. ....	37
<b>9 lentelė.</b> Atsitiktinių miškų modelių vertinimo kriterijų palyginimas.....	41
<b>10 lentelė.</b> Atraminių vektorių mašinų modelių vertinimo kriterijų palyginimas.....	43
<b>11 lentelė.</b> Neuroninių tinklų modelių vertinimo kriterijų palyginimas. ....	43
<b>12 lentelė.</b> Sudarytų modelių vertinimo kriterijų palyginimas. ....	46
<b>13 lentelė.</b> Priklausomojo kintamojo pasiskirstymas mokymo bei testavimo duomenų imtyse.....	47
<b>14 lentelė.</b> Pašalinto disbalanso mokymo imties priklausomo kintamojo pasiskirstymas.....	47
<b>15 lentelė.</b> Atsitiktinių miškų ir neuroninių tinklų modelių palyginimas. ....	48
<b>16 lentelė.</b> Sukčiavimo rizikos vertinimo modelių rezultatų palyginimas.....	50

## Paveikslų sąrašas

<b>1 pav.</b> Kinijos tarpusavio skolinimosi platformų skaičius vnt. (2007–2020). Sudaryta Rodenbaugh (2020) .....	11
<b>2 pav.</b> TSPO skaičius ir per metus išmokėtų vartojimo kreditų suma Lietuvoje (mln. Eur.). Sudaryta remiantis Lietuvos banko duomenimis 2022.....	13
<b>3 pav.</b> Apgaulės trikampis. Sudaryta remiantis D. R. Cressey’io teorija (1970) .....	15
<b>4 pav.</b> Kokybinių kintamųjų pasiskirstymas duomenų rinkinyje.....	26
<b>5 pav.</b> Kintamųjų pasiskirstymas duomenų rinkinyje pagal lytį.....	27
<b>6 pav.</b> „SAVY” sukčiaujančių klientų pasiskirstymas pagal skolintų pinigų panaudojimo tikslą .....	28
<b>7 pav.</b> „SAVY” klientų skolintų pasiskirstymas pagal šeimyninę padėtį (kairėje) ir finansiškai išlaikomų asmenų skaičių (dešinėje) .....	28
<b>8 pav.</b> „SAVY” sukčiaujančių klientų pasiskirstymas pagal viešai deklaruotą turtą.....	29
<b>9 pav.</b> „SAVY” sukčiaujančių klientų pasiskirstymas pagal darbinę sritį.....	29
<b>10 pav.</b> „SAVY” sukčiaujančių klientų pasiskirstymas pagal įgytą išsilavinimą.....	30
<b>11 pav.</b> „SAVY” sukčiaujančių klientų pasiskirstymas pagal lytį.....	30
<b>12 pav.</b> Geografinis „SAVY” sukčiaujančių klientų pasiskirstymas.....	31
<b>13 pav.</b> Atsitiktinių miškų algoritmo schema. Sudaryta remiantis (Dmitrievsky, 2018).....	33
<b>14 pav.</b> Atraminių vektorių mašinų veikimo vizualizavimas (Schultebrucks, 2017). .....	34
<b>15 pav.</b> Dirbtinio neuroninio tinklo schema (Verikas ir Gelžinis, 2008).....	34
<b>16 pav.</b> Pirmųjų 10 pagrindinių komponentų paaiškinamos dispersijos dalis .....	39
<b>17 pav.</b> Suminė komponentų paaiškinama priklausomojo kintamojo dispersijos dalis.....	39
<b>18 pav.</b> Kiekybinių tyrimo duomenų koreliacijų matrica. ....	40
<b>19 pav.</b> Modelių sudarytų su pirmuoju duomenų rinkiniu ROC ir EER kreivių grafikai.....	45
<b>20 pav.</b> Modelių sudarytų su antruoju duomenų rinkiniu ROC ir EER kreivių grafikai.....	45
<b>21 pav.</b> Modelių sudarytų su trečiuoju duomenų rinkiniu ROC ir EER kreivių grafikai.....	46
<b>22 pav.</b> Atsitiktinių miškų modelio su subalansuota imtimi ROC kreivė.....	48
<b>23 pav.</b> Kintamųjų svarba.....	50

## Įvadas

**Tyrimo aktualumas.** Finansinių technologijų (FinTech) sektorius itin sparčiai auga ir vis daugiau reikšmės įgyja visoje finansų rinkoje (Bytautas ir Morkūnas, 2023). FinTech apibrėžiamas kaip technologinės naujovės, taikomos tradicinių finansinių paslaugų kontekste, kurios lemia esamą finansinių paslaugų rinkos evoliuciją (Dhar ir Stein, 2017). Naujų technologijų pritaikymas teikiant finansines paslaugas leidžia rasti naujiems verslo modeliams, procesams bei produktams. Vienas iš inovatyvių FinTech verslo modelių – tarpusavio skolinimosi platformos (angl. *peer-to-peer* – P2P). Šios platformos suveda lėšų stokojantį asmenį su asmeniu, turinčiu laisvų neįdarbintų pinigų, be finansų rinkos tarpininko. Tarpusavio skolinimosi platformos yra patogi tradicinių bankų alternatyva, dažnu atveju leidžianti skolintis pinigus pigiau bei greičiau, taip pat suteikianti galimybę apsaugoti sukauptas lėšas nuo nuvertėjimo.

Tarpusavio skolinimosi platformų skaičius pasaulyje sparčiai auga ir iki 2030 metų prognozuojama, kad P2P išaugs 4,3 karto lyginant su 2022 metais (Precedence Research, 2022). Didėjant šiai rinkai, neišvengiamai didėja ir iššūkių, susijusių su investuotų, sutelktų pinigų išsaugojimu. Apžvelgiant bendrą finansinių technologijų rinkos padėtį, „sukčiavimų atliekant mokėjimo operacijas skaičius 2021 metais buvo beveik 7 kartus, o vertė – daugiau nei 8 kartus didesni nei 2019 m.“ (Bytautas ir Morkūnas, 2023). Taigi auganti tarpusavio skolinimosi platformų rinka neišvengiamai susiduria ir su didesniu bandymų pasisavinti skolintas lėšas skaičiumi, kaip ir visas FinTech sektorius.

*World Bank Group* 2021 metais atlikto vartotojų rizikos tyrimo finansinių technologijų sektoriuje duomenimis buvo nustatyta, kad tarpusavio skolinimosi platformos dalyviui praradus investuotas lėšas dėl sukčiavimo atvejų gali padidėti jo pačio kredito rizika, nesugebant laiku atgauti investuotų lėšų ir nebegalint laiku vykdyti turimų asmeninių finansinių įsipareigojimų. Taigi įvairūs sukčiavimo atvejai daro tiek tiesioginę, tiek netiesioginę įtaką visai finansų rinkai. Dėl to yra svarbi tinkama reguliatoriaus priežiūra bei gebėjimas iš anksto nustatyti galimą sukčiavimo riziką. Atsižvelgiant į tai, šiame darbe keliamas **problema** – kaip įvertinti sukčiavimo riziką tarpusavio skolinimosi platformose?

**Tyrimo tikslas** – sudaryti sukčiavimo rizikos vertinimo modelį remiantis tarpusavio skolinimosi platformų duomenimis.

**Tyrimo objektas** – sukčiavimo rizikos vertinimo modelis.

### Darbo uždaviniai:

1. išanalizuoti tarpusavio skolinimosi platformas bei jų reguliavimą;
2. išnagrinėti literatūros šaltiniuose išskiriamus sukčiavimo rizikos veiksnius bei naudotus sukčiavimo rizikos vertinimo modelius;
3. sudaryti sukčiavimo rizikos vertinimo modelį, remiantis Lietuvos tarpusavio skolinimosi platformos UAB „Bendras finansavimas“ duomenimis;

4. atlikti gautų rezultatų palyginimą su mokslinėje literatūroje pateikiamais sukčiavimo rizikos modelių įverčiais.



## **1. Privačių asmenų tarpusavio skolinimosi sukčiavimo atvejų kredito rizikos vertinimo teoriniai aspektai**

Šioje darbo dalyje apibrėžiama privačių asmenų tarpusavio skolinimosi platformos samprata bei reguliavimas skirtingose užsienio šalyse. Išanalizuoti skirtingi užsienių bei lietuvių autorių siūlomi kredito rizikos bei sukčiavimo apibrėžimai. Išskirti sukčiavimo riziką didinantys veiksniai. Siekiant darbo tikslo, palyginti užsienio literatūros autorių jau naudoti rizikos vertinimo modeliai.

### **1.1. Privačių asmenų tarpusavio skolinimosi platformos bei jų reguliavimas**

Finansų rinka, tai ekonominė erdvė, sudaranti galimybę susitikti finansinių lėšų stokojantiems asmenims ir asmenims, turintiems laisvų, neįdarbintų pinigų. Skaitmenizuojantis pasauliui ir finansinėms technologijoms tampant vis populiareesnėmis, dalis finansų rinkos taip pat turėjo galimybę pereiti į skaitmeninį pasaulį. Internetinis tarpusavio skolinimasis, tai būdas individualiems investuotojams teikti smulkias paskolas individualiems skolininkams be finansų institucijos tarpininkavimo (Emekter, R., Tu, Y., Jirasakuldech, B., et al., 2015). 2005 m. Jungtinėje Karalystėje pradėjo veikti pirmoji internetinė tarpusavio skolinimosi platforma fiziniams asmenims – „Zopa“. Šiai įmonei įsikūrus ir pradėjus savo verslą, siūlant klientams galimybę užsidirbti iš įdarbintų laisvųjų lėšų, taip pat klientams, stokojantiems apyvartinių lėšų, siūlant galimybę pasiskolinti palankesnėmis sąlygomis, įmonės veikla buvo įvardijama kaip „nišinė“, „trumpalaikė“ ar net „rizikinga“ (Bednorz, 2020).

Šiandien privačių asmenų tarpusavio skolinimasis įvardijamas kaip inovatyvus bei nusistovėjusią skolinimosi tvarką keičiantis būdas. Šis naujas verslo modelis laikomas banko paskolos alternatyva, sudaranti galimybę finansavimą gauti iš kelių ar net keliolikos skirtingų investuotojų. Nauda iš daugelio investuotojų yra ta, kad „dėl konkurencijos tarp potencialių rėmėjų pastebimai mažinamos palūkanų normos, tuo tarpu bankai dėl Centrinio banko nustatyto limito tokios mažos palūkanų normos pasiūlyti negali“ (Alešiūnaitė ir Malakaitė, 2019). Naujasis finansavimo šaltinis, P2P skolinimasis, kredito gavėjams patrauklus savo galimybe pasiūlyti mažesnes palūkanų normas nei bankai, o lėšų nestokojantiems asmenims – galimybe apsaugoti savo sukauptą kapitalą nuo nuvertėjimo, jį skolinant kreditų gavėjams per tarpusavio skolinimosi platformas, siekiant gauti gražos bei užsitikrinti pasyvias pajamas ateityje.

Gaunama graža už skolintas lėšas gali svyruoti nuo kelių procentų iki keliolikos ar net keliasdešimt, atsižvelgiant į pasirinktos investicijos rizikingumą. Žemo rizikos lygio paskolų gražos norma bus taip pat žema, esant aukštesnei rizikai, galima tikėtis didesnės palūkanų normos (gražos), mokamos investuotojui (Suryanarayana, 2020). Renkantis investiciją P2P platformoje, svarbu suprasti, kas yra rizika, kokio tipo ji gali būti ir kaip gali paveikti galutinį turimo finansinio portfelio dydį.

Pagrindinė paskolos finansuotojui tenkanti rizika – skolininko finansinių įsipareigojimų nevykdymas laiku arba visiškai nevykdymas, tapus nemokiu dėl pasikeitusios finansinės būklės. Ši rizika dažnai analizuojama moksliniuose šaltiniuose ir įvardijama, kaip kredito rizika. Kredito rizika – tikimybė, kad teisiškai vykdytina sutartis gali tapti bevertė, nes kita sandorio šalis nevykdys įsipareigojimų ar nutrauks savo veiklą (Anderson ir Edvard, 2013). Kitaip tariant, tai galima finansinė rizika, tenkanti asmeniui,

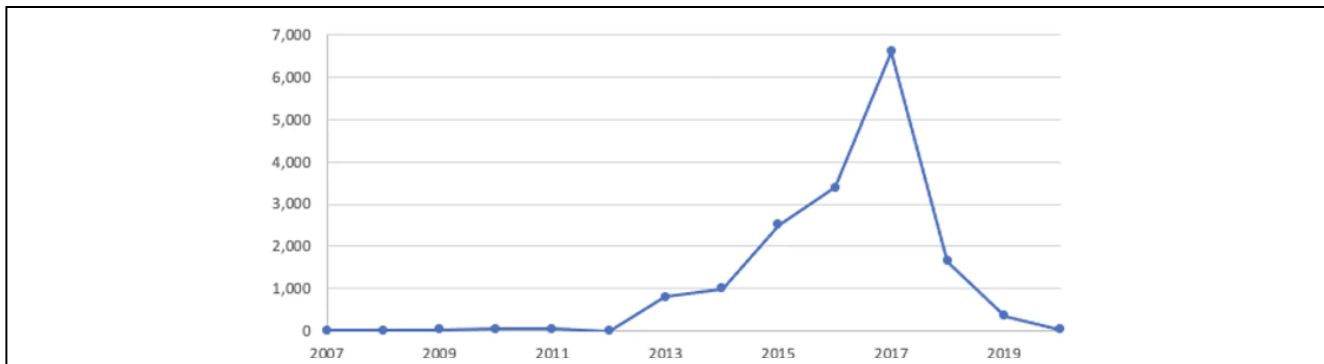
organizacijai, valstybei ar kitam subjektui, susijusi su finansinio sandorio daliniu ar visišku įsipareigojimų nevykdymu.

Kita pagrindinė tarpusavio skolinimo platformų investuotojams tenkanti rizika – sukčiavimo rizika (Wang et al. 2018). Šiandieniam pasaulyje neretai pasitaiko sukčiavimo (apgaulės) atvejų, kai siekiama gauti kreditą, bet intencijos jį mokėti ateityje nėra. Mokslinėje literatūroje sukčiavimas apibrėžiamas kaip „neteisingas ar nusikalstamai apgaulingas veiksmas, sukeliantis ekonominius nuostolius į sukčiavimo veiksmus nukreiptai šaliai, o finansinę ar asmeninę naudą – vykdančiajai“ (Rai, 2017). Galima aptikti ir apibrėžimų, kurie sukčiavimą įvardina kaip tyčinį veiksma, kurio metu pasisavinama kitam asmeniui priklausanti nuosavybė ar pinigai, naudojant neteisėtas priemones (ACFE, 2016). Lietuvos Respublikos baudžiamajame kodekse (2022) sukčiavimas įvardijamas, kaip nusikalstama veikla, kuria kaltininkas apgaule savo ar kitų asmenų naudai įgyja svetimą turtą ar turtinę teisę, išvengia turtinės prievolės arba ją panaikina. Europos sąjungos direktyvoje (2017) sukčiavimas apibrėžiamas kaip tyčinis apgaulės veiksmas, kuriuo siekiama asmeninės naudos arba nuostolių padarymo kitai šaliai. Taigi analizuojant pateikiamus apibrėžimus, sukčiavimą galima suprasti kaip veiksma, kurio metu siekiama asmeninės naudos neteisėtais metodais. Nagrinėjamos problemos aspektu, sukčiavimą laikome, kaip neteisėtais veiksmais įgytas skolintas lėšas be ketinimo jas gražinti ir sumokėti už įgytą paslaugą.

Sukčiavimo rizika, lyginant su kredito, yra kur kas kompleksiškesnė, mažiau ištyrinėta ir skirtingose rinkose sukčiavimą gali skatinti skirtingi veiksniai, priklausantys nuo skolininkų įvairių tiek asmeninių, tiek demografinių ar kitų savybių (Agarwal ir Sharma 2014). Investuotojams svarbu įvertinti abejas rizikas, siekiant pasirinkti optimalų investavimo sprendimą, atsižvelgiant į savo galimus patirti nuostolius.

Augant tarpusavio skolinimo platformų skaičiui ir populiarumui visame pasaulyje, siekiant sumažinti galimas kredito bei sukčiavimo rizikas, iškyla poreikis užtikrinti tinkamą šio finansavimo šaltinio reguliavimą. Visi finansų rinkos dalyviai turi būti prižiūrimi ir kontroliuojami, nes šalies finansų rinka yra neatsiejama nuo valstybės ekonominės padėties (Vorobjova ir Kartašova, 2015; Abbasi et al. 2021). Dėl griauamo vartotojų pasitikėjimo teisėtais verslais, finansinis sukčiavimas gali sutrikdyti įprastą ekonominę veiklą bei daryti neigiamą įtaką ekonomikos efektyvumui (Brenner et al., 2020)

Užsienio finansų rinkoje, kaip jau minėta, P2P pirmiausia atsirado bei įsitvirtino Jungtinėje Karalystėje. 2019 metų duomenimis, pagal išduotų paskolų portfelio dydį sutelktinio finansavimo platformose pirmaujančios šalys buvo šios: Kinija (58 491 mln. USD.); Jungtinės Amerikos Valstijos (24 068 mln. USD); Jungtinė Karalystė (1 800 mln. USD); Vokietija (233 mln. USD); Prancūzija (229 mln. USD). Nepaisant lyderiavimo išduotų kreditų rinkoje P2P srityje, tarpusavio skolinimosi rinka Kinijoje pradėjo trauktis ir 2020 metų lapkričio 27-ąją dieną, Kinijos bankininkystės ir draudimo reguliavimo komisija paskelbė, jog iš beveik 5000 P2P platformų neliko nei vienos gyvuojančios tarpusavio skolinimosi platformos (Huang ir Wang, 2021) pagal suteiktų kreditų skaičių lyderiavusioje šalyje (žr. 1 pav.)



**1 pav.** Kinijos tarpusavio skolinimosi platformų skaičius vnt. (2007–2020). Sudaryta Rodenbaugh (2020)

Žlugus P2P rinkai Kinijoje, kitose valstybėse šis verslo modelis ir toliau sėkmingai plėtojamas ir prognozuojama, kad pasaulinė P2P rinka iki 2030 metų išaugs 4,3 karto, lyginant su 2022 metais ir sieks 705,81 mlrd. USD (Precedence Research, 2022). Todėl svarbu išanalizuoti P2P žlugimo priežastis, kad to būtų galima išvengti kitose stabiliose, augančiose rinkose.

Mokslininkai H. N. Pontel'is ir L. Huang'as (2022), Gu et al. (2022) prieina prie bendros išvados – P2P skolinimosi reguliacinė aplinka Kinijos rinkoje palengvino finansinių nusikaltimų atlikimą ir sukčiavimas buvo pagrindinis P2P skolinimosi rinkos žlugimo veiksnys. Tarpusavio skolinimosi platformos nukrypo nuo pagrindinio verslo modelio – buvimo tarpininku tarp kredito gavėjo ir kredito teikėjo, dėl to sukčiavimui tenkančią riziką pradėjo prisiimti ne investuotojai, o tarpusavio skolinimo platformos operatoriai (TSPO) (Huang Y., 2022). Siekiant nustatyti, kodėl tik Kinijos P2P rinkoje buvo nukrypta nuo pirminio verslo modelio, atliktas lyderiaujančių pagal išduotų P2P paskolų sumą valstybių tarpusavio skolinimosi platformų reguliavimo palyginimas (žr. 1 lentelę).

**1 lentelė.** TSPO reguliavimo lyderiaujančiose valstybėse apžvalga. Sudaryta remiantis: Huang ir Wang, (2021); Nemoto, Huang ir Storey (2019); Chiu (2016).

Šalis	Regulatorius	Reguliavimas
Kinija	Kinijos vyriausybė	Iš pradžių P2P veikla nebuvo reguliuojama, tačiau vėliau, pradėjus įmonėms patirti nuostolių, tarpusavio skolinimosi platformų veikla buvo apribojama tik tarpininkavimo modeliu.
Jungtinės Amerikos Valstijos	Vertybinių popierių ir biržos komisija (angl. <i>Securities and Exchange Commission, SEC</i> ), Vartotojų finansinės apsaugos biuras, Federalinė prekybos komisija.	JAV reikalavo, kad tarpusavio skolinimosi platformos atitiktų esamus finansų rinkos saugumo reglamentavimus.
Jungtinė Karalystė	Karalystės finansinių paslaugų priežiūros institucija. (angl. <i>Financial Conduct Authority, FCA</i> )	Dialogu paremtas reguliavimas. Buvo įsteigta reguliavimo „smėlio dėžė“, leidžianti „finansines naujoves įgyvendinti eksperimentiniais būdais pagal reguliavimo patvirtinimo ir stebėjimo parametrus“ (Chiu, 2016)

Šalis	Regulatorius	Reguliavimas
Vokietija	Federalinė finansų priežiūros institucija (BaFin)	Tarpusavio skolinimosi veikla reglamentuojama kartu su visų finansinių technologijų veikla Vokietijos bankininkystės įstatymu ir Vokietijos investicinės institucijos aktu. Atskirų įstatymų P2P skolinimui nėra.
Prancūzija	Prancūzijos teisė	Kaip ir Vokietijoje, P2P skolinimosi įstatymų nėra. Šių platformų veikla reglamentuojama bankininkystės įstatymais.

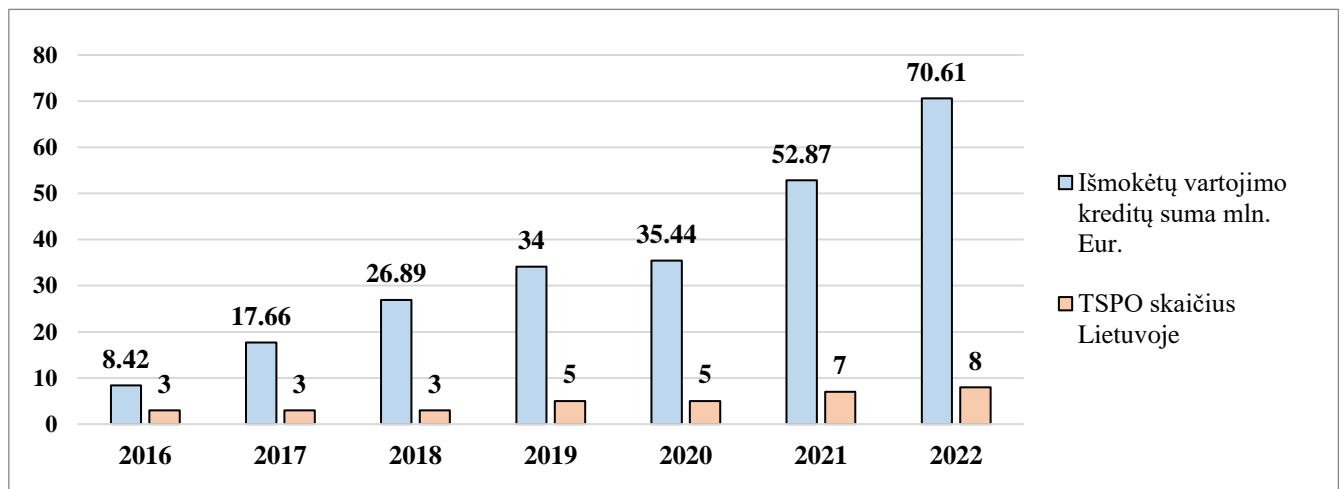
Kinijoje pradėjus plėstis P2P rinkai, vyriausybė neskyrė daug pastangų reguliuoti šį naują verslo modelį, literatūros šaltiniuose teigiama, kad vyriausybė buvo „atitraukusi rankas“ nuo reguliavimo ir tik prasidėjus rinkos krizei ėmėsi griežto kontroliavimo (Huang ir Wang, 2021). Jungtinėse Amerikos Valstijose, Vokietijoje bei Prancūzijoje pradėjus kurtis tarpusavio skolinimosi platformoms, jos buvo įpareigojamos laikytis jau vyraujančių finansų rinkų reglamentavimų. Tuo tarpu valstybėje, kurioje startavo P2P verslo modelis – Jungtinėje Karalystėje, buvo sudaryta galimybė steigti reguliacines „smėlio dėžes“ (angl. *Sand-box*). Reguliacinė „smėlio dėžė“ – galimybė išbandyti inovatyvias idėjas realioje aplinkoje, prižiūrint reguliatoriaus ir teikiant konsultacijas (Lietuvos bankas, 2018). Taigi P2P sektorius Jungtinėje Karalystėje buvo ne tik prižiūrimas, bet ir sudarytos palankios sąlygos toliau plėtoti tarpusavio skolinimosi platformų verslo modelį.

Taigi matome, kad Kinijoje TSPO reguliavimas buvo, tačiau operatorių veikla pradėta reguliuoti per vėlai. 2017 metais Kinijoje išduotų kreditų suma buvo pasiekusi piką ir sudarė 327 800 mln. USD, tuo tarpu Jungtinių Amerikos Valstijų P2P sektoriaus išduotų paskolų portfelis sudarė 17 340 mln. USD, Jungtinės Karalystės – 6 005 mln. USD (Huang. Y. 2022). Kinijos P2P sektoriui pasiekus piką, tuo pat metu platformoms nebuvo taikomi tokie patys apribojimai kaip ir kitoms finansinėms įstaigoms, tai sudarė sąlygas „šešėliniam“ verslui, dėl to, siekdami kuo didesnės rinkos dalies, P2P rinkos dalyviai prisiėmė nepamatuotą riziką, teigdami investuotojams, kad užtikrina šimtaprocentinį investuotos sumos grąžinimą (Jiang et al., 2021). Didelė dalis Kinijos P2P platformų susidūrė su sukčiavimo atvejais, kai klientai pasisavinę pinigus pradingdavo, o TSPO buvo įpareigoti išpirkti investuotojų skolą išmokant investuotą pradinę sumą ir perimti investiciją. Kinijoje pirmasis reguliavimo raštas Centrinio Banko buvo išleistas 2015 metais (Shen W. Ir Yu T., 2019), o pirmoji P2P platforma įsteigta 2006 m. Dėl negriežto P2P reguliavimo dalis investuotojų taip pat tapo sukčiavimo aukomis, kai TPSO, siekdami išsaugoti bent dalį pelno, pradėjo užsidarinėti ir „pabėgti“ su investuotojų lėšomis (Gu et al.). Taigi galima teigti, jog pagrindinės Kinijos P2P rinkos žlugimo priežastys buvo nepakankamas ir pavėluotas reguliavimas valstybės atžvilgiu, kuris sudarė palankias sąlygas sukčiavimo atvejams. Todėl yra svarbus ne tik tinkamas bei savalaikis valstybės reguliavimas, bet ir tikslus galimos sukčiavimo rizikos įvertinimas, kuris Kinijos atveju galėjo išsaugoti visą tarpusavio skolinimosi rinką šalies mastu.

Europoje tarpusavio skolinimo rinka iki šiol perspektyviai auga. 2021 metų pabaigoje ji sudarė 13,67 mlrd. USD. Per artimiausius septynerius metus prognozuojama, kad rinka išaugs 5,7 karto ir iki 2028 metų sudarys 78,11 mlrd. USD (*Verified Market Research*, 2021). Europos augančioje tarpusavio

skolinimo rinkoje neišvengiamai didėja ir apgaulės atvejų, tarp kurių vis dažniau pasitaiko ir stambių apgaulių. Jungtinėje Karalystėje veikusi skolinimosi platforma „Funding Circle“ tapo stipraus sukčiavimo auka klientų duomenims vertinti priėmusi suklastotus dokumentus. Platformos nuostolis sudarė 27 mln. EUR. ir akcijų vertės kritimą (Labunskiy, 2023). Kiti dažniausiai pasitaikantys sukčiavimo atvejai susiję su netinkamu klientų identifikavimu bei pinigų pervedimu ne į pareiškėjo sąskaitą. Vis dėl to, Europos finansinės rinkos sistema stipriai prižiūrima reguliatoriaus ir šiuo metu tarpusavio skolinimo rinkos sektorius laikomas stabilu.

Lietuvoje, kaip Europoje tarpusavio skolinimosi platformų rinka augo ir iki 2022 m. pabaigos Lietuvoje įregistruoti aštuoni tarpusavio skolinimosi platformų operatoriai (žr. 2 pav.).



**2 pav.** TSPO skaičius ir per metus išmokėtų vartojimo kreditų suma Lietuvoje (mln. Eur.). Sudaryta remiantis Lietuvos banko duomenimis 2022.

Lietuvos TSPO išduodamų kreditų suma Eurais kasmet auga. 2022 metais per aštuonias P2P platformas Lietuvoje buvo paskolinta 70,61 mln. Eur., t.y. 33.5 proc. daugiau nei 2021 metais (žr. 2 pav.). Didėjant TSPO užimamos finansų rinkos daliai svarbu užtikrinti ir tinkamą reguliavimą, siekiant sumažinti nusizengimų bei galimų sukčiavimo atvejų skaičių. Visi Lietuvos finansų rinkos dalyviai turi atskaitomybę Lietuvos bankui, kurio pareiga yra: “Vienodų sąlygų užtikrinimas /.../. Bet kokia prielaida incidentui atsitikti gali menkinti pasitikėjimą visu finansų sektoriumi“ (Zalatorius, 2017). Lietuvoje tarpusavio skolinimasis buvo pradėtas reglamentuoti 2015 metais prieš pradėdant steigti pirmosioms P2P platformoms. Tarpusavio skolinimosi platformų veikla reglamentuojama Lietuvos Respublikos vartojimo kredito įstatymu. Įstatymas nurodo, kad P2P skolinimosi metu tiek investuotojas, tiek vartojimo kredito gavėjas yra fizinis asmuo, kuriems tarpininkauja tarpusavio skolinimosi platforma.

Tarpusavio skolinimosi platformų operatorius taip pat turi reguliatoriaus nurodytą įpareigojimą įvertinti klientų kreditingumą, gebėjimą vykdyti finansinius įsipareigojimus laiku ir pateikti surinktą informaciją investuotojams (Lietuvos bankas, 2013). TSPO atsižvelgiant į Lietuvos banko „Dėl vartojimo kredito gavėjų mokumo vertinimo ir atsakingojo skolinimo nuostatų patvirtinimo“ (2013) nutarimą įprastai viešai pateikia kreditų gavėjų gaunamų pajamų, mėnesinių įsipareigojimų dydį, šeimyninę padėtį, šeimos

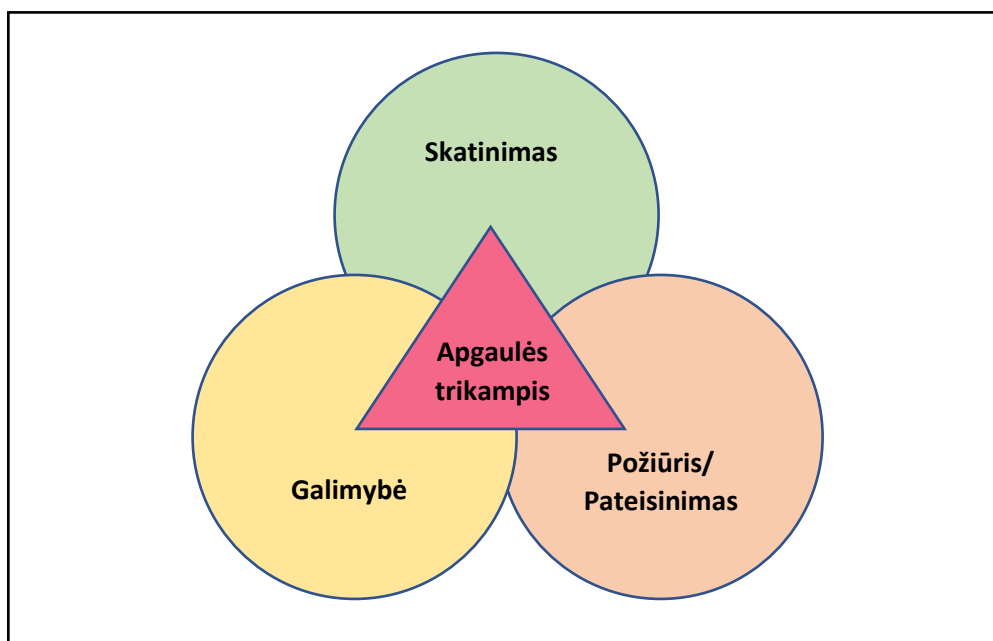
sudėtį, vartojimo kredito naudojimo paskirtį. Pagal turimus kliento duomenis bei „Creditinfo Lietuva“ (kredito biuro kaupiančio bei analizuojančio Lietuvos fizinių ir juridinių asmenų kreditingumo duomenis) duomenų bazėje pateikiamą informaciją tarpusavio skolinimosi platformoje pateikiamas ir kliento nemokumo rizikos vertinimas priskiriant klientui kreditingumo reitingą. Dažniausiai reitingas pateikiamas raidėmis nuo A (aukščiausias saugumo lygis) iki E (įsipareigojimų nevykdymas). Reitingas suteikiamas atsižvelgiant į asmens mokėjimų istoriją, kredito istorijos trukmę, ekonominius duomenis bei kreditavimosi aktyvumą (MANO CREDIT INFO, 2022). Kreditingumo reitingas atitinkamai nurodo ir kredito rizikos lygį. Investuotojas turi pasirinkimą tiek vadovautis P2P skolinimo platformos operatoriaus atliktu kliento kreditingumo vertinimu priimant sprendimą, tiek pats atlikti kliento mokumo bei galimų kredito bei sukčiavimo rizikų vertinimą remiantis pateikiamais duomenimis.

Taigi tarpusavio skolinimasis yra dar pakankamai naujas finansų rinkos dalyvis kaip alternatyva tradicinėms pinigų skolinimo įstaigoms, atlikdamas finansų tarpininko vaidmenį tarp pinigų skolinčio ir besiskolinančio asmens. Nepakankamas finansų rinkų dalyvių reguliavimas sudaro palankias aplinkybes sukčiavimo atvejams. Esant nepakankamai griežtam reguliavimui, lėšas skolinančiam asmeniui yra svarbu įvertinti ne tik investicijai tenkančią kredito, bet ir sukčiavimo riziką, ir atsižvelgiant į vertinimą priimti tinkamiausią investavimo sprendimą.

## **1.2. Sukčiavimo riziką didinančių veiksnių analizė**

Analizuojant mokslinėje literatūroje atliktus tyrimus (Sahin ir Duman, 2011; Xu, Jennifer, Yong, ir Michael 2015; Ki ir Yo, 2017 ir kt.), atliekant sukčiavimo rizikos vertinimą pirmasis žingsnis prieš modelio sudarymą yra duomenų surinkimas ir svarbiausių statistiškai reikšmingų faktorių nustatymas, kurie suteikia galimybę sėkmingai identifikuoti sukčiavimo atvejus. Į modelį įtraukiant nereikšmingus kintamuosius modelis praranda tikslumą (Date, 2023), todėl yra būtina atsakingai atrinkti veiksnus iš viso turimo duomenų rinkinio.

Sukčiavimo vertinimo moksliniuose darbuose dažnai sutinkamas „Apgaulės trikampis“ 1970 m. pasiūlytas D. R. Cressey'io (žr. 3 pav.).



**3 pav.** Apgaulės trikampis. Sudaryta remiantis D. R. Cressey'io teorija (1970)

Pagrindinė Apgaulės trikampio idėja – sukčiavimas sąlygojamas trijų sąlygų: skatinimo, galimybės bei požiūrio/pateisinimo. Asmuo išorinės aplinkos gali būti skatinamas atlikti apgaulės veiksmą. Tai gali būti psichologinis skatinimas norint pakenkti kitam arba asmeninės naudos siekiantis skatinimas, jei neteisėtai įgytomis lėšomis bus dalijamasi. Antroji apgaulės sąlyga - išvelgiamos neteisėto pasipelnymo galimybės. Galimybės skatina užsiimti apgaulinga veikla, taigi įmonių jau aptartoje reguliacinėje aplinkoje neturi būti palikta neapgalvotų aspektų, kurie skatintų pasinaudoti „sistemos silpna vieta“ siekiant gauti asmeninės naudos. Trečioji apgaulės sąlyga – požiūris/pateisinimas. Šių sąlygų atsiradimą sąlygoja asmeninės asmens charakterio savybės. Ar asmuo sukčiavimą suvokia kaip nusikalstamą veiklą ar kaip neišvengiamai reikalingą veiksmą, iš kurio gaunama daugiau naudos nei žalos.

Taigi atsižvelgiant į jau 50 metų gyvuojančią apgaulės trikampio koncepsiją, vertinant sukčiavimo riziką į modelius dažniausiai įtraukiami klientų elgesio veiksniai, tokie kaip klientų požiūris, motyvacija, gebėjimas sukčiauti, pasitenkinimas esama finansine padėtimi, godumas ir pan. (Rahamn et al. ,2014; Avortri ir Agbanyo, 2020) bei sudaromi ir analizuojami elgesio modeliai. Vis dėl to, tarpusavio skolinimosi platformose, įprastai prie pateikiamos informacijos apie lėšas renkančią paskolą nėra pateiktas skolininko klausimynas apie asmenines savybes: ar jis linkęs į godumą, ar patenkintas savo finansine padėtimi ir kiti klientą apibūdinantys klausimai. Dėl to elgesio modeliai netinka P2P sukčiavimo rizikai vertinti pirminėje stadijoje sprendžiant ar klientui galima suteikti kreditą pagal jo pateiktus atsakymus. Atsižvelgiant į tai, vertinant tarpusavio skolinimo platformoms tenkančią sukčiavimo riziką reikalinga sudaryti ne naudotojo elgesio analizės, bet sukčiavimo rizikos vertinimo modelį.

Sudarant sukčiavimo rizikos vertinimo modelius mokslinės literatūros šaltiniuose naudoti sukčiavimo rizikos veiksniai pateikiami 2-oje lentelėje.

## 2 lentelė. Sukčiavimo rizikos veiksniai.

Autoriai	Veiksniai
Sahin ir Duman'as (2011)	Kreditinės kortelės panaudojimo vieta, laikas, operacijų kiekis, tipas, suma.
Xu, Jennifer, Lu, ir Chau (2015)	Paskolos prašymo ir mokėjimų įrašų dažniai, išsilavinimas, profesiniai pasiekimai, įsipareigojimų ir pajamų santykis.
Ki ir Yoon (2017)	Didelis operacijų dydis, suma ir pirkimo operacija nereikalaujanti asmens tapatybės patvirtinimo.
Huangb ir Essa (2017)	Kredito reitingas, kredito suma, kliento metinės pajamos, įsipareigojimų ir pajamų santykis, kredito užklausų skaičius per pastaruosius 6 mėnesius, turimų įsipareigojimų skaičius.
Wang (2018)	Finansinė kliento informacija (pajamos, įsipareigojimai); Darbovietės informacija (gaunamų pajamų tipas, pajamų trukmė); Istorinė kredito informacija (Buvusių paskolų/kreditų suma, kreipimosi dėl kreditų/lizingų/paskolų skaičius); Demografinė informacija (šeimos narių skaičius ir pan.).
Zhang, Zerui, ir Shaoli Jin. (2020)	Paskolos suma; vidutinė tikėtina grąžos norma; vidutinis paskolos terminas; paskolos likutinė vertė; kiek laiko naudojamas P2P platformomis mėnesiais, Skolinimų skaičius; kredito istorija; investuotojų skaičius, vidutinė investicija, skolininkų skaičius.
Barua et al. (2021)	Palūkanų norma, atlyginimas, kredito suma, pateiktų dokumentų teisinis patvirtinimas.
María Óskarsdóttir et al. (2022)	Draudžiamasis įvykis, suma, draudėjas, draudėjo praeities pretenzijų skaičius.

**Suma.** Beveik visuose tyrimuose, kaip vienas pagrindinių veiksnių išskiriama operacijos/kredito suma. Sukčiaujantys asmenys yra linkę pirkti brangesnius daiktus, imti kreditus didesnėmis sumomis (Ki ir Yoon, 2017). Ši sukčiavimo logika suprantama, klientų ketinimu pasisavinti kuo didesnę svetimų lėšų kiekį su kuo mažiau operacijų, taip sumažinant tikimybę, kad sukčiavimas bus pastebėtas bei sustabdytas. Pavyzdžiui, vieną kartą pasiskolinus 100 eurų ir negrąžinus, tikimybė kitą kartą gauti kreditą toje pačioje P2P platformoje sumažėja. Todėl pirmu skolinimusi, neketinant grąžinti lėšų, paranku pasisavinti kuo didesnę piniginę sumą.

**Pajamos.** Kitas dažnai į tyrimus įtraukiamas veiksnys - uždirbama kliento pinigų suma, kuri taip pat gali daryti įtaką sukčiavimui. Kuo mažesnės kliento pajamos, tuo didesnis nepasitenkinimas esama finansine situacija, kuris gali motyvuoti sukčiauti (Rahamn et al. ,2014).

**Įsipareigojimai.** Atitinkamai kaip ir kliento pajamos, dideli turimi įsipareigojimai skatina gauti papildomų pajamų senesniems įsipareigojimams dengti. Jei klientas įsisuka į skolinimosi ratą, skolinasi pinigus senoms skoloms dengti, tai sudaro didelę tikimybę, kad žmogus taps nemokiu.

**Istorinė kredito informacija.** Dar vienas bendras veiksnys literatūros šaltiniuose - kliento kredito istorija. Jei matoma, kad klientas nėra linkęs vykdyti įsipareigojimų laiku, neatmetama tikimybė, kad naująjį kreditą klientas mokėti intencijos taip pat neturės. Dėl nemokaus kliento istorinių kreditingumo duomenų, jam teikiamas paskolos pasiūlymas su didesnėmis palūkanų normomis lyginant su tvarkingą



kredito istoriją turinčiu, laiko patikrintu, mokių klientu (Lerner, 2022). Atsižvelgiant į tai, **palūkanų norma**, tai pat gali būti veiksnys indikuojantis polinkį į sukčiavimą.

**Lytis, amžius, šeimyninė padėtis, darbo stažas** – veiksniai, kurie pasak Yingying, Ying‘o ir Yan‘o (2019) daro įtaką kliento ekonominei situacijai, kuri atitinkamai arba skatina, arba neskatina kliento linkti į sukčiavimą. Vyrai išskiriami kaip labiau linkę rizikuoti bei prisiimti papildomą tenkančią riziką, kaip ir šeimyninio gyvenimo nevedantys asmenys (vieno asmens namų ūkis). Trumpas darbo stažas taip pat įvardijamas kaip sukčiavimo riziką didinantis veiksnys, nes įdarbinimas gali būti fiktyvus. Kuo ilgiau klientas dirba toje pačioje įmonėje tuo tvaresnėmis jo pajamas galime laikyti (Wang, 2018).

**Išsilavinimas.** Lee‘o ir Soberon‘o-Ferrer‘o (2005) atlikto tyrimo rezultatai parodė, kad sukčiavimui įtaką daro ne tik minėtieji veiksniai kaip, amžius, vieno asmens namų ūkis, finansinė padėtis, bet ir žemas išsilavinimas. Mažiau išsilavinę asmenys inicijavo reikšmingai daugiau sukčiavimo atvejų nei didesnį išsilavinimą turintys asmenys. Tačiau autoriai nustatė, kad nei lytis nei rasė reikšmingo poveikio asmens polinkiui į sukčiavimą – neturi.

**Skolintų pinigų panaudojimas.** Avoyem‘is et al. (2017) teigia, kad sukčiavimo rizika dažniausiai susijusi su tuo kaip ir kur skolinti pinigai išleidžiami. Ar jie naudojami kitoms skoloms dengti, ar išleidžiami iš karto tiesiai iš banko sąskaitos, ar išgryninami. Pinigų išleidimo tipo analizavimas gali padėti iš anksto identifikuoti sukčiaujančią asmenį.

Taip pat į tyrimus įtraukiami ir kiti veiksniai kaip kreipimosi dėl kredito dažnis – kuo dažniau klientas kreipiasi dėl papildomos kredito sumos, tuo labiau didėja tikimybė, kad jis nustos vykdyti įsipareigojimus laiku. Pateiktų dokumentų teisinis patvirtinimas – jei dokumentai nepatvirtinti parašais, kyla tikimybė, kad jie yra suklastoti ir kreditą ketinama pasiimti apgaule (Barua et al. 2021). Kliento naudojimosi paskolų tarpusavio platforma laikotarpis – kuo naujesnis klientas, tuo mažiau jis pažintas, tuo didesnė tikimybė, kad kreditas nebus sugražintas, itin jei jo kredito istorija nėra „švari“ (Zhang, Zerui, ir Shaoli., 2020).

Taigi daug įvairių veiksnių gali sąlygoti kredito sukčiavimo riziką. Ir nors dažniausiai sukčiavimui nustatyti yra naudojami elgesio vertinimo modeliai, atsižvelgiant į prieinamą informaciją, yra svarbu sudaryti sukčiavimo analizės modelį remiantis techninėmis paskolos charakteristikomis (kredito suma, laikotarpis, palūkanų norma ir t.t.) bei kliento asmenine, demografinė informacija (pajamos, įsipareigojimai, šeimos narių skaičius, amžius, kredito istorija ir t.t.). Sudarius sukčiavimo analizės modelį galima daryti prielaidas, dėl kokių gyvenimo aplinkybių asmenys yra linkę į sukčiavimą, susiformuoti sukčiaujančiojo asmens profilį ir vengti teigti paskolas klientams atitinkantiems apsibrėžtaji profilį.

Mokslinėje literatūroje randama prieštarų nuomonių - autoriai išskiria skirtingus veiksnius lemiančius polinkį į sukčiavimą. Skirtingų faktorių išskyrimas gali būti suprantamas dėl nuolat kintančio sukčiavimo modelio, kuris vis progresuoja siekiant atrasti naujas apgaulės galimybes. Taip pat prieštarų veiksnių įtraukimas į modelius indikuoja ir kiekvienoje finansų rinkoje esančią unikalią reguliacinę sistemą, kuri gali palikti daugiau arba mažiau galimybių sukčiavimui. Taigi darbo antroje dalyje, atsižvelgiant į

išanalizuotą literatūrą, išrinkti reikšmingi veiksniai, kurie įtraukiami į sukčiavimo rizikos vertinimo modelį.

### 1.3. Sukčiavimo rizikos vertinimo modelių analizė

Sukčiavimas nėra lengvai pastebimas, tačiau visose sferose gali pridaryti didelių nuostolių. Dėl šios priežasties yra svarbu identifikuoti galimą sukčiavimo riziką ir užkirsti kelią apgaulei įvykdyti. Mandrijauskaitė (2018) teigia, kad sukčiavimo rizikos matavimas yra itin sudėtingas ir kompleksiškas procesas. Pasak Sorournejad'o et al. (2016) egzistuoja du sukčiavimo aptikimo būdai – sukčiavimo analizė ir naudotojo elgesio analizė. Sukčiavimo analizė analizuoja istorinius duomenis ir ieško nuo normos nukrypstančių veiklos rodiklių. Naudotojo elgesio analizė yra nekontroliuojama metodika priklausanti nuo kliento elgesio modelio (Pratuzaitė ir Maknickienė, 2018). Šie metodai skiriasi tuo, kad sukčiavimo analizės metu naudojami istoriniai duomenys ir sukčiavimo rizikos aptikimas priklauso nuo jau turimų sukčiavimo įrašų. Tuo tarpu naudotojo elgesio analizė gali nukrypti į netipinių sukčiavimų atpažinimą.

3 lentelėje pateikiami sukčiavimo rizikos vertinimo modeliai pasiūlyti nagrinėtos mokslinės literatūros autorių.

#### 3 lentelė. Rizikos vertinimo modeliai

Autoriai	Tyrimas	Naudotas vertinimo modelis
Kurt, Kenneth, ir Rajendra (1995)	Valdymo sukčiavimo aptikimas: neuroninio tinklo metodas.	Dirbtiniai neuroniniai tinklai.
Sahin ir Duman (2011)	Kreditinių kortelių sukčiavimo aptikimas naudojant sprendimų medžius ir atraminių vektorių mašinas.	Sprendimų medžiai ir atraminių vektorių mašinos.
Xu, Jennifer J., Yong Lu, ir Chau. (2015)	P2P skolinimo sukčiavimo aptikimas: didelių duomenų metodas.	Sprendimų medžiai, atraminių vektorių mašinos, neuroniniai tinklai.
Ki ir Yoon (2017)	Pirkimo tankumu pagrįsta internetinė sukčiavimo kredito kortelėmis aptikimo sistema.	Puasono skirstinio funkcija.
Wang (2018)	Nesažiningų vartotojų aptikimas P2P finansų rinkoje.	Atsitiktiniai miškai ir gradiento didinimo sprendimų medžiai.
Zhang, Zerui, ir Shaoli Jin. (2020)	P2P rinkos rizikos vertinimas ir matavimas.	Faktorinė analizė, klasterinė analizė.
Barua et al. (2021)	Sukčiavimas: paskolos įsipareigojimų nevykdymo tikimybės numatymas naudojant CatBoost algoritmą.	CatBoost algoritmas.
Óskarsdóttir et al. (2022)	Socialinio tinklo analizė, skirta prižiūrimam sukčiavimui aptikti draudime.	Logistinė regresija, atsitiktiniai miškai.

F. Kurt'as, O. C. Kenet'as. Ir S. Rajendra (1995) moksliniame darbe tyrė valdymo sukčiavimo atvejį. Valdymo sukčiavimas suprantamas, kaip vadovų piktybinis siekis pasipelninti klastojant įmonės dokumentus. Šiam sukčiavimui tirti buvo naudoti apklausų duomenys, kurie buvo apjungti į faktorius dėl nustatytos išvados, kad „vertinant riziką turi būti vertinami veiksmų rinkiniai, nes vienas veiksmas nebūtinai gali rodyti padidėjusią sukčiavimo riziką“. Tyrimui atlikti pasitelkti dirbtiniai neuroniniai tinklai. „Dirbtinis neuroninis tinklas yra matematinų modelių rinkinys, kuriuo bandoma emuliuoti įvairias biologinių sistemų savybes.“ (Bernotas E., 2009). Šie modeliai yra sudaryti iš tarpusavyje susijusių dirbtinių neuronų, kurie yra sujungti įvairių stiprumų jungtimis. Gavę informaciją dirbtiniai neuronai ją apdoroja ir pateikia išvesties rezultata. Autoriai tyrime lygina du dirbtinių neuroninių tinklų algoritmus - apibendrintą prisitaikančią neuroninių tinklų architektūrą (angl. *Generalized adaptive neural network architecture*, GANNA) ir prisitaikančią loginį tinklą (angl. *Adaptive Logic Network*, ALN). Sudarius modelius bei palyginus tikslumo matavimus, gautų modelių tikslumo rezultatai abiem atvejais statistiškai reikšmingai nesiskiria (GANNA tikslumas – 89%; ALN – 90%).

Sprendimų medžių bei atramininių vektorių mašinų tikslumas nusakant kredito kortelių sukčiavimo riziką buvo palygintas Sahin ir Duman'o (2011) moksliniame darbe. Darbas atliktas su realiais asmenų duomenimis, darbo metu pašalintas klasių disbalansas. Vadeika (2018) sprendimų medžius apibūdina kaip medžio struktūros algoritmą, kuris atvaizduoja sprendimus. Sprendimų medžių metodas gali susidurti su persimokymu, dėl ko gali reikėti atlikti medžio genėjimą. Sprendimų medžiai buvo sudaromi remiantis trim algoritmais: Klasifikavimo ir regresijos medžiai (angl. *Classification and Regression Tree* C&RT), Chi kvadrato automatinės sąveikos detektorius (angl. *Chi-square Automatic Interaction Detector*, CHAID) ir C5.0. C&RT prognozuoja kintamojo reikšmę atsižvelgiant į kitų kintamųjų reikšmes. Račys (2014) kaip pagrindinį šio algoritmo privalumą išskiria tai jog „atliekant skaičiavimus, daroma visų galimų imčių padalijimų analizė ir išrenkamas geriausias padalijimas“. Pagal CHAID metodą „kiekviename žingsnyje nustatomas stipriausią sąveiką su priklausomu kintamuoju turintis nepriklausomas kintamasis“ (Pukėnas, 2009). C5.0 algoritmo metu sprendimų medžiai generuojami atsižvelgiant į mokymo duomenų rinkinius. Palyginus visus sprendimų medžių algoritmus, didžiausiu tikslumu nustatant sukčiavimo riziką pasižymėjo C&RT algoritmas. Atraminų vektorių mašinos (angl. *Support vector machines*, SVM) – suranda tarp dviejų klasių tokią sprendimo ribą, kad atstumas tarp klasių būtų didžiausias (Vadeika, 2018). Tyrime SVM buvo naudota su polinominio, sigmoidinio, tiesinio ir spindulio paremta branduolio funkcija. Geriausias rezultatus iš šių funkcijų parodė linijinio branduolio funkcija (modelio tikslumo matas – 93.08%). Lyginant atsitiktinių miškų ir atraminų vektorių metodus, tyrime kaip tikslesnis sukčiavimo rizikos vertinimo modelis išskiriamas – atsitiktinių miškų metodas.

Xu, J. Jennifer, L. Yong'as ir M. Chau'as. (2015) darbe tyrė tarpusavio skolinimosi platformos sukčiavimo riziką Kinijoje. Autoriai pabrėžia, kad Kinijoje nėra nacionalinės gyventojų kreditingumo duomenų bazės, kas neleidžia patikrinti ar gaunami duomenys apie kredito gavėjus yra patikimi, o tai sudaro palankias sąlygas sukčiavimui plisti. Visų pirma siekiant sudaryti sukčiavimo vertinimo modelį autoriai išskyrė veiksmus, kurie atspindi didesnę polinkį į sukčiavimą. Vėliau išskirtiems faktoriams analizuoti buvo pasitelkiami atsitiktinių miškų, XGBoost bei giliųjų neuroninių tinklų algoritmai. Sudarytų modelių tikslumui įvertinti buvo naudojami preciziškumo ir išsamumo matai (atitinkamai angl. *precision ir recall*), tikslumas (angl. *accuracy*), plotas po ROC kreive (angl. *Area Under Curve - AUC*)

ir  $F$  įvertis (angl. *F score*). Visų modelių tikslumo matas svyruoja nuo 0.95 iki 0.96, išsamumo matas nuo 0.30 iki 0.48, preciziškumo nuo 0.13 iki 0.15, AUC nuo 0.932 iki 0.939,  $F$  score nuo 0.19 iki 0.22. Modelių įvertinimo rodikliai reikšmingai nesiskiria, tačiau nežymiai geriausius rezultatus parodė XGBoost algoritmas.

Y. Ki ir J.W. Yoon'as (2017) darbe tyrė sukčiavimą susijusį su neteisėtu kreditinių kortelių panaudojimu. Autoriai išskiria tris pagrindinius bruožus, kurie būdingi sukčiavimui: greitis – neteisėti veiksmai atliekami kaip galima greičiau; kiekis – siekiama sudaryti kuo didesnius sandorius, siekiant gauti maksimalią naudą, prekybininko kategorija – ieškoma paslaugų tiekėjų, kurių taikomos asmens duomenų apsaugos priemonės silpniausios. Norint įvertinti riziką visų pirma apskaičiuojami išskirtų veiksmų dažnumai remiantis istoriniais kliento duomenimis, vėliau įvertinama rizikos tikimybė, naudojant Puasono skirstinio funkciją. Puasono procesas yra gerai žinomas tikimybinio modeliavimo metodas, skirtas matuoti įvykių skaičiaus  $N(t)$  tikimybę laike  $t$ . Sudaryto modelio tikslumas vertinamas tikslumo ir išsamumo matais, plotu po ROC (angl. *Receiver operating characteristic*) kreive. Tyrimo metu nustatyta, kad sukčiavimo riziką tiksliausiai galima įvertinti turint bent tris kreditinės kortelės istorinius įrašus. Kadangi sukčiavimo rizikos modelyje naudojami kintamieji yra jau išduotų kreditinių kortelių istoriniai duomenys, kaip pagrindinis šio metodo trūkumas išskiriamas negalėjimas nustatyti sukčiavimo rizikos naujiems klientams. Kredito kortelių davėjai turi finansiškai išgalėti patirti sukčiavimo nuostolį pirmus du kartus kortelės gavėjui nevykdžius įsipareigojimų laiku.

H. Wang'as 2018m. sudarė modelį nesąžiningiems vartotojams aptikti P2P skolinimo finansų rinkoje. Kaip ir Xu, J. Jennifer, L. Yong'o ir M. Chau'o (2015) bei Sahin ir Duman'o (2011) darbuose, tyrime remtasi sprendimų medžiais. Naudotas atsitiktinių miškų algoritmas bei gradiento didinimo sprendimų medžiai (angl. *Gradient boosting decision tree*). Šie metodai pasirinkti dėl lengvai koreguojamų parametrų bei aukšto rezultatų tikslumo. Atsitiktinis miškas yra mokymosi algoritmas, kuris apibendrina regresijos arba klasifikavimo medžių grupės rezultatus (Wang, 2018). Gradiento didinimo sprendimų medžiai įvardijami kaip viena sėkmingiausių gradiento didinimo algoritminių paradigmu. Darbe sudaryto modelio tikslumui vertinti pasirenkamas AUC matas. Atlikus tyrimą nustatyta, kad geriausiai turėtiems duomenims tinka gradiento didinimo sprendimų medžių algoritmas  $AUC = 0.88$ , tuo tarpu atsitiktinių miškų  $AUC = 0.83$ .

Zhang'as, Zerui, ir J. Shaoli. 2020 metais moksliniame tyrime vertino tarpusavio skolinimosi platformų riziką. Šiai rizikai vertinti buvo atlikta faktorinė analizė, kuria sugrupuojamos panašiausios tarpusavio skolinimosi platformos pagal bendrus požymius bei atlikta klasterinė analizė tyrimo tikslumui patikrinti. Faktorinė analizė leidžia susimąžinti dideles duomenų imtis, išskiriant keletą „bendrujų faktorių, kuriuose sukonzentruota informacija be nuostolių pakeičia informaciją, užfiksuotą dešimtyse požymių.“ (Šidlauskaitė, 2010). Atlikus faktorinę analizę buvo atlikta ir klasterinė analizė, leidžianti suskirstyti objektus į klasterius pagal panašius požymius. Ar duomenys tinkami faktorinei analizei buvo tikrinama pagal Kaizerio-Mejerio-Olkino kriterijų (KMO) bei Bartlett'o sferiškumo testą. Iš turėtų duomenų po faktorinės analizės buvo išskirti 5 apibendrinti faktoriai. Klasterizavimas buvo atliktas  $k$ -vidurkių metodu. Pagal galimą sukčiavimo rizikos dydį tyrimo metu išskirti 6 P2P platformų klasteriai.

Barua et al. 2021 metais atliekant sukčiavimo tikimybės vertinimą naudojo CatBoost algoritmą. CatBoost yra sprendimų medžių gradiento didinimo algoritmas, kuris pasižymi didesniu tikslumu už logistinės regresijos, gradiento didinimo bei šiek tiek didesniu tikslumu už atsitiktinių miškų algoritmus (Hembram et al. 2020). Barua et. Al. (2021) išskiriami Catboost algoritmo privalumai – gerai tvarkosi su kategoriniais kintamaisiais, yra greitesnis lyginant su kitais gradiento didinimo algoritmais, puikiai tinka didelėms duomenų imtims. CatBoost algoritmo gauti rezultatai buvo palyginti su atsitiktinių miškų bei gradiento didinimo metodais ir padaryta išvada, jog geriausius rezultatus parodė Catboost algoritmas.

Óskarsdóttir‘o et al. (2022) nagrinėjo draudimo išmokų sukčiavimo riziką. Darbe sudaryti modeliai naudojantis logistine regresija ir atsitiktiniais miškais. Logistinė regresija yra statistinis metodas, naudojamas tam tikro įvykio ar klasės tikimybei modeliuoti. Autoriai modeliui sukurti pasirinko logistinę regresiją dėl jos populiarus naudojimo draudimo srityje. Atsitiktinių miškų metodas buvo pasirinktas, nes jie įvardijami kaip „galingi ansamblio modeliai, galintys aptikti sudėtingus duomenų modelius“ (Breiman, 2001). Tyrimų tikslumui įvertinti, kaip ir kiti autoriai, Óskarsdóttir et al. (2022) naudojo AUC kreivę bei tikslumo ir išsamumo kreivę (angl. *precision–recall curve*). Tyrimo metu, kaip reikšmingiausi veiksniai didinantys sukčiavimo riziką išskiriami: operacijos suma, teistumas, amžius, kontraktų skaičius.

Išanalizavus atliktus 1995-2022 m. tyrimus, pastebima, kad dažniausiai sutinkami modeliai sukčiavimo rizikai nustatyti įvairiose srityse yra atsitiktinių miškų, neuroninių tinklų bei atraminių vektorių metodų algoritmu pagrįsti modeliai. 4 lentelėje pateikiamas palyginimas tarp populiariausių modelių.

**4 lentelė.** Atsitiktinių miškų, neuroninių tinklų ir atraminių vektorių metodų palyginimas. Sudaryta remiantis: Wainberg, Alipanahi ir Frey (2016); Płoński (2019).

Metodas	Tikslumas	Mokymo trukmė	Skaičiavimų trukmė	Apribojimas duomenims
Atsitiktiniai miškai	Aukštas	Greita mokymo trukmė	Greita skaičiavimo trukmė	Yra geresni kategoriškiems duomenims. Reikalauja mažesnės duomenų įvesties nei neuroniniai tinklai.
Neuroniniai tinklai	Neuroninių tinklų tikslumas paprastai nusileidžia atsitiktinių medžių modelių tikslumui.	Lėtesnė skaičiavimo trukmė	Lėtesnė skaičiavimo trukmė	Reikalauja didesnės duomenų įvesties nei SVM.
Atraminių vektorių mašinos	Aukštas, dažnai pranoksta atsitiktinių miškų algoritmą.	Įprastai ilgesnė mokymo trukmė.	Įprastai užtrunka ilgiau nei atsitiktiniai miškai ar neuroniniai tinklai.	Reikalauja mažesnės duomenų įvesties nei neuroniniai tinklai.

**Atsitiktinių miškų metodas.** Metodas buvo pritaikytas daugumoje analizuotų straipsnių. Šis modelis gali būti taikomas klasifikavimo bei regresijos uždaviniams spręsti. Atsitiktiniai miškai įprastai pasižymi aukštu tikslumu, greita mokymo bei skaičiavimo trukme (4 lentelė). Šio modelio principas yra - sudaryti

daug sprendimų medžių, paprastai nuo 100 iki 1000 sprendimų medžių iš mokymo imties. (Misius, 2019). Kiekvienam sprendimų medžiui priskiriama vis kita, individuali duomenų imtis, kas leidžia išvengti medžių genėjimo bei persimokymo problemos (Horning, 2010). „Testuojamas mėginys klasifikuojamas su visais gautais medžiais. Kiekvienas medis jį priskiria tam tikrai klasei. Apskaičiuojami gautų klasių pasikartojimo dažniai ir mėginys priskiriamas tai klasei, kurios pasikartojimo dažnis buvo didžiausias” (Meržvinskaitė, 2016). Horning‘as (2010) taip pat pabrėžia pagrindinį šio modelio trūkumą sprendžiant regresijos uždavinį – jei mokymo duomenų rinkinys neapima visų duomenų galimų pavyzdžių, tai pritaikius modelį pilnai duomenų imčiai taip pat nebus tikslaus įvertinimo tai duomenų kategorijai, kurios neapėmė mokymo imtis.

**Neuroninių tinklų metodas.** Neuroninis tinklas yra tarpusavyje sujungtų mazgų rinkinys, sukurtas imituoti žmogaus smegenų veiklą (Yufeng et al., 2004). Tinklas sudarytas iš įvesties, paslėptųjų bei išvesties sluoksnių. Neuroninių tinklų modeliai dažnai pasižymi aukštu tikslumu dėl savo kompleksiško bei galimybės turėti daug laisvų parametrų, kas suteikia jiems lankstumo prisitaikant prie labai sudėtingų duomenų, tačiau įprastai jų tikslumas nusileidžia atsitiktinių miškų modelių tikslumui. Kitas neuroninių tinklų trūkumas – sudėtingas algoritmo parametrų nustatymas bei reikalingas didesnis mokymo duomenų kiekis nei įprastai (Grigaliūnas, 2017). Kuo daugiau duomenų įvedama į tinklą, tuo geriau duomenys bus apibendrinti ir prognozės bus atliktos tikslesnės su mažiau klaidų.

**Atraminių vektorių mašinų metodas.** Metodo principas – duomenų aibėje sukurti tokią hiperplokštumą, kuri duomenis atskirtų į dvi klases (TICHONOV, 2018). Duomenys atskiriami į klases taip, kad atstumas tarp artimiausių skirtingų klasių objektų būtų didžiausias. SVM dažniausiai pasižymi aukštesniu tikslumu nei atsitiktinių miškų metodas. Skirtumas tarp pastarųjų minėtųjų modelių yra toks, kad SVM naudoja branduolio metodą, kad išspręstų netiesines problemas, o sprendimų medžiai įvesties erdvėje išveda hiper stačiakampius toms pačioms problemoms spręsti. Visgi sprendimų medžiai yra geresni kategoriniams duomenims ir jie geriau išlaiko kolinearumą nei SVM (Varghese, 2018). Taip pat Ray‘a (2023) Atraminių vektorių metodo trūkumus išskiria šiuos: netinka didelėms duomenų imtims dėl ilgos skaičiavimo trukmės, tiesiogiai nepateikia tikimybių įvertinimų.

Apibendrinant išanalizuotus modelius, naudotus sukčiavimo atvejų analizės vertinime, galime išskirti atsitiktinių miškų algoritmo pranašumą, kadangi metodas verčia atsižvelgti į visus galimus sprendimo rezultatus ir suteikia galimybę atsekti kiekvieną kelią iki padarytos išvados. Jis sukuria išsamią pasekmių analizę kiekvienoje šakoje ir nustato sprendimų mazgus, kuriuos reikia toliau analizuoti (Woodruff, 2019). Atsitiktinių miškų metodas nereikalauja didelės duomenų įvesties bei palyginus pasižymi trumpesne mokymo bei skaičiavimo trukme. Visgi literatūroje sutinkama teiginių, kad atraminių vektorių metodas dažnai pranoksta atsitiktinių miškų algoritmo tikslumą, o neuroniniai tinklai tiksliai sugeba prognozuoti kompleksinius modelius ir atpažinti dėsnius, kurių kiti algoritmai nepastebi. Dėl šių priežasčių privačių asmenų skolinimosi platformos sukčiavimo atvejų kreditų rizikos vertinimas atliekamas naudojant atsitiktinių miškų, atraminių vektorių mašinų bei neuroninių tinklų metodus siekiant juos palyginti ir išskirti geriausią metodą analizuojamai problemai spręsti.

Taigi analizuojamoji problema – sukčiavimo rizikos aptikimas, yra aktuali bei vis didėjanti dėl augančios tarpusavio skolinimosi platformų rinkos. Siekiant finansinio stabilumo tiek atskirų įmonių, tiek šalies

ekonominiu mastu, svarbu gebėti iš anksto įvertinti galimus sukčiavimo atvejus bei laiku užkirsti jiems kelią. Išsikeltai problemai spręsti, svarbu tinkamai suprasti tarpusavio skolinimosi platformų veiklą bei esamą reguliacinę sistemą, todėl darbe keliamas pirmasis uždavinys – išanalizuoti tarpusavio skolinimosi platformas bei jų reguliavimą. Lietuvos finansų rinkos dalyvių reguliatorius – Lietuvos bankas, griežtai prižiūri finansinę veiklą iš finansinių paslaugų teikėjo pusės. Todėl dažniausiai susiduriama su sukčiavimo atvejais inicijuotais kredito gavėjo. Siekiant įvertinti kokie veiksniai skatina klientus pasisavinti skolintas lėšas be intencijos jas mokėti ateityje, darbo antrasis uždavinys – išnagrinėti literatūros šaltiniuose išskiriamus sukčiavimo rizikos veiksnius bei naudotus sukčiavimo rizikos vertinimo modelius. Išanalizavus kitose rinkose naudotus sukčiavimo modelius, sudaromas naujasis modelis remiantis Lietuvos tarpusavio skolinimosi platformos UAB „Bendras finansavimas“ antrinėje rinkoje prieinamais paskolų duomenimis. Trečiojo darbo uždavinio tikslas – sudaryti sukčiavimo rizikos vertinimo modelį pritaikytą Lietuvos rinkos tarpusavio skolinimosi platformoms, leidžiantį iš anksto identifikuoti klientus linkusius į svetimų lėšų pasiskolinimą be ketinimo jas grąžinti.

## 2. Sukčiavimo atvejų nustatymo tyrimo metodologija

Šiame skyriuje pateikiama tyrimo objekto UAB „Bendras finansavimas“ apžvalga, atlikta tyrimo duomenų žvalgomoji analizė. Aprašoma tiriamojo darbo metodologija, kuria remiantis siekiama išsikelto tyrimo tikslo įgyvendinimo.

### 2.1. Tarpusavio skolinimosi platformos „SAVY“ analizė

Lietuvoje 2022 metais Lietuvos banko pateikiamais duomenimis buvo registruoti 8 tarpusavio skolinimosi platformos operatoriai, tarp jų ir UAB „Bendras Finansavimas“ (toliau – „SAVY“). „SAVY“ tai pirmoji tarpusavio skolinimosi platforma Lietuvoje įregistruota 2015 metais ir 2016 metais įrašyta į viešąjį tarpusavio skolinimo platformos operatorių sąrašą. Ši platforma siūlo sutelktinį investavimą į kitų žmonių paskolas, yra lietuviška bei kol kas veikia tik Lietuvos rinkoje (Smart Invest, 2021).

„SAVY“ investavimo platformoje galime pastebėti panašumą su Kinijos rinkoje buvusiu pasiūlymu klientams – 100 procentine paskolintos sumos grąžinimo garantija. „SAVY“ leidžia investuoti į paskolas su investuotojų fondu ir sumažinti investavimo riziką. Investuojant į paskolą su investuotojų fondu, investuotojui mokamos mažesnės palūkanos nei investuojant be fondo. Palūkanų skirtumas pervedamas į investuotojų fondą, iš kurio TSPO įsipareigoja kredito gavėjo nemokumo atveju kredito davėjui kompensuoti visą investuotą sumą bei iki tos dienos uždirbtas palūkanas. „SAVY“ neištinka Kinijos TSPO likimas dėl to, nes operatorius nerizikuoja savo asmeniniu turtu, kompensacinės lėšos mokamos tik iš investuotojų fondo. Kredito davėjams renkantys investuoti su investuotojų fondu, lėšų grąžinimui užtikrinti, reikia atkreipti dėmesį, kad lėšos už vėluojančias paskolas išmokamos tik esant pakankamam pinigų kiekiui fonde, todėl pinigų grąžinimas gali užtrukti.

UAB „Bendras finansavimas“ per praėjusius metus nepaisant padidėjusios vidutinės svartinės naujai išduotų paskolų palūkanų normos, nuo 14,8% iki 16,4%, stabiliai augino savo išduotų paskolų portfelį ir iki 2023 metų pradžios jau per „SAVY“ platformą yra sufinansuota paskolų daugiau nei už 68 mln. Eur. (Žiugžda, 2023). Kaip ir minėta darbo pradžioje, didesni finansinių operacijų kiekiai neatsiejamai susiduria ir su didesniu sukčiavimo skaičiumi. Tyrimo metu siekiant sudaryti sukčiavimo rizikos vertinimo modelį, naudojami viešai prieinami UAB „Bendras finansavimas“ paskolų duomenys pateikiami antrinėje paskolų rinkoje. Antrinėje rinkoje investuotojai susipažindami su paskolų informacija gali priimti sprendimą perpirkti investiciją iš kito investuotojo. Viso tyrimo metu surinkta duomenų apie 865 paskolas. Paskolas apibūdinantys kintamieji pateikiami 5-oje lentelėje.

**5 lentelė.** Turimas UAB „SAVY“ duomenų rinkinys

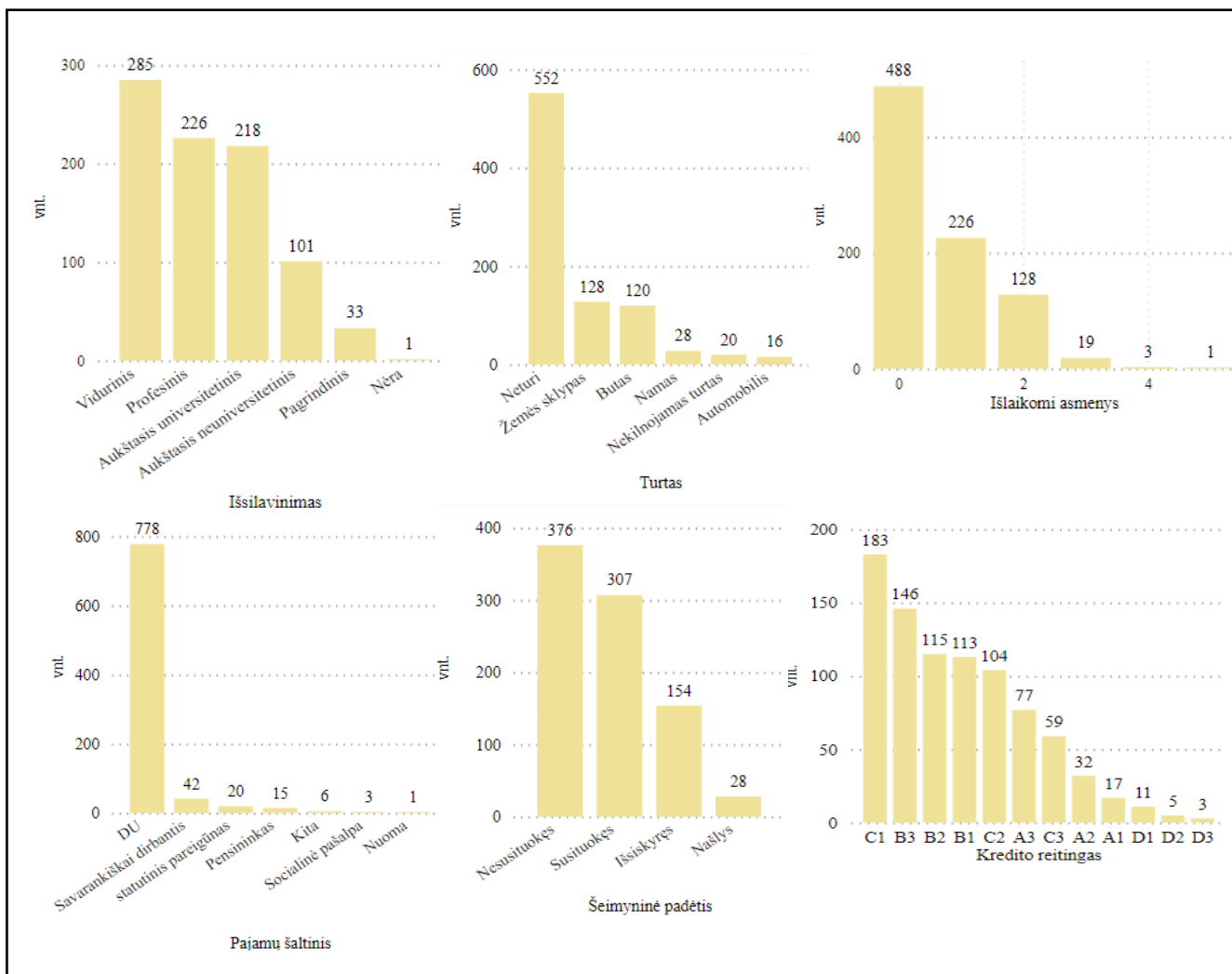
Nr.	Kintamasis	Aprašymas
1.	Suma	Kredito gavėjo skolintasi suma (eur.)
2.	Terminas	Laikotarpis mėnesiais, kuriam sudaryta sutartis.
3.	„SAVY“ kredito reitingas*	„SAVY“ apskaičiuojamas ir suteikiamas kredito reitingas



Nr.	Kintamasis	Kintamojo pavadinimas duomenų rinkinyje	Aprašymas
4.	Nemokumo rizika	Rizika	„SAVY“ apskaičiuota galima nemokumo rizika procentais remiantis kredito reitingu.
5.	Metinės palūkanos	Met_Paluk	Kredito laikotarpiu mokamų metinių palūkanų dydis kredito davėjui.
6.	Atlikti mokėjimai	Atlikti_mok	Kredito gavėjo atliktų mėnesinių įmokų skaičius.
7.	Vėlavimas	Velavimas	Ar klientas šiuo metu vėluoja atlikti mokėjimą ar ne. 0 – Moka laiku. 1 – Vėluoja.
8.	Ikiteisminis tyrimas	Teismas	Ar dėl kliento vėlavimo yra pradėtas iki teisminis tyrimas. 0 – Nėra. 1 – Yra.
9.	Sukčiavimas	Sukciavimas	Klientas laikomas sukčiaujančiu, jei atliko ne daugiau 1 kredito įmoką ir dėl jo vėlavimo yra pradėtas iki teisminis tyrimas. 0 – Nesukčiaujantis. 1 – Sukčiaujantis.
10.	Amžius	Amzius	Kredito gavėjo amžius metais.
11.	Lytis	Lytis	Kredito gavėjo lytis 0 – Vyras. 1 – Moteris.
12.	Gyvenamoji vieta	Gyv_vieta	Kredito gavėjo gyvenamasis miestas/miestelis/kaimas.
13.	Tikslas*	Tikslas	Skolintų lėšų panaudojimo tikslas.
14.	Pajamų šaltinis*	Paj_Saltinis	Kredito gavėjo pajamų šaltinis.
15.	Sritis*	Darbo_Sritis	Kredito gavėjo darbo sritis.
16.	Pajamų trukmė	Paj_trukme	Kredito gavėjo tvarių pajamų trukmė metais.
17.	Stažas	Stazas	Darbo stažo trukmė metais.
18.	Išsilavinimas*	Isilavinimas	Kredito gavėjo išsilavinimas.
19.	Šeimyninė padėtis*	Seim_Padetis	Kredito gavėjo šeimyninė padėtis.
20.	Išlaikomi asmenys	Islaik_asmenys	Finansiškai išlaikomų asmenų skaičius.
21.	Vėlavimas kartais	Velavimas_Savy_k	Kiek kartų per paskutinius 6 mėn. Kredito gavėjas vėlavo grąžinti paskolą SAVY platformoje.
22.	Vėlavimas dienomis	Velavimas_Savy_d	Kiek dienų per paskutinius 6 mėn. Kredito gavėjas vėlavo grąžinti paskolą „SAVY“ platformoje.
23.	Turtas*	Turtas	Kredito gavėjo turimas turtas viešai deklaruotas prieinamose duomenų bazėse.
24.	Pajamos	Pajamos	Kredito gavėjo mėnesinės pajamos (eur.).
25.	Įsipareigojimai	Isipareigojimai	Kredito gavėjo mėnesiniai įsipareigojimai (eur.).
26.	Įsipareigojimų ir pajamų santykis	DSTI	Kredito gavėjo mėnesinių įsipareigojimų ir pajamų santykis.
27.	Registruota skola	Anks_skola	Ar klientas turi anksčiau registruotų skolų. 0 – Neturi. 1 – Turi.

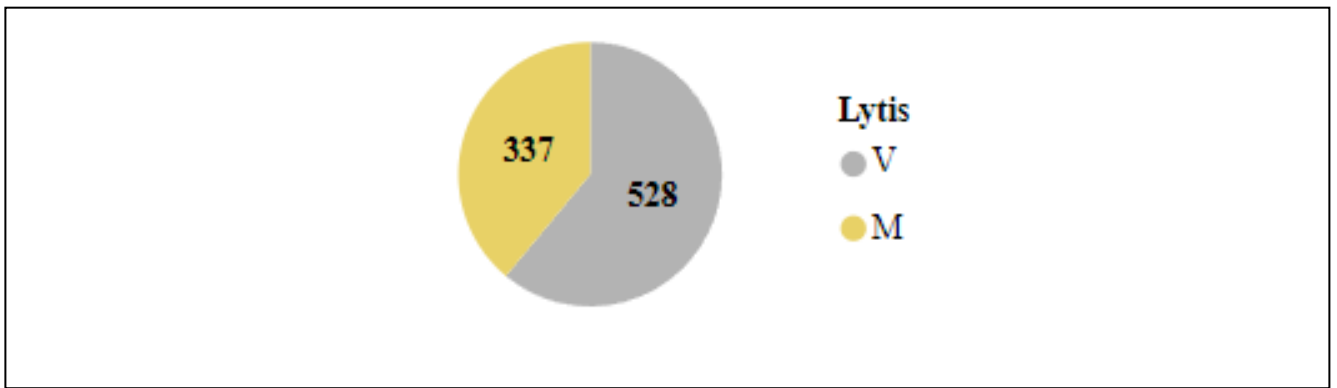
Kintamieji pažymėti žvaigždute (\*) buvo užkoduoti skaičiais, kurių paaiškinimai pateikiami priede nr.1. Kintamieji nr. 6, 7 ir 8 (atitinkamai: atlikti mokėjimai, vėlavimas, ikiteisminis tyrimas) pašalinami iš duomenų rinkinio, nes jie beveik tapatūs (nr. 6) ir visiškai tapatūs (nr. 7, nr. 8) su priklausomu kintamuoju – „Sukčiavimas“ (nr. 9).

4 paveiksle pateikiamas turimo duomenų rinkinio kokybinių kintamųjų pasiskirstymas.



4 pav. Kokybinių kintamųjų pasiskirstymas duomenų rinkinyje.

Iš diagramų viršuje matome, kad UAB „Bendras finansavimas“ dominuojantys klientai yra neturintys viešai registruoto turto, neturintys finansiškai išlaikomų asmenų bei dirbantys pagal darbo sutartį asmenys. Tarp klientų taip pat vyrauja vyriškos lyties atstovai (žr. 5 pav.). Vidutinis klientų amžius – 37-38 metai, uždirbamų pajamų mediana – 1000 eur. ir vidutinis klientų įsipareigojimų ir pajamų (angl. *debt service-to-income ratio* – DSTI) santykis duomenų imtyje yra 31.53 proc. Iš visų analizuojamų klientų 61.38 proc. asmenų kaip paskolos tikslą nurodė refinansavimo paslauga, o 22.42 proc. paskolų prašymų buvo skirti vartojimo išlaidoms.

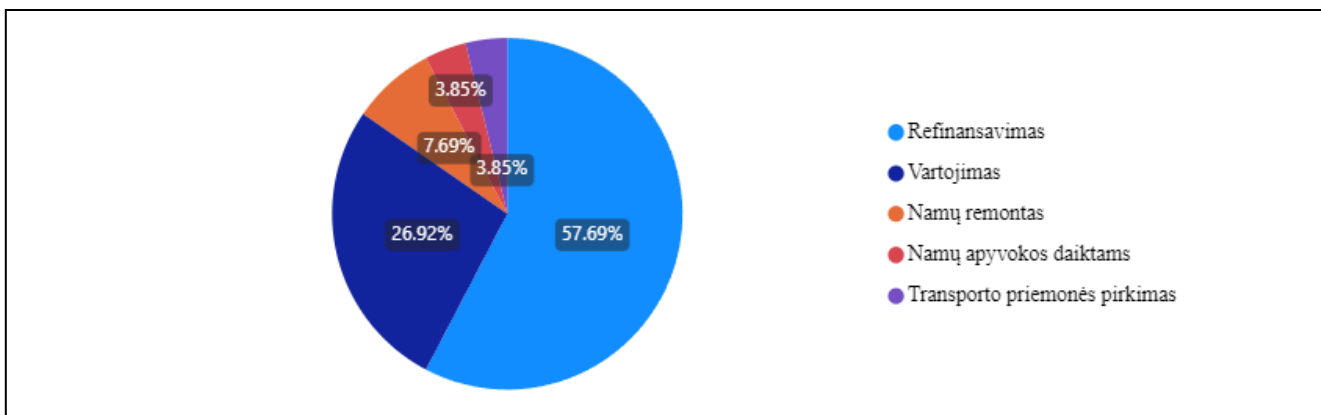


**5 pav.** Kintamųjų pasiskirstymas duomenų rinkinyje pagal lytį.

Siekiant susidaryti tarpusavio skolinimosi platformose besiskolinančio asmens vaizdą, galima klientą apibrėžti kaip vidutines pajamas uždirbantį asmenį, dažnu atveju jau turintį skolinių įsipareigojimų kitose įmonėse, neturintį viešai registruoto turto bei siekiantį refinansuoti turimus įsiskolinimus arba pasiskolinti papildomų lėšų vartojimo reikmėms.

Tyrimo priklausomas kintamasis yra - sukčiavimas. Tiriamojo darbo dalyje sukčiaujančiais laikomi kredito gavėjai atlikę ne daugiau nei vieną kredito įmoką ir perduoti iki teismo tyrimo procesui dėl nemokumo. T.y. asmenys pasisavinę lėšas be intencijos jas gražinti. Kredito rizikos samprata glaudžiai susijusi su sukčiavimu, kadangi įvertinus sukčiavimo tikimybę, galime nustatyti ir galimą kredito riziką. Dėl to siekiant sumažinti tenkančią kredito riziką ir ateityje išvengti finansinių nuostolių, yra svarbu laiku identifikuoti galimus sukčiavimo atvejus ir jiems užkirsti kelią dar ankstyvoje galimo sukčiavimo stadijoje.

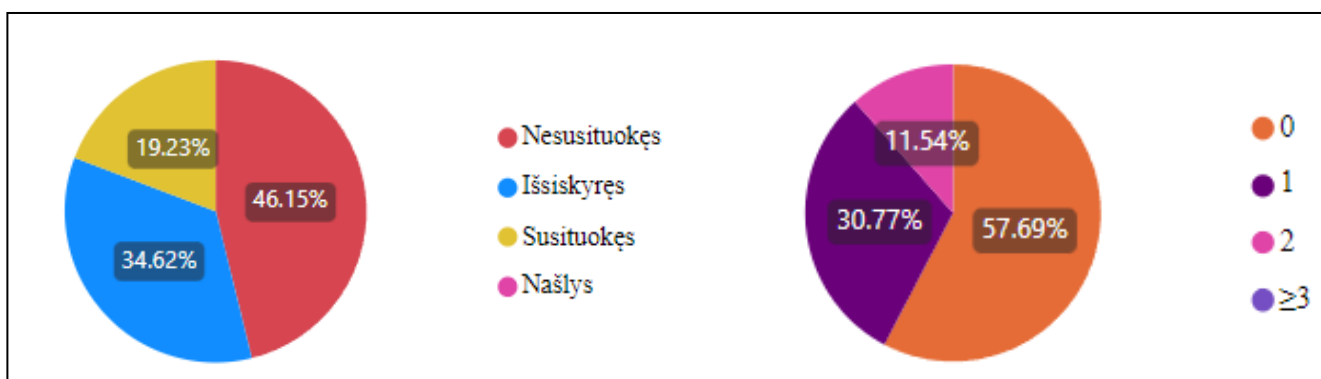
Tarp tyrimo duomenų buvo išskirti 26 sukčiavimo atvejai. Pagal kokybinius kintamuosius analizuojant sukčiavimo atvejus, pastebima, kad daugiau nei pusė, 57.69 proc. visų sukčiavimo atvejų kaip kredito tikslą įvardino – Esamų įsipareigojimų refinansavimą (žr. 6 pav.). Refinansavimas turi dvi pagrindines priežastis: pirmoji – skolininkai siekia padengti esamus įsipareigojimus palankesnėmis sąlygomis ir susimąžinti už kreditą mokamų palūkanų dalį, įmokų dydžius ar pakoreguoti įsiskolinimo terminą, antroji – artėjant vieno įsiskolinimo mokėjimo dienai, neturint laisvų lėšų, prisiimami nauji finansiniai įsipareigojimai, kuriais padengiami seni. Antroji refinansavimo priežastis gali būti inicijuojama sąmoningai, jau įsisukus į įsipareigojimų ratą tikintis naują įsipareigojimą vėl refinansuoti kitu. Vis dėlto, Lietuvos kreditų rinka yra griežtai reguliuojama, kreditoriams įvertinus, kad klientas nemokus, arba klientui jau pasiekus 40 procentų įsipareigojimų ir pajamų santykio ribą, naujas kreditas skolininkui nebebus suteikiamas, kas gali lemti nesugebančio planuoti finansinių išteklių kliento nemokumą.



**6 pav.** „SAVY” sukčiaujančių klientų pasiskirstymas pagal skolintų pinigų panaudojimo tikslą

Antroji pagal dydį sukčiavimo atvejais pasisavintų pinigų paskirtis – Vartojimo paskola (26.92 proc.). Vartojimo paskolos naudojamos kasdieniams poreikiams tenkinti arba smulkiems pirkiniams įsigyti. Trečia pagal dydį sukčiavimo grupė teigė naudosianti lėšas namų remontui (7.69 proc.). Ir po vieną sukčiavimo atvejį aptikta paskolose kurios skolintas lėšas ketino panaudoti namų apyvokos daiktų bei transporto priemonei pirkti. Tarp sukčiavimo atvejų nebuvo paskolų, su tokiais lėšų panaudojimo tikslais kaip: Medicininėms išlaidoms, nekilnojamajam turtui pirkti, nekilnojamajam turtui vystyti, studijoms ar verslui finansuoti. Taigi pagal paskolos tikslą didžiausią riziką kelia paskolos, kuriomis siekiama kompensuoti trūkstamą lėšų kiekį kasdieniame gyvenime (turimų įsipareigojimų refinansavimas bei vartojimo paskolos).

7-ame paveiksle pateikiamas sukčiavimo atvejų pasiskirstymas pagal šeimyninę padėtį (kairėje) bei finansiškai išlaikomų asmenų skaičių (dešinėje).



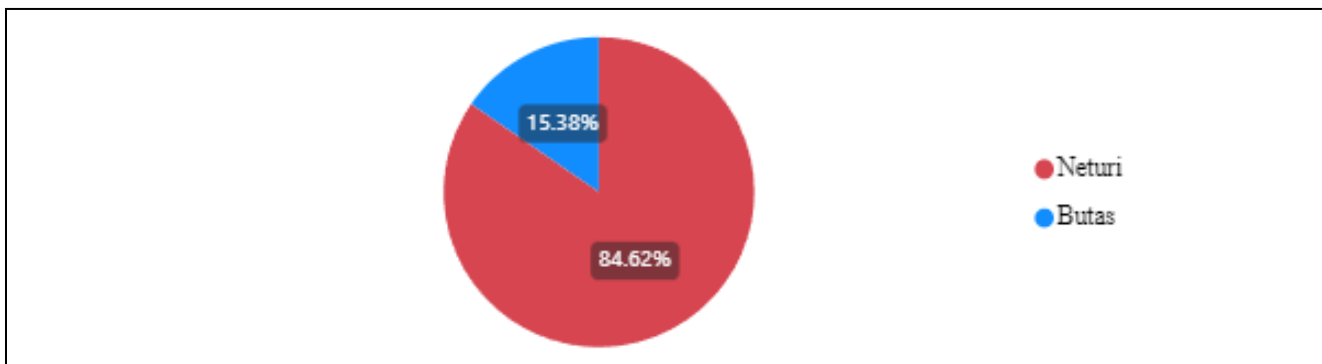
**7 pav.** „SAVY” klientų skolintų pasiskirstymas pagal šeimyninę padėtį (kairėje) ir finansiškai išlaikomų asmenų skaičių (dešinėje)

Taigi tarp sukčiavimo atvejų dominuoja nesusituokę asmenys (46.15 proc.), šiek tiek mažesnę dalį sudaro išsiskyrusieji (34.62 proc.), dar mažesnę dalį tarp sukčiaujančiųjų užima susituokę klientai (19.23 proc.). Šiuo atveju netekusių sutuoktinio asmenų tarp sukčiaujančiųjų nebuvo. Apibendrinant rezultatus, galima

teigti, kad vieniši asmenys, neturintys sutuoktinio, sudaro 80.77 proc. visų sukčiavimo atvejų. Susituokę asmenys būna dviese atsakingi už bendrus finansus, taigi vienam susidūrus su pajamų sumažėjimu ar kitu finansiniu pasikeitimu, kito asmens pajamos dažniausiai išlieka stabilios, kas gali būti laikoma su finansiniais sunkumais susidūrusio asmens „finansine pagalve“, dėl ko poreikis sukčiauti lieka nuošalyje.

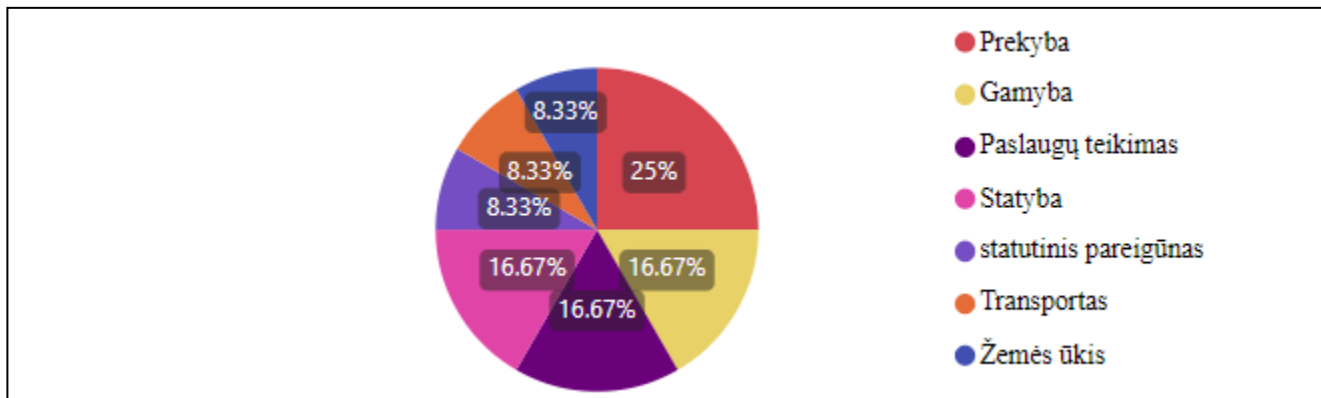
Iš septintojo paveikslo dešinėje pusėje esančios diagramos matome, kad taip pat tarp sukčiaujančiųjų asmenų didelę dalį sudaro asmenys, kurie neturi finansiškai išlaikomų asmenų (57.69 proc.). Tai galima sieti su šeimyninės padėties pasiskirstymu, kadangi nesutuokę asmenys dažniausiai finansiškai išlaikomų asmenų neturi. 30.77 proc. tarp sukčiaujančiųjų sudaro asmenys turintys tik vieną finansiškai išlaikomą asmenį, o turintys du, tik – 11.54 proc. Daugiau nei 2 asmenis išlaikančiųjų tarp sukčiaujančiųjų – nebuvo.

Aštuntame paveiksle pateikiama sukčiavusiųjų asmenų turimo turto pasiskirstymo diagrama. Matome, kad didžioji dalis (84.62 proc.) sukčiavusiųjų neturėjo viešai registruoto turto. Galima daryti prielaidą, kad „SAVY“ sukčiaujantys asmenys iš tiesų gali susidurti su finansinėmis problemomis, kas lemia polinkį į sukčiavimą.



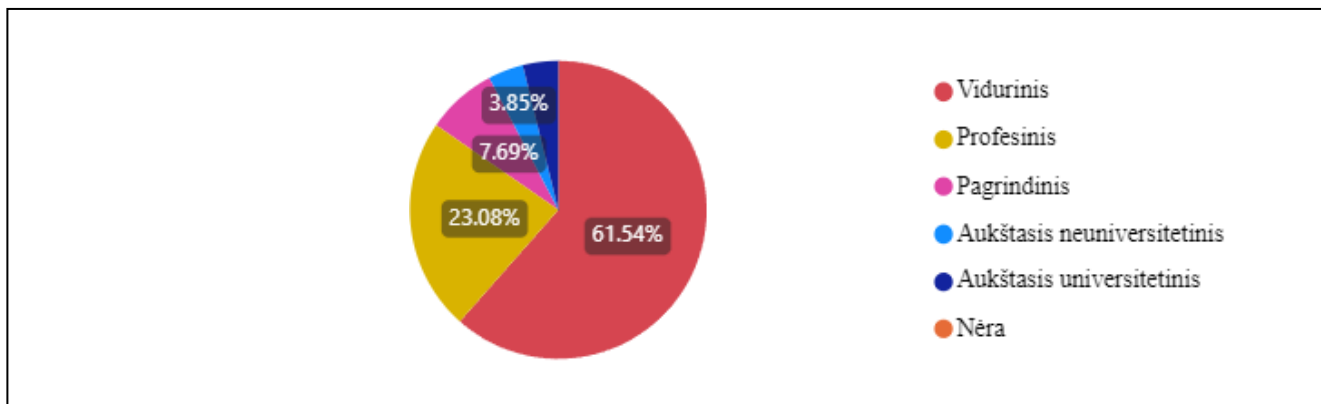
**8 pav.** „SAVY“ sukčiaujančių klientų pasiskirstymas pagal viešai deklaruotą turtą.

Devintame paveiksle pateikiamas tyrimo objekto - sukčiaujančių klientų, pasiskirstymas pagal darbinę sritį. Pagal šį kintamąjį neišsiskiria nei viena darbinė sritis, kaip turinti daugiausiai į sukčiavimą „SAVY“ platformoje linkusių darbuotojų.



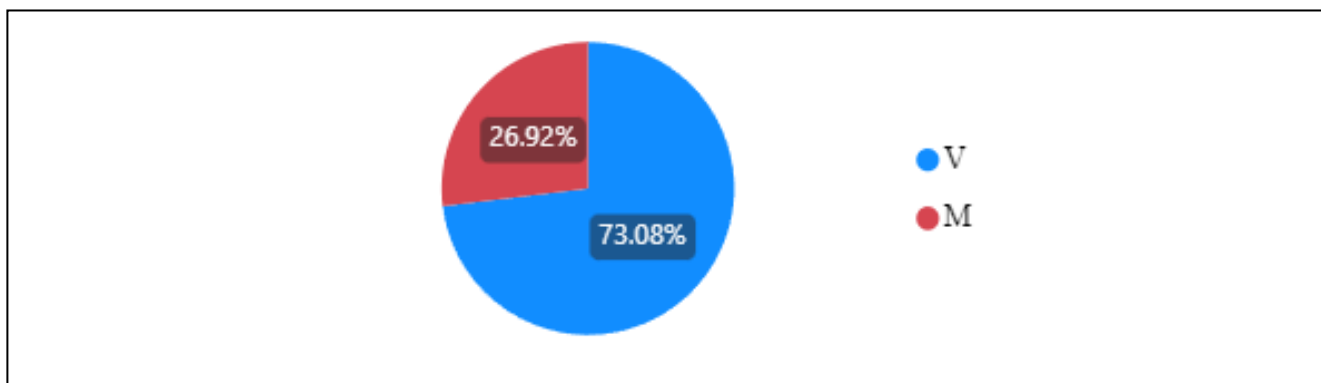
**9 pav.** „SAVY“ sukčiaujančių klientų pasiskirstymas pagal darbinę sritį.

Dešimtas paveikslas leidžia daryti prielaidą ar kliento išsilavinimas gali lemti polinkį į sukčiavimą. „SAVY“ paskolų duomenimis matome, kad daugiau nei puse sukčiavimo atvejų (61.54 proc.) priklauso klientams įgijusiems tik vidurinį išsilavinimą. Visoje duomenų imtyje šio išsilavinimo skolininkų grupė sudarė 32.94 proc. iš visos duomenų imties. Antroji pagal dydį tarp sukčiaujančiųjų dalis – klientai įgiję profesinį išsilavinimą (23.08 proc.). 7.69 proc. sukčiavimo atvejų sudaro klientai turintys pagrindinį išsilavinimą, ir po 3.85 proc. sudaro klientai įgiję aukštąjį neuniversitetinį bei aukštąjį universitetinį išsilavinimą. Taigi apžvelgiant duomenų pasiskirstymą, galime daryti prielaidą, kad kuo aukštesnis kliento išsilavinimas, tuo mažiau jis linkęs į sukčiavimą.



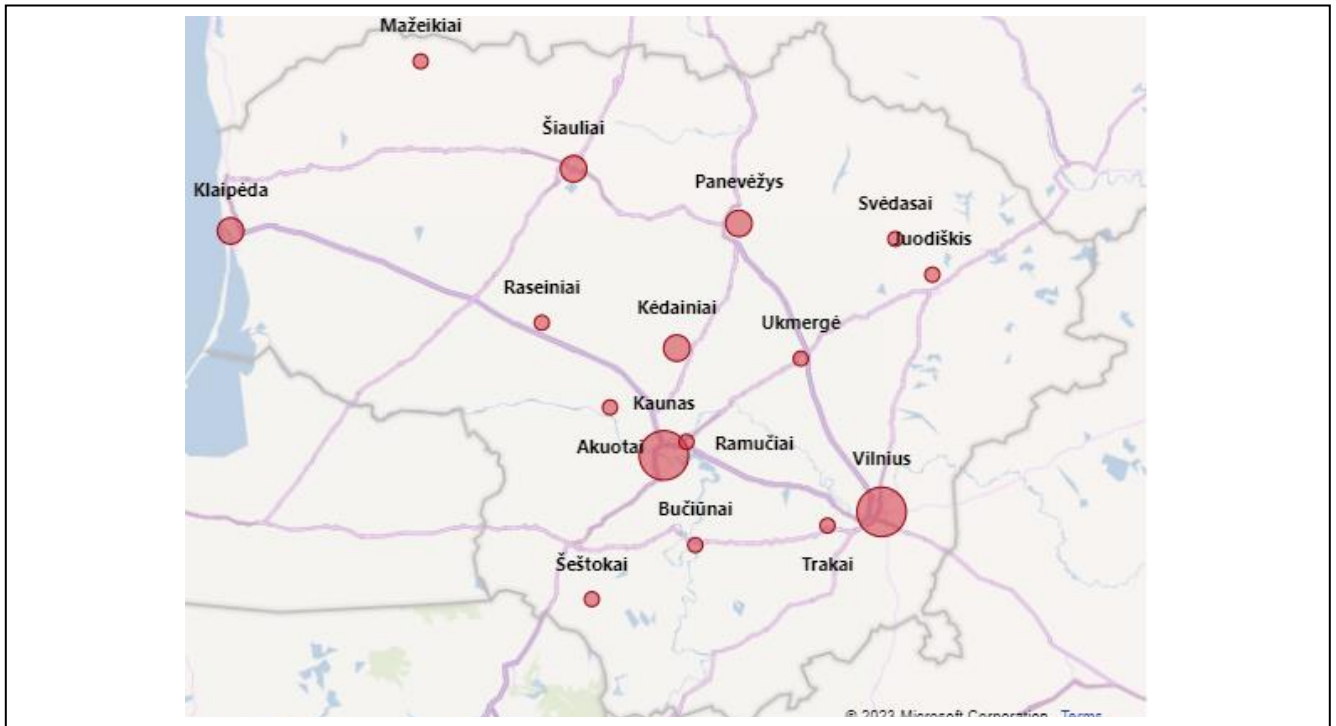
**10 pav.** „SAVY“ sukčiaujančių klientų pasiskirstymas pagal įgytą išsilavinimą.

Pagal lytį sukčiaujančiųjų tarpe didesnė dalis vyrų. Vyrų sudaro 73.08 proc., visų sukčiavimo atvejų, tuo tarpu moterys tik 26.92 proc. (žr. 11 pav.). Vis dėl to atsižvelgiant į tai, kad vyrai sudarė 61.04 proc. visos duomenų imties, negalime formuluoti prielaidos, kad labiau linkę į sukčiavimą.



**11 pav.** „SAVY“ sukčiaujančių klientų pasiskirstymas pagal lytį.

Dvyliktame paveiksle atvaizduotas „SAVY“ sukčiavimo atvejų geografinis pasiskirstymas. Iš vietovių išsiskiria Vilnius bei Kaunas, šiuose miestuose gyvena po 4 klientus, kurie buvo pripažinti kaip sukčiaujantys „SAVY“ platformoje. Taip pat po du sukčiavimo atvejus buvo: Kėdainiuose, Klaipėdoje bei Panevėžyje. Kituose miestuose atvejai pavieniai. Kadangi sukčiavimas pastebėtas didžiuosiuose Lietuvos miestuose, kur gyventojų populiacija didžiausia, daroma prielaida, kad gyvenamoji vieta nelemia klientų polinkio į sukčiavimą.



12 pav. Geografinis „SAVY“ sukčiaujančių klientų pasiskirstymas.

Taigi apžvelgiant išnagrinėtus sukčiavimo atvejų duomenis, galime formuluoti prielaidą – didesnę riziką būti sukčiaujančiais turi asmenys siekiantys gauti refinansavimo paskolą, sudarantys vieno asmens namų ūkį, neturintys viešai registruoto turto bei įgiję tik vidurinę išsilavinimą.

## 2.2. Tyrimo metodologija

**Tyrimo problema.** Auganti tarpusavio skolinimosi platformų rinka neišvengiamai susiduria su didesne sukčiavimo rizika, kuri daro tiek tiesioginį poveikį P2P rinkos stabilumui tiek netiesioginę įtaką visai šalies ekonomikai. Sukčiavimo atvejus identifikuoti sudėtinga, kadangi atsiranda vis naujų sukčiavimo formų. Siekiant atpažinti šiandienius sukčiavimo atvejus darbu sudaromas naujas modelis galimo sukčiavimo identifikavimui.

Atsižvelgiant į keliamą problemą, **tyrimo tikslas** – sudaryti sukčiavimo rizikos vertinimo modelį remiantis tarpusavio skolinimosi platformų duomenimis.

**Tyrimo objektas** – sukčiavimo rizikos vertinimo modelis.

**Tyrimo metodai:**

1. duomenų surinkimas ir apdorojimas;
2. pagrindinių komponentų analizė;
3. atsitiktinių miškų, atraminių vektorių mašinų bei neuronų tinklų tyrimo modelių sudarymas.
4. rezultatų palyginimas.

Kaip ir minėta, siekiant atlikti analizę, pirmiausia reikia išskirti svarbiausius faktorius. Duomenims tvarkyti pritaikyta faktorinė analizė, pagrindinių komponentų būdu (angl. *Principal Components Analysis - PCA*), kaip ir Zhang, Zerui, ir Shaoli Jin. (2020) tyrime “ Rizikos vertinimas ir matavimas P2P rinkoje“ (angl. *Risk Assessment and Measurement of P2P Industry*).

**Pagrindinių komponentų analizė** išskiria ir apibrėžia faktorius į kuriuos apjungiami turimų kintamųjų svarbi informacija atsižvelgiant į duomenų panašumus. Ši analizė tinkama tik kiekybiniam duomenims, todėl atlikta tik jiems. Pagrindinių komponentų analizės tikslas suskirstyti stebimus kintamuosius į grupes, kurios turi vienijantį faktorių, taip siekiant sumažinti modelyje esančių kintamųjų skaičių bei multikolinearumo problemą (Čekanavičius ir Murauskas, 2002).

Siekiant pritaikyti pagrindinių komponentų analizę, būtina patikrinti ar duomenys tinkami PCA analizei. Dėl to darbo metu apskaičiuoti Cronbach'o alfa, Bartlett'o sferiškumo, Kaizer'io-Mejer'io-Olkin'o kriterijai bei koeficientai ir koreliacinės matricos determinantas. Kriterijų reikšmės ir vertinimas pateikiami 6 lentelėje.

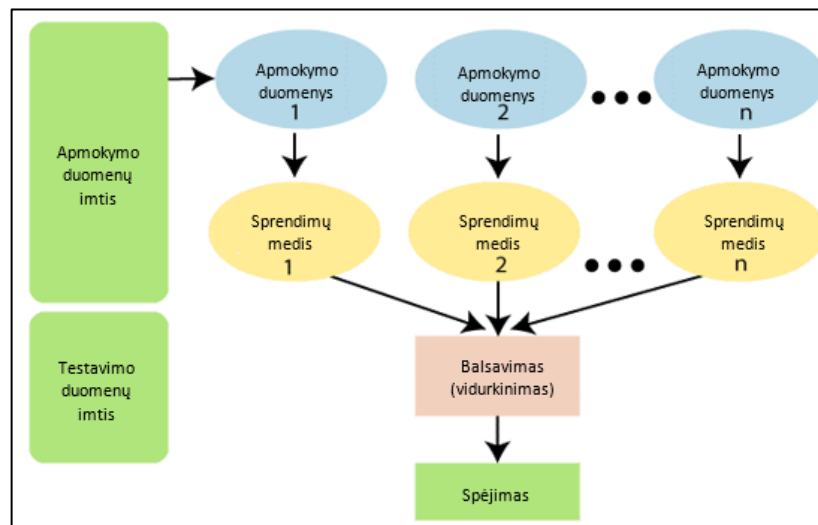
**6 lentelė.** Pagrindinių komponentų analizės vertinimo rodikliai (sudaryta remiantis: Pūkėnas, 2009; Stankūnaitė, 2009; Shrestha, 2021).

Vertinimo rodikliai	Koeficiento reikšmė	Pageidaujama norma	Šaltinis
Cronbach'o alfa	Įvertina ar kintamieji pakankamai apibrėžia nagrinėjamą veiksni	> 0,7	Pūkėnas (2009)
Bartlett'o sferiškumo kriterijus	Tikrina hipotezę, kad visi stebimi kintamieji yra nekoreliuoti.	p-reikšmė < $\alpha$	Pūkėnas (2009)
Kaizer'io-Mejer'io-Olkin'o kriterijus	KMO mato reikšmė rodo ar kintamųjų porų koreliacija yra paaiškinama kitais kintamaisiais	KMO >0.9 – Puiku 0.8<KMO<0.9 – Gerai 0.7<KMO<0.8 – Patenkinamai 0.6<KMO<0.7 – Vidutiniškai 0.5<KMO<0.6 - Blogai KMO<0,5 – Faktorinė analizė nepriimtina	Stankūnaitė (2009)
Koreliacinės matricos determinantas	Įvertina ar duomenyje nėra multikolinearumo	>0,0001	Shrestha (2021)



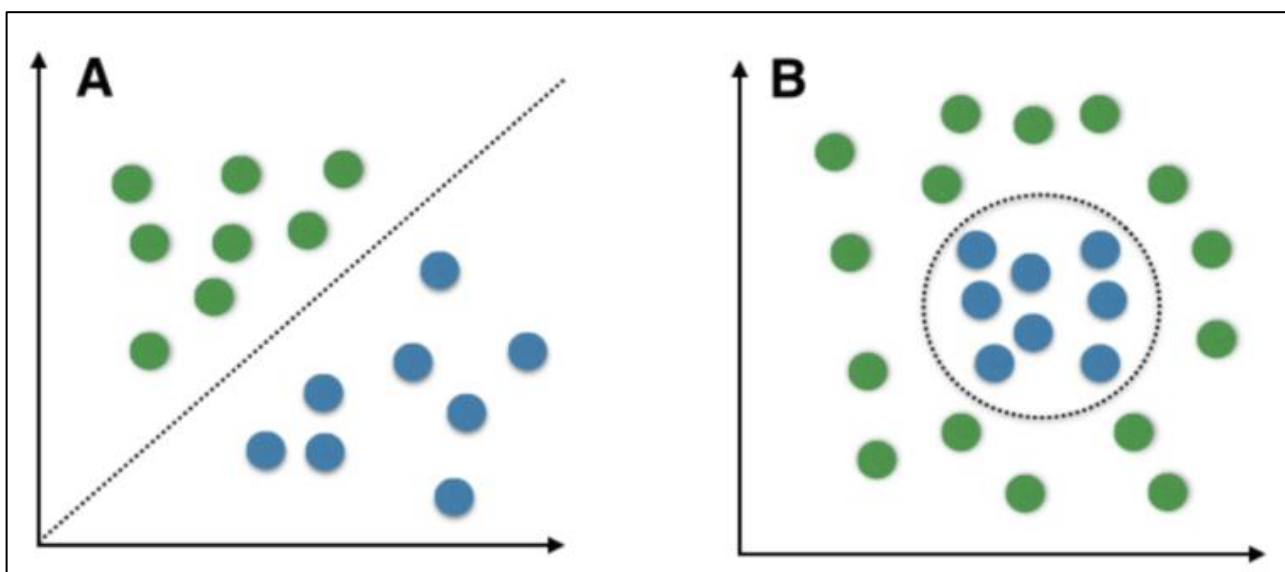
Išskyrus pagrindines komponentes sudaryomi atsitiktinių miškų, atraminių vektorių mašinų bei neuroninių tinklų modeliai.

**Atsitiktiniai miškai** – algoritmas sudarytas iš grupės sprendimų medžių naudojamas duomenų klasifikacijai. Algoritmo vykdymo metu pradinė duomenų imtis padalijama į mokymo bei testavimo dalis. Iš mokymui skirtų duomenų naudojant atsitiktinę imtį sukuriama atskiri atsitiktinio miško medžiai. „Testuojamas mėginys klasifikuojamas su visais gautais medžiais. Kiekvienas medis jį priskiria tam tikrai klasei, (Meržvinskaitė, 2016). Atsitiktinių miškų algoritmas remiasi „daugumos balsavimu“ (angl. *Majority Voting*), todėl po testuojamojo kintamojo klasifikavimo, apskaičiuojami gautų įverčių dažniai ir kintamasis priskiriamas tai klasei, kurios pasikartojimas buvo didžiausias (žr. 13 pav.)



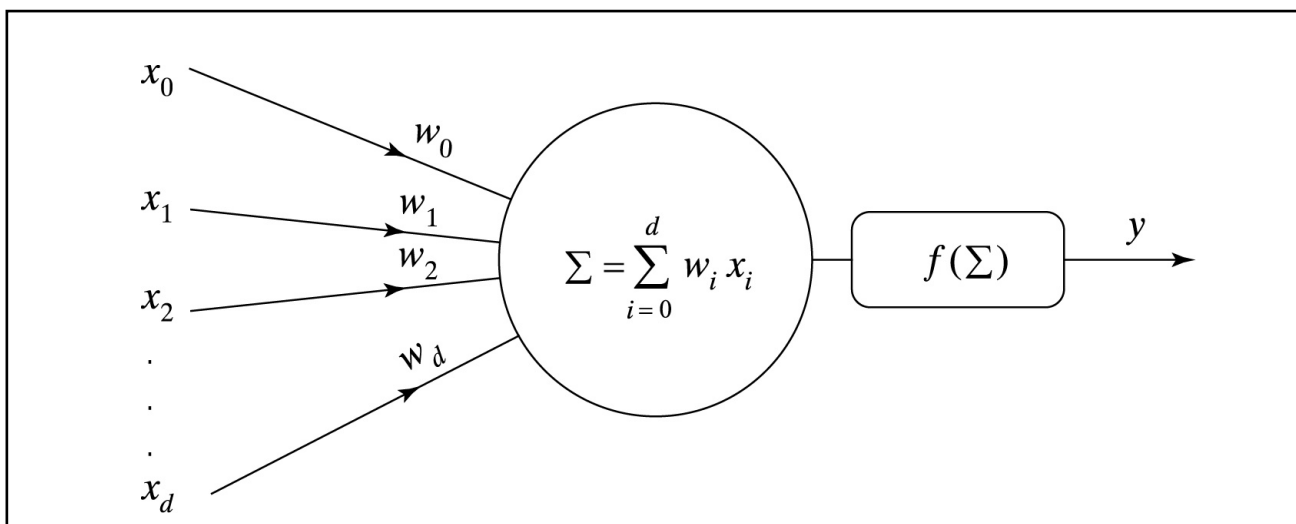
13 pav. Atsitiktinių miškų algoritmo schema. Sudaryta remiantis (Dmitrievsky, 2018)

**Atraminių vektorių mašinos** – metodas, kuris klasifikuoja požymius juos atskiriant hiperplokštuma (žr. 14 pav. A.). Jei nėra plokštumos, kuria galima atskirti požymius, taškai yra projektuojami į aukštesnės dimensijos erdvę ir naudojamas SVM su branduolio (angl. *Kernel*) funkcija (žr. 14 pav. B.), kuri naudojama tiesiškai neatskiriamiems duomenims (Buivyte, 2016).



14 pav. Atraminų vektorių mašinų veikimo vizualizavimas (Schultebrucks, 2017).

**Neuroniniai tinklai** – „informacijos apdorojimo struktūra, imituojanti kai kuriuos gyvųjų organizmų smegenyse vykstančius informacijos perdavimo procesus“ (Verikas ir Gelžinis, 2008). Dirbtinių neuroninių tinklų metodas taikomas tiek klasifikavime, tiek prognozavime ar optimizavime. Neuroninį tinklą apibūdina įvesties signalai ( $x_n$ ), neuronų svoriai ( $w_n$ ), perdavimo funkcija ir neuronų išėjimo signalas ( $y$ ) (žr. 15 pav.). Kiekvienam įvesties signalui priskiriamas svoris (perdavimo koeficientas), apskaičiavus neuronų perdavimo funkciją suformuojamas neuronų išėjimo signalas. (Bernotas, 2009).



15 pav. Dirbtinio neuroninio tinklo schema (Verikas ir Gelžinis, 2008).

Sudarytųjų modelių tikslumas palyginamas remiantis vertinimo kriterijais pateiktais 7 lentelėje.

**7 lentelė.** Modelių vertinimo kriterijai. Sudaryta remiantis: Pukėnas (2009) ir Varghese (2018).

Vertinimo kriterijus	Skaičiavimas	Ką parodo?
ROC (angl. <i>Receiver operating characteristic</i> )	Kreivė priklausanti nuo modelio specifiškumo ir jautrumo.	Ar regresijos modelis yra geras, kai jo sprendimo priėmimo slenkstis kinta.
AUC (angl. <i>Area under the curve</i> )	Plotas po ROC kreive	Modelio gebėjimą teisingai klasifikuoti dvi klases.
Teisingai Teigiama (angl. <i>True Positive - TP</i> )	Modelio rezultatas	Kiek teigiamų reikšmių įvertinta teigiamai.
Klaidingai Neigiama (angl. <i>False Negative - FN</i> )	Modelio rezultatas	Kiek teigiamų reikšmių įvertinta neigiamai.
Klaidingai Teigiama (angl. <i>False Positive - FP</i> )	Modelio rezultatas	Kiek neigiamų reikšmių įvertinta teigiamai.
Teisingai Neigiama (angl. <i>True Negative - TN</i> )	Modelio rezultatas	Kiek neigiamų reikšmių įvertinta neigiamai.
Tikslumas (angl. <i>Accuracy - ACC</i> )	$\text{Tikslumas} = \frac{TT + TN}{TT + KN + KT + TN}$	Bendras modelio tikslumo vertinimo matas.
Preciziškumas ( <i>precision</i> )	$p = \frac{TT}{TT + KT}$	Santykį tarp TT ir visų teigiamai prognozuotų reikšmių.
Išsamumas ( <i>recall</i> )	$r = \frac{TT}{TT + KN}$	Modelio gebėjimą teisingai atskirti realias objekto būsenas.
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)	$F1 - \text{įvertis} = \frac{2TT}{2TT + KT + KN}$	Modelio tikslumo vertinimo matas teigiamai klasei. Reikšmė < 0.5 – netenkinama.
Lygių paklaidų dydis (angl. <i>Equal Error Rate - EER</i> )	Taškas kuriame klaidingo atmetimo ir klaidingo priėmimo rodikliai lygūs.	Modelio klaidų lygį. Kuo mažesnė EER reikšmė, tuo didesnis klasifikatoriaus tikslumas.

Paprastai sukčiavimo aptikimas yra labai sudėtingas dėl klasių disbalanso. Duomenų rinkinyje sukčiaujančių asmenų su nesukčiaujančiais santykis yra 1:32. Todėl modelio sudaryme yra labai svarbu sukurti modelį, kuris tiksliai įvertintų mažumos klasę.

**Stratifikuotos k-dalių kryžminės patikros metodas** (angl. *Stratified k-fold cross validation*)  
 Naudojant stratifikuotą k-dalių kryžminę patikrą duomenys padalijami į k lygių dalių, taip, kad kiekvienoje dalyje priklausomojo kintamojo santykis išlaikomas tapatus pradinėje duomenų imtyje esančiam klasių santykiui. Taikant šį metodą viena k dalis yra paliekama testavimui, o visos kitos (k-1) naudojamos mokymui, vėliau procesas pakartojamas testavimui pasirinkant kitą k dalį ir iteracijos vyksta

tol, kol ištestuojamos visos  $k$  dalys. Tai leidžia pritaikyti modelį nesubalansuotų duomenų rinkiniui ir gauti aukštus tikslumo matus (Abreu et al., 2018). Atsižvelgiant į tai modeliai sudaryti naudojant stratifikuotą  $k$ -dalių kryžminę patikrą.

**Atsitiktinis imties retinimas (angl. *Random under-sampling*)** – tai procesas, „kurio metu pašalinami stebejimai iš didesniosios klasės (angl. *majority class*) atsitiktiniu būdu arba taikant tam tikras technikas.“ (Juodzevičiūtė, 2020). Siekiant rasti geriausią sukčiavimo rizikos vertinimo modelio sudarymą nesubalansuotiems duomenims, duomenų rinkiniui paruošti panaudojama atsitiktinio imties retinimo technika, kuri sumažina didesniosios klasės įrašus tiek, kad būtų pasiektas imties balansas.

Tyrimas atliktas R-studio programa, versija - 4.2.3. Duomenims aprašyti vizualizacijos sugeneruotos Power BI Desktop įrankiu. Tyrimo reikšmingumo lygmuo ( $p$ -reikšmė) – 0.05.

### 3. Tyrimų rezultatai ir jų aptarimas

Šiame skyriuje aprašyti bei išanalizuoti tyrimo metu gauti rezultatai. Visų pirma vadovaujantis tyrimo metodologija tyrimo duomenims buvo atliktas Kaizer'io-Mejer'io-Olkin'o, Bartlett'o sferiškumo vertinimas, apskaičiuota Cronbach'o alfa reikšmė siekiant įvertinti ar tyrimo duomenims gali būti atlikta faktorinė analizė pagrindinių komponentių metodu. Sukurti atsitiktinių miškų, atraminių vektorių mašinų bei neuroninių tinklų modeliai su išskirtomis pagrindinėmis komponentėmis ir be jų. Modeliai palyginti ir išrinktas geriausias modelis tyrimo duomenims klasifikuoti. Gauto geriausio modelio tikslumo įverčiai palyginti su kitų autorių sudarytų sukčiavimo vertinimo modelių įverčiais.

#### 3.1. Pagrindinių komponentių analizė

Gauti duomenų tinkamumo faktoriams išskirti pagrindinių komponentių būdu testų rezultatai pateikiami 8-oje lentelėje.

**8 lentelė.** Duomenų rinkinio tikimo pagrindinių komponentių analizei rodikliai.

Kaizer'io-Mejeri'o-Olkin'o kriterijus		0.66
Bartlett'o sferiškumo kriterijus	$\chi^2$	7017.989
	L.l.s.	300
	<i>p</i> -reikšmė	0,000

Kaizer'io-Mejer'io-Olkin'o kriterijaus reikšmė – 0.66. Literatūros šaltiniuose kaip minėta metodologinėje dalyje, išskiriami keli matavimo režiai pagal KMO priimtinumą. Kuo kriterijaus reikšmė aukštesnė tuo labiau duomenų rinkinys pritaikomas faktorinei analizei. Tyrimo duomenys pagal Kaizerio-Mejerio-Olkin'o kriterijų pagrindinių komponentių analizei tinka vidutiniškai (patenka į 0.6 – 0.7 vertinimo režį). Taigi daroma išvada, kad pagal Kaizer'io-Mejer'io-Olkin'o kriterijų duomenų rinkinys tinkamas PCA, tačiau reikia griežčiau įvertinti kitus rodiklius.

Bartlett'o sferiškumo kriterijaus hipotezės:

$H_0$ : Visų imčių dispersijos lygios.

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2.$$

$H_a$ : Bent dvi dispersijos tarpusavyje nelygios.

$$H_a: \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_k^2.$$

Apskaičiuota Bartlett'o sferiškumo kriterijaus p-reikšmė mažesnė už tyrimo reikšmingumo lygmenį ( $p = 0.000 < 0.05$ ), kas leidžia atmesti tikrinamą hipotezę ir priimti alternatyviąją. Todėl daroma išvada, jog duomenyse yra koreliuotų kintamųjų, kurie gali būti apjungiami į faktorius.

Kitas kriterijus leidžiantis nustatyti, ar duomenys gali būti naudojami faktorinei analizei – Cronbach'o alfa koeficientas. Koeficientas parodo kintamųjų vidinį suderinamumą, o tai svarbu įvertinti, kai ketinama kintamuosius apjungti į bendrą faktorių. Atlikus patikimumo analizę R-Studio programa gaunama Cronbach alfa koeficiento reikšmė – 0.73. Įvertinus gautus patikimumo duomenis, jei būtų išmestas kuris nors iš kintamųjų, Cronbach alfa koeficientas statistiškai reikšmingai nepadidėtų, todėl pakeitimai duomenų faile neatliekami.

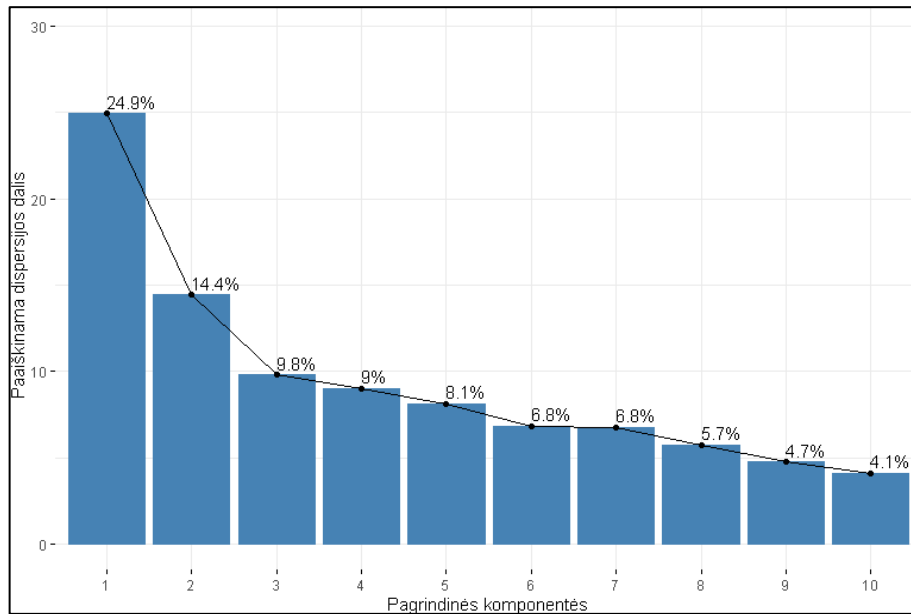
Paskutinis analizuojamas kriterijus – korealiacijų matricos determinantas. Determinanto reikšmė turi būti nuo 0.0001 iki 1, kitu atveju duomenyse kintamieji gali būti multikolinearūs, dėl ko reikia pašalinti iš imties stiprias koreliacijas turinčius kintamuosius. Tyrimo imties apskaičiuota koreliacijos determinanto reikšmė – 0.0002. Determinanto reikšmė nežymiai peržengia patenkinamą slenkstį ir daroma išvada, kad pagal koreliacijų matricos determinanto reikšmę tyrimo duomenų imtis yra tinkama faktorinei analizei atlikti.

Taigi remiantis Cronbach'o alfa, Bartlett'o sferiškumo, Kaizer'io-Mejer'io-Olkin'o kriterijais, bei korealiacijų matricos determinanto reikšme pagrindinių komponentių analizę galime taikyti tyrimo duomenims. Vis dėlto visų vertintų kriterijų reikšmės indikuoja, kad pagrindinių komponentių analizę gali būti taikoma, tačiau nėra puikiai tinkanti.

Atsižvelgiant į tai darbe sudaryta po tris atsitiktinių miškų, atraminių vektorių mašinų bei neuroninių tinklų modelius su trimis skirtingais duomenų rinkiniais:

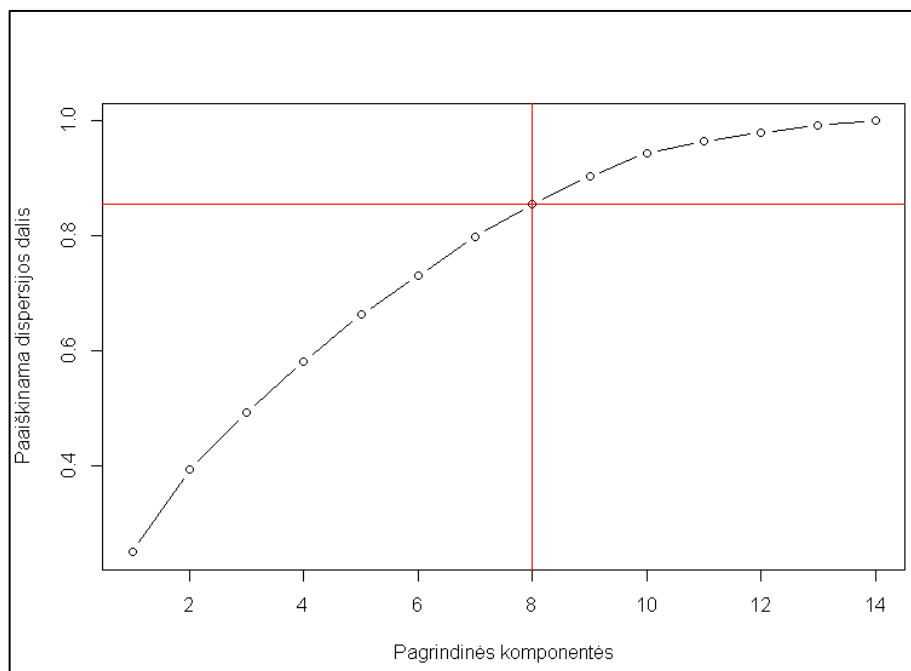
1. Pirminis tyrimo duomenų rinkinys su kiekybiniais ir kokybiniais kintamaisiais (I);
2. Pagrindinių komponentių analizės metu gautas duomenų rinkinys tik su komponentėmis (II);
3. Apjungtas pagrindinių komponentių ir tyrimo kokybinių kintamųjų duomenų rinkinys (III).

Pagrindinių komponentių analizės metu algoritmas išskiria tiek komponentių, kiek įtraukiama į modelį kintamųjų. Visos komponentės bendrai paaiškina 100 proc. priklausomojo kintamojo. R-Studio programoje nusibraižomas pagrindinių komponentių paaiškinamos informacijos kiekio grafikas (žr. 16 pav.). Pateiktame paveiksle vaizduojama kiek informacijos apie priklausomą kintamąjį paaiškina pirmosios 10 komponentių individualiai. Daugiausiai informacijos paaiškina pirmoji komponentė – 24.9 proc., antroji - 14.4 proc., trečioji – 9.8 proc.



**16 pav.** Pirmųjų 10 pagrindinių komponentių paaškinamos dispersijos dalis

„Komponentės, kurios yra už taško, kuriame staigus kreivės kritimas pereina į palaipsninį mažėjimą, yra laikomos nereikšmingomis” (Pūkėnas, 2009). Iš grafiko matoma, kad po pirmųjų 8 komponentių kreivės kritimo tendencija sumažėja ir pereina į palaipsninį mažėjimą. Iš kiekybinių kintamųjų išskirtos aštuonios pirmosios komponentės paaškina 85.58 proc. priklausomojo kintamojo sklaidos. Suminė komponentių paaškinamos informacijos dalis pateikiama 17 paveiksle.



**17 pav.** Suminė komponentių paaškinama priklausomojo kintamojo dispersijos dalis.

Taigi atsižvelgiant į komponentių grafikus, sudarytuose sukčiavimo rizikos vertinimo modeliuose naudotos pirmos aštuonios pagrindinės komponentės, kurios paaiškina 85.58 proc. informacijos.

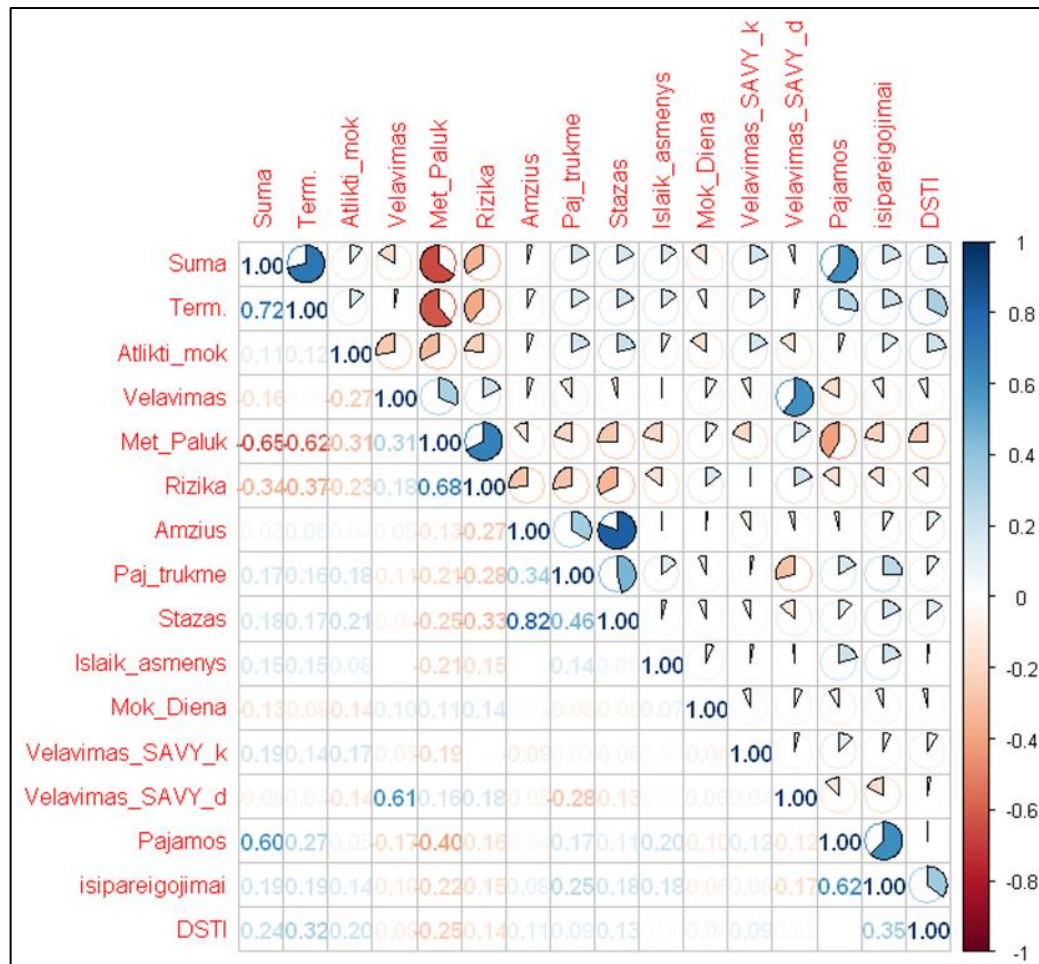
### 3.2. Mašininio mokymosi metodų taikymas

#### 3.2.1. Duomenų imties paruošimas

Prieš modelių sudarymą yra svarbu susitvarkyti duomenis. Tyrimo kokybiniai kintamieji R programoje paverčiami faktoriais, kad jie būtų tinkamai interpretuojami. Sudaryta kiekybinių kintamųjų koreliacijų matrica siekiant išanalizuoti ar duomenyse nėra multikolinearumo (glaudaus koreliacinio ryšio tarp nepriklausomų kintamųjų, darančių bendrą įtaką rezultatui (Jakeliūnaitė ir Skvarciany, 2019)) (žr. 18 pav.).

Iš koreliacijų matricos matome, kad stipriausią teigiamą koreliaciją turi:

1. Darbo stažas – amžius (0.82);
2. Kredito terminas – kredito suma (0.72).



18 pav. Kiekybinių tyrimo duomenų koreliacijų matrica.



Mokslinėje literatūroje, koreliacijos koeficientui esant didesniai nei 0.7 - 0.8, siūloma įtarti multikolinearumo problemą ir esant būtinybei pašalinti iš duomenų imties vieną kintamąjį iš koreliuojančios poros. Šio tyrimo atveju koreliuojantys veiksniai apjungiami į pagrindines komponentes, dėl to reikiamybės šalinti kintamųjų iš modelio – nėra.

Duomenų rinkinyje visi kintamųjų įrašai trūkstančių reikšmių neturi.

Sudarant duomenų vertinimo modelį yra svarbu gebėti patikrinti modelio tikslumą. Dėl galimybės patikrinti modelio tinkamumą tyrimo duomenims vertinti, tyrimo metu duomenų failui pritaikoma stratifikuota k-dalių kryžminė patikra. Šis metodas vienas iš dažniausiai taikomų klasifikavimo algoritmų vertinimo metodų dėl tinkamumo nesubalansuotiems duomenims. Tyrimo metu sudaryti modeliai su penkiais ( $k=5$ ) kryžminės patikros blokais, dėl geriausių parodytų rezultatų.

### 3.2.2. Atsitiktinių miškų metodas

Siekiant geriausio modelio sudarymo, atsitiktinių miškų metodu sudaryti trys modeliai:

1. Naudojant tyrimo duomenų rinkinį;
2. Naudojant aštuonias pagrindines komponentes;
3. Naudojant pagrindines komponentes ir kokybinius tyrimo kintamuosius.

Geriausi atsitiktinių miškų modeliai gauti su parinktais parametrais:

- Medžių skaičius – 10 (*ntree*);
- Atsitiktinių kintamųjų skaičius šakojant medį – 2 (*mtry*).

Sudarytų modelių rodiklių palyginimas pateikiamas 9 lentelėje

**9 lentelė.** Atsitiktinių miškų modelių vertinimo kriterijų palyginimas

Duomenų failas	Pirminiai tyrimo duomenys (I)	Pagrindinės komponentės (II)	Pagrindinės komponentės ir kokybiniai kintamieji (III)
Vertinimo kriterijus			
Tikslumas (angl. <i>Accuracy</i> – ACC)	<b>0.53</b>	0.47	0.38
Teisingai Teigiama (angl. <i>True Positive</i> - TP)	21	21	<b>24</b>
Klaidingai Neigiama (angl. <i>False Negative</i> – FN )	5	5	<b>2</b>
Klaidingai Teigiama (angl. <i>False Positive</i> - FP)	<b>398</b>	447	526
Teisingai Neigiama (angl. <i>True Negative</i> - TN)	<b>441</b>	392	313
Preciziškumas	<b>0.05</b>	0.04	0.04
Išsamumas	0.80	0.81	<b>0.92</b>
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)	<b>0.09</b>	<b>0.09</b>	0.08
Lygių paklaidų lygis (angl. <i>Equal error rate</i> – EER)	<b>29.47</b>	35.12	29.70

Pritaikius apmokytus modelius testavimo bloko duomenims, atsitiktinių miškų modelis geriausiai rezultatus parodė naudojant pirminį duomenų rinkinį su kokybiniais ir kiekybiniais kintamaisiais. Pastebėtina, kad kitiems dviems tyrimo duomenų rinkiniams pagal tikslumo matą atsitiktinių miškų metodas yra netinkamas. Tačiau, kadangi ACC įvertis apibrėžiamas kaip netinkamas modeliams su duomenyse esančiu disbalansu vertinti, būtina išanalizuoti ir kitus parametrus siekiant įvertinti modelio tinkamumą. Pirmojo modelio aukščiausias preciziškumas (0.05) parodo, kad 5 proc. iš visų teigiamai prognozuotų reikšmių suklasifikuota teisingai. Modelis 95 proc. visų spėjimų, kad klientas bus sukčiaujantis, suklasifikavo klaidingai. Išsamumo matas didžiausias sudarytame trečiajame modelyje – 0.92, vadinasi modelis 92 proc. sukčiaujančių klientų atpažino teisingai ir priskyrė juos rizikos grupei. Įvertis F-1 yra modelio tikslumo matas skaičiuojamas atsižvelgiant į preciziškumo ir išsamumo matus. Pagal šį įvertį išsiskiria pirmasis bei antrasis atsitiktinių miškų modelis, kurio tikslumo matas įvertinantis kiek kartų modelis atliko teisingą klasifikavimą yra – 0.09. Daroma išvada, kad modelis su pradiniu duomenų rinkiniu klasifikavimo uždavinį sprendžia tiksliausiai lyginant su modeliais, kuriuose naudoti antrasis ir trečiasis duomenų rinkiniai.

### 3.2.3. Atraminų vektorių mašinų metodas

Sudarytų atraminų vektorių mašinų modelių vertinimo kriterijų palyginimas pateikiamas 10 lentelėje. Atraminų vektorių mašinų radialinis sigma (*svmRadialSigma*) metodas atlieka modelio skaičiavimus su įvairiomis *Cost* bei *Sigma* reikšmėmis ir į modelį įtraukia tas reikšmes, su kuriomis gaunamas tiksliausias modelis.

Parinkti automatinio derinimo modelio parametrai paraštėms nustatyti:

Pirmo modelio:

- $\text{Sigma } (\sigma) = 0.0252;$
- $C (\text{Cost}) = 0.25.$

Antro modelio:

- $\text{Sigma } (\sigma) = 0.0394;$
- $C (\text{Cost}) = 0.25.$

Trečio modelio:

- $\text{Sigma } (\sigma) = 0.0101;$
- $C (\text{Cost}) = 2.$

**10 lentelė.** Atraminų vektorių mašinų modelių vertinimo kriterijų palyginimas.

Vertinimo kriterijus	Duomenų failas	Pirminiai tyrimo duomenys (I)	Pagrindinės komponentės (II)	Pagrindinės komponentės ir kokybiniai kintamieji (III)
Tikslumas (angl. <i>Accuracy</i> – ACC)		0.58	0.45	<b>0.59</b>
Teisingai Teigiama (angl. <i>True Positive</i> - TP)		17	<b>20</b>	19
Klaidingai Neigiama (angl. <i>False Negative</i> – FN )		9	<b>6</b>	7
Klaidingai Teigiama (angl. <i>False Positive</i> - FP)		352	473	<b>344</b>
Teisingai Neigiama (angl. <i>True Negative</i> - TN)		487	366	<b>495</b>
Preciziškumas		0.04	0.04	<b>0.05</b>
Išsamumas		0.65	<b>0.77</b>	0.73
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)		0.08	0.08	<b>0.10</b>
Lygių paklaidų lygis (angl. <i>Equal error rate</i> – EER)		35.82	38.14	<b>30.02</b>

Atraminų vektorių metodo tikslumas geriausias naudojant duomenų failą su pagrindinėmis komponentėmis ir kategoriniais kintamaisiais (ACC - 0.59). Šis modelis 4 proc. iš visų teigiamai prognozuotų reikšmių suklasifikavo teisingai. Atsižvelgiant į gebėjimą teisingai identifikuoti sukčiaujančius asmenis, matoma, kad geriausiai atvejus klasifikuoja SVM modelis su pagrindinių komponentių duomenų rinkiniu – 77 proc. visų sukčiavusiųjų buvo priskirti prie rizikos grupės. Naudojant pirmąjį duomenų rinkinį, teisingai suklasifikuota 65 proc. sukčiavimo atvejų, o trečiąjį - 73 proc. Naudojant atraminų vektorių metodą, visų atvejų klasifikavimo didesniu tikslumu pasižymėjo modelis su trečiuoju duomenų failu (F1- įvertis – 0.10).

### 3.2.4. Neuroniniai tinklai

Trečiasis pritaikytas modelis – neuroniniai tinklai. Sudarytų modelių vertinimo kriterijų palyginimas pateikiamas 11 lentelėje. Naudoti derinimo parametrai:

- Paslėptasis sluoksnis (*hiddenSize*) – 4;
- Maksimalus iteracijų skaičius (*maxit*) – 999.

**11 lentelė.** Neuroninių tinklų modelių vertinimo kriterijų palyginimas.

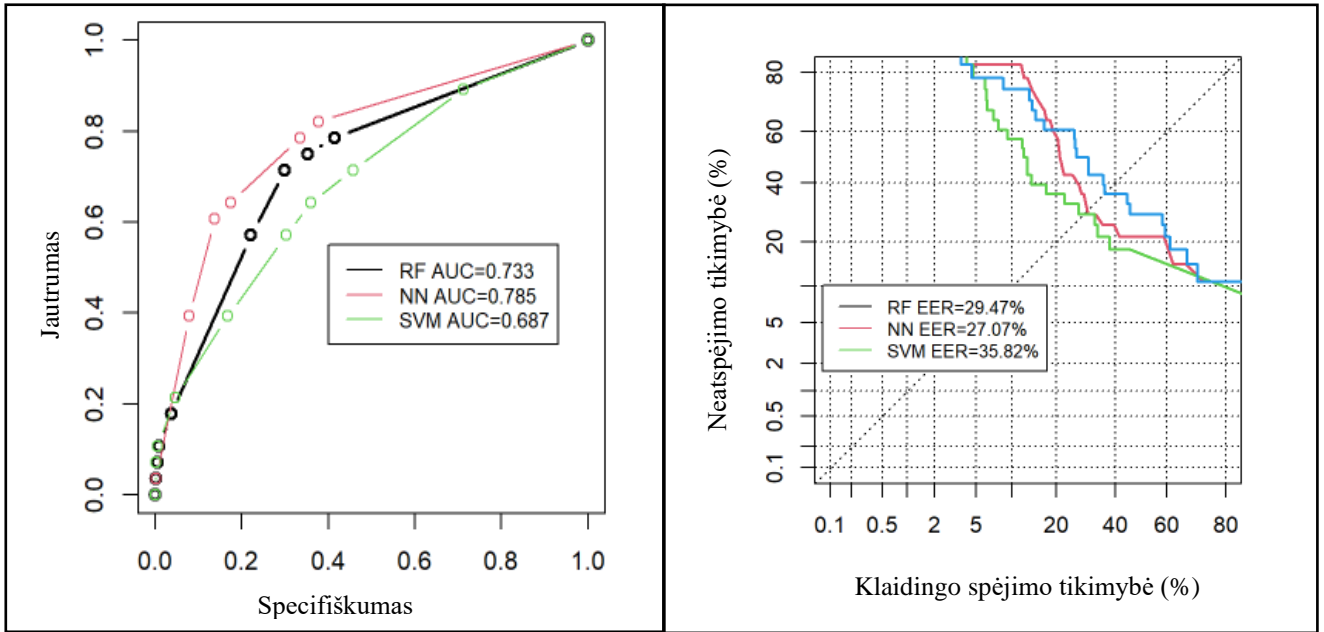
Vertinimo kriterijus	Duomenų failas	Pirminiai tyrimo duomenys (I)	Pagrindinės komponentės (II)	Pagrindinės komponentės ir kokybiniai kintamieji (III)
Tikslumas (angl. <i>Accuracy</i> – ACC)		0.93	<b>0.96</b>	0.93
Teisingai Teigiama (angl. <i>True Positive</i> - TP)		3	3	<b>5</b>
Klaidingai Neigiama (angl. <i>False Negative</i> – FN )		23	23	<b>21</b>

Vertinimo kriterijus	Duomenų failas	Pirminiai tyrimo duomenys	Pagrindinės komponentės	Pagrindinės komponentės ir kokybiniai kintamieji
Klaidingai Teigiama (angl. <i>False Positive</i> - FP)		34	<b>33</b>	37
Teisingai Neigiama (angl. <i>True Negative</i> - TN)		805	<b>806</b>	802
Preciziškumas		0.08	0.08	<b>0.12</b>
Išsamumas		0.11	0.12	<b>0.19</b>
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)		0.10	0.10	<b>0.15</b>
Lygių paklaidų lygis (angl. <i>Equal error rate</i> – EER)		35.82	42.71	<b>25.86</b>

Analizuojant sudarytų modelių įverčius, matoma, kad tikslumas (ACC – 0.96) geriausiai antrojo modelio su pagrindinių komponentių duomenų failu. Tačiau tiksliausiai skirtingas klases klasifikuoja trečiasis modelis su pagrindinių komponentių bei kokybinių kintamųjų duomenų failu. Preciziškumas – 0.12 rodo, kad 12 proc. iš suklasifikuotų kaip sukčiaujančių, asmenų buvo iš tikrųjų linkę į sukčiavimą. Taigi 0.88 proc. atvejų priskirtų sukčiavimui, buvo priskirti klaidingai. Modelio išsamumo matas rodo, kad modelis sugeba nustatyti 19 proc. sukčiavimo atvejų teisingai. F1-įverčio reikšmė – 0.15 rodo, kad bendras modelio tikslumas atsižvelgiant tiek į mažumos (sukčiaujančiųjų) tiek į daugumos klases, yra 15 proc.

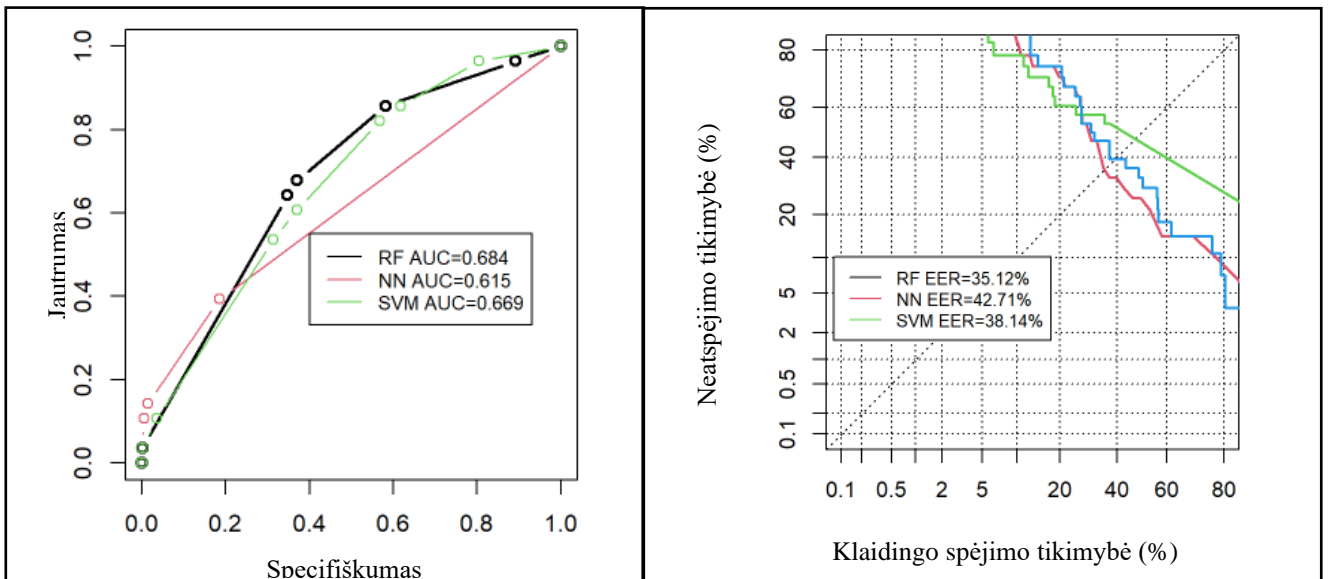
Apibendrinant visų sudarytų modelių rezultatus atsitiktinių miškų metodas geriausiai įverčius parodė su pirminiu duomenų failu. Atraminių vektorių mašinų bei neuroninių tinklų modelių didžiausi tikslumo įverčiai gauti naudojant trečiąjį duomenų failą su pagrindinėmis komponentėmis bei kokybiniais kintamaisiais.

Siekiant visų modelių palyginimo nubraižomos ROC kreivės bei apskaičiuojami AUC matai. 19 paveiksle pateikiamos ROC kreivės bei lygių klaidų paklaidos modelių sudarytų naudojant pradinį duomenų rinkinio failą. Didžiausias plotas po kreive neuroninių tinklų modelio – 0.785, antroje vietoje atsitiktinių miškų (AUC – 0.733), trečioje atraminių vektorių metodas (AUC – 0.687). Atitinkamai didžiausia paklaida atraminių vektorių mašinų (EER – 35.82 proc.), mažiausia neuroninių tinklų (EER – 27.07 proc.).



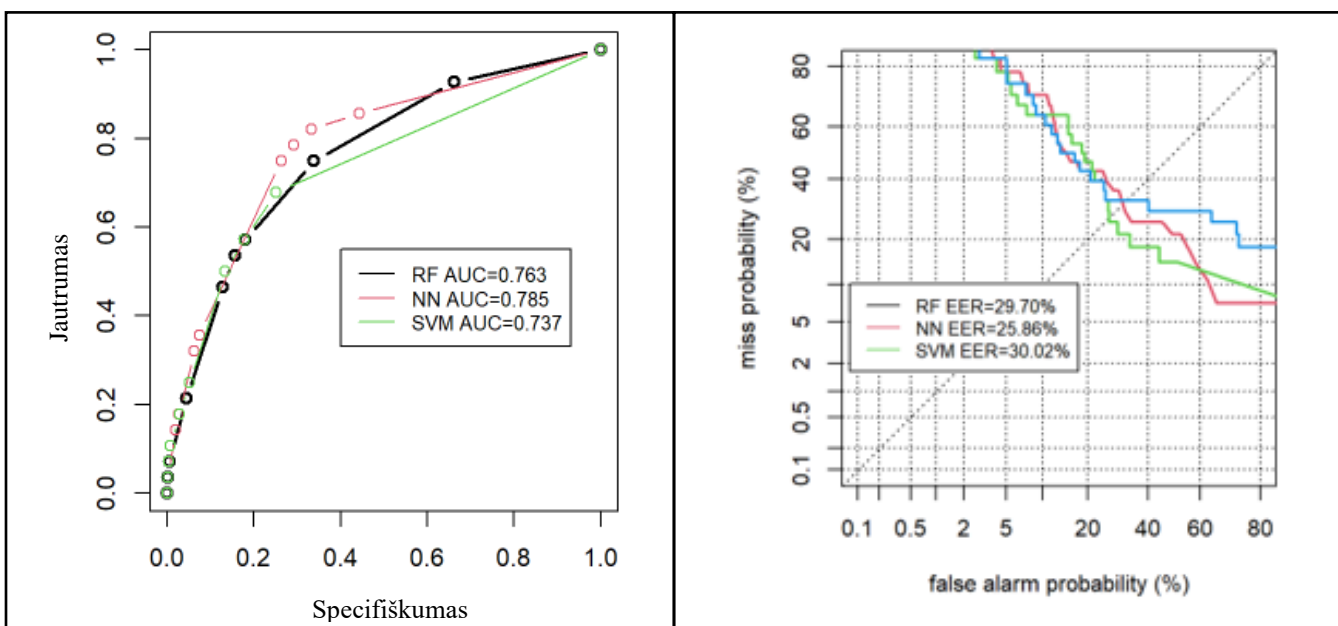
19 pav. Modelių sudarytų su pirmuoju duomenų rinkiniu ROC ir EER kreivių grafikai.

20 paveiksle pateikiamos ROC kreivės ir EER modelių sudarytų naudojant išskirtas pagrindines komponentes. AUC matai žymiai mažesni, geriausius rezultatus parodė atsitiktinių miškų modelis (AUC – 0.684; EER – 35.12 proc.), atraminių vektorių mašinų AUC – 0.669, ERR – 38.14, o neuroninių tinklų tikslumas mažiausias AUC - 0.615; ERR – 42.71 proc.)



20 pav. Modelių sudarytų su antruoju duomenų rinkiniu ROC ir EER kreivių grafikai.

21 paveiksle pateikiamos ROC kreivės modelių sudarytų naudojant tyrimo metu išskirtas pagrindines komponentes bei kokybinius tyrimo kintamuosius. Šiuo atveju, pagal plotą po ROC kreive geriausius rezultatus rodo neuroninių tinklų metodas (AUC – 0.785), lygių paklaidų lygis taip pat mažiausias (ERR – 25.86), antroje vietoje atsitiktiniai miškai (AUC – 0.763; ERR – 29.70), trečioje atraminių vektorių mašinos (AUC - 0.737, ERR – 30.02).



**21 pav.** Modelių sudarytų su trečiuoju duomenų rinkiniu ROC ir EER kreivių grafikai.

12 lentelėje palyginami visų sudarytų mašininio mokymosi metodų geriausi modeliai. Pagal gautus rezultatus geriausias tikslumas atsižvelgiant į sukčiaujančių asmenų klasifikavimą ir nesukčiaujančių atskirai yra sudaryto dirbtinių neuroninių tinklų metodo. Šis metodas su mažiausia paklaida sugebėjo identifikuoti tikslinę grupę – sukčiaujančius asmenis iš viso duomenų rinkinio.

**12 lentelė.** Sudarytų modelių vertinimo kriterijų palyginimas.

Vertinimo kriterijus	Modelis	Atsitiktiniai miškai (I)	Atraminiai vektoriai (III)	Neuroniniai tinklai (III)
Tikslumas (angl. <i>Accuracy</i> – ACC)		0.53	0.59	<b>0.93</b>
Plotas po kreive (AUC)		0.73	0.74	<b>0.79</b>
Teisingai Teigiama (angl. <i>True Positive</i> - TP)		21	19	<b>5</b>
Klaidingai Neigiama (angl. <i>False Negative</i> – FN)		<b>5</b>	7	21
Klaidingai Teigiama (angl. <i>False Positive</i> - FP)		398	344	<b>37</b>
Teisingai Neigiama (angl. <i>True Negative</i> - TN)		441	495	<b>802</b>
Preciziškumas		0.05	0.05	<b>0.12</b>
Išsamumas		<b>0.80</b>	0.73	0.19

Vertinimo kriterijus	Modelis	Atsitiktiniai miškai (I)	Atraminiai vektoriai (III)	Neuroniniai tinklai (III)
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)		0.09	0.10	<b>0.15</b>
Lygių paklaidų lygis (angl. <i>Equal error rate</i> – EER)		29.47	30.02	<b>25.86</b>

Tyrimo duomenų imtis maža ir nesubalansuota, tai gali sąlygoti mažus modelių tikslumo matus. Siekiant patikrinti ar mažoje duomenų imtyje pašalinus priklausomojo kintamojo disbalansą, naudojant metodologinėje dalyje aprašytą atsitiktinio imties mažinimo techniką, galima gauti modelį su tikslesniais įverčiais, sudaromas tiriamojo darbo teorinėje dalyje kaip reikalaujantis mažiausiai duomenų įvesties išskirtas atsitiktinių miškų modelis su trečiąja duomenų imtimi ją skaidant į mokymo bei testavimo.

### 3.2.5. Atsitiktiniai miškai. Priklausomojo kintamojo disbalanso šalinimas

Pradinė tyrimo imtis padalinama į mokymo bei testavimo imtis. Imtys dalinamos santykiu 1:2. Mokymo imčiai skiriama 2/3 duomenų rinkinio, testavimo imčiai – 1/3. Priklausomojo kintamojo pasiskirstymas mokymo ir testavimo duomenų imtyse pateikiamas 13 lentelėje.

**13 lentelė.** Priklausomojo kintamojo pasiskirstymas mokymo bei testavimo duomenų imtyse.

	Mokymo imtis	Testavimo imtis
Sukčiaujantis	16	10
Nesukčiaujantis	564	275

Tarp duomenų imties priklausomojo kintamojo klasių, kaip jau minėta, yra didelis disbalansas. Naudojant mašininio mokymosi algoritmus duomenyse esantis disbalansas sąlygoja modelio rezultatų nepatikimumą. Pritaikant klasifikavimo metodus (atsitiktiniai miškai, atraminių vektorių klasifikatorius, neuroniniai tinklai) nesubalansuotiems duomenims, gaunamas aukšto tikslumo modelis, tačiau žemo klasifikuojant mažumos klasę. Dėl to tyrimo metu pašalintas klasių disbalansas naudojant atsitiktinio imties retinimo techniką. Šio metodo tikslas atsitiktiniu būdu atrinkti ir pašalinti didžiosios klasės įrašus iš mokymo imties, kad būtų pasiektas pasirinktas imties balansas. Gauta naujo mokymo duomenų rinkinio priklausomo kintamojo pasiskirstymas pateikiamas 14 lentelėje.

**14 lentelė.** Pašalinto disbalanso mokymo imties priklausomo kintamojo pasiskirstymas.

	Mokymo imtis
Sukčiaujantis	16
Nesukčiaujantis	21

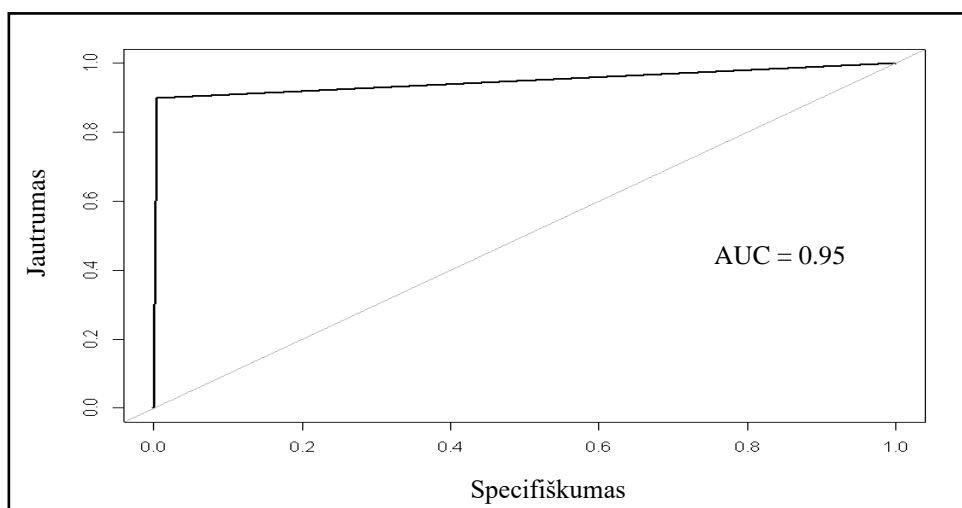
Naujajame mokymo duomenų rinkinyje priklausomas kintamasis pasiskirsto santykiu 57% ir 43%. Duomenų rinkinys sutvarkytas, priklausomojo kintamojo klasių disbalanso nėra. Naujoji mokymo imtis naudojama atsitiktinių miškų modeliui sudaryti.

Mokymo duomenų rinkinyje atsitiktinių miškų klasifikatorius puikiai suklasifikavo priklausomojo kintamojo klases (ACC = 0.97). Pritaikius modelį testavimo imčiai, gauti modelio įverčiai palyginami su geriausio sudaryto neuroninių tinklų modelio įverčiais 15 lentelėje.

**15 lentelė.** Atsitiktinių miškų ir neuroninių tinklų modelių palyginimas.

Vertinimo kriterijus	Modelis	Atsitiktiniai miškai (III)	Neuroniniai tinklai (III)
Tikslumas (angl. <i>Accuracy</i> – ACC)		<b>0.97</b>	0.93
Plotas po kreive (AUC)		<b>0.95</b>	0.79
Teisingai Teigiama (angl. <i>True Positive</i> - TP)		9	5
Klaidingai Neigiama (angl. <i>False Negative</i> – FN )		1	21
Klaidingai Teigiama (angl. <i>False Positive</i> - FP)		1	37
Teisingai Neigiama (angl. <i>True Negative</i> - TN)		274	802
Preciziškumas		<b>0.9</b>	0.12
Išsamumas		<b>0.99</b>	0.19
Išsamumo ir preciziškumo harmoninis vidurkis (F1-įvertis)		<b>0.9</b>	0.15

Sudaryto naujojo atsitiktinių miškų modelio tikslumo matai visi geresni už anksčiau sudarytų modelių matus. Tikslumas – 0.97, preciziškumas – 0.9, išsamumas – 0.99 (99 proc. sukčiavusiųjų atvejų suklasifikuota teisingai). Išsamumo ir preciziškumo harmoninis vidurkis – 0.90. Taigi atsitiktinių miškų modelis mokymui naudojant 2/3 duomenų imties su pašalintu duomenų disbalansu parodė geriausius rezultatus. 22 paveiksle pateikiama ROC kreivė. Ploto po kreive dydis taip pat išsiskiria iš visų sudarytų modelių (AUC = 0.95).



**22 pav.** Atsitiktinių miškų modelio su subalansuota imtimi ROC kreivė.



Palyginus gautus rezultatus, daroma išvada, kad turint mažą nesubalansuotą duomenų imtį tikslingiausia modelio sudarymui duomenų imtį skaidyti į mokymo ir testavimo imtis, o mokymo imtį apdoroti daugumos klasės kintamųjų šalinimo metodu.

Apibendrinant visus išanalizuotus modelių tinkamumo įverčius, daroma išvada, kad tinkamiausias metodas UAB „Bendras finansavimas“ klientų galimam sukčiavimui identifikuoti yra – pagrindinių komponentų išskyrimas, sujungimas su kokybiniais duomenų failo kintamaisiais bei atsitiktinių miškų klasifikatoriaus pritaikymas pašalinus imčių disbalansą. Šis metodas pasižymi aukščiausiu tikslumu tiek sukčiavusiems asmenims klasifikuoti, tiek nesukčiavimo atvejams teisingai nustatyti. Prasčiausiai turėtiesiems duomenims klasifikuoti tiko atraminių vektorių mašinų metodas. Šio detektoriaus sudarytų modelių klaidų tikimybė didžiausia, o klasių klasifikavimo tikslumas mažiausias.

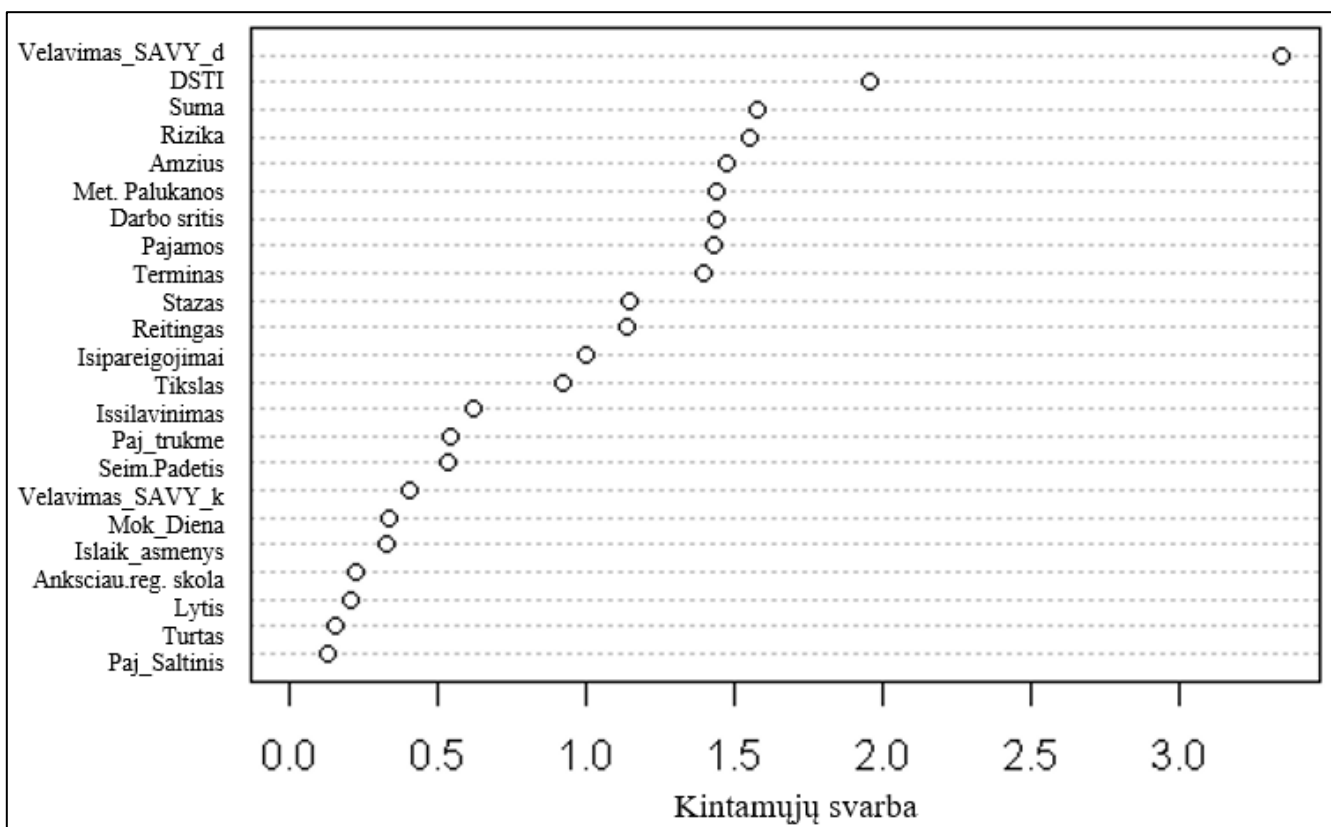
### **3.3. Sukčiavimą lemiančių veiksnių svarba**

Tiksliausias atsitiktinių miškų modelis, kaip svarbiausius veiksnius identifikuojančius klientų polinkį į sukčiavimą išskiria:

1. Vėlavimo dienų skaičių tarpusavio palyginimo platformoje per pastaruosius 6 mėnesius;
2. Įsipareigojimų ir pajamų santykį;
3. Pasiimto įsiskolinimo sumą;
4. Įvertintą nemokumo riziką;
5. Amžių.

23 paveiksle matoma, kad mažiausią įtaką sukčiauti daro:

1. Pajamų šaltinio kintamasis;
2. Viešai registruotas turtas;
3. Lytis;
4. Anksčiau turėtos blogo skolos;
5. Finansiškai išlaikomų asmenų skaičius.



23 pav. Kintamųjų svarba.

Iš kintamųjų svarbos, matoma, kad kuo daugiau praityje klientas buvo linkęs nevykdyti savo įsipareigojimų laiku, tuo didesnė tikimybė, kad jis bus sukčiaujantis. Analizuojant pasirinkimą investuoti į tarpusavio skolinimo platformoje esančią paskolą ar ne, verta atkreipti dėmesį į besiskolinamą kliento pinigų sumą, įgyjamą įsipareigojimų ir pajamų santykį po naujo įsipareigojimo gavimo bei kliento amžių. Visi šie faktoriai iš analizuotų veiksnių turi stipriausią įtaką galimam sukčiavimui nustatyti. Taip pat naudingas rodiklis pateikiamas TSPO platformų puslapiuose yra jau apskaičiuota bei pateikiama klientų nemokumo rizika. Tai vienas iš pirmųjų penkių didžiausių svarbumą turinčių kintamųjų modelyje.

### 3.4. Rezultatų palyginimas

Išskyrus patikimiausią modelį, sudaryto sukčiavimo rizikos vertinimo modelio tikslumo parametrai palyginami su mokslinėje literatūroje aprašytų sukčiavimo rizikos vertinimo modelių tikslumo įverčiais. Palyginimas pateikiamas 16-liktoje lentelėje.

16 lentelė. Sukčiavimo rizikos vertinimo modelių rezultatų palyginimas

Autoriai	Metodas	Tikslumas	Išsamumas	Preciziškumas
Tyrimo metu sudarytas geriausias modelis	Atsitiktiniai miškai	0.97	0.9	0.99
Kurt, Kenneth ir Rajendra (1995)	Prisitaikantis loginis tinklas	0.9	-	-
Sahin ir Duman (2011)	Atraminų vektorių mašinos	0.93	-	-

<b>Autoriai</b>	<b>Metodas</b>	<b>Tikslumas</b>	<b>Išsamumas</b>	<b>Preciziškumas</b>
Xu, Jennifer, Lu ir Chau. (2015)	Atsitiktiniai miškai	0.96	0.40	0.15
Ki ir Yoon (2017)	Puasono procesas	-	0.73	0.34
Wang (2018)	Gradiento didinimo sprendimų medžiai	0.88	-	-

Pagal turimus duomenis, vertinant sukčiavimo riziką didžiausiu tikslumu pasižymėjo sudaryti atsitiktinių miškų modeliai (ACC 0.97 bei 0.96) Toliau pagal tikslumą seka – Atraminių vektorių mašinos (ACC – 0.93), Prisitaikantis loginis tinklas (ACC – 0.9), gradiento didinimo sprendimų medžiai (ACC – 0.88). Išsamumo ir preciziškumo matus pateikia ne visi sukčiavimo vertinimo modelių autoriai, tačiau matoma, kad naudojant Puasono procesą teisingai suklasifikuota buvo 73 proc. sukčiavimo atvejų, vis dėlto nesukčiavimo atvejų didžioji dalis identifikuota klaidingai (Preciziškumas – 0.34). Tyrimo metu sudaryto modelio rezultatai geresni, išsamumo mato rodiklis – 0.9, vadinasi modelis teisingai suklasifikuoja 90 proc. sukčiavimo atvejų. Lyginant preciziškumo rodiklius, taip pat tikslumu išsiskiria tyrimo metu sudarytas atsitiktinių miškų sukčiavimo rizikos vertinimo modelis.

Apibendrinant tyrimo rezultatus, daroma išvada, kad sukčiavimo rizikos vertinimo modelio sudarymo technika yra stipriai priklausoma nuo turimo duomenų rinkinio. Tyrimo metu išanalizuoti bei praktiškai išbandyti trys mašininio mokymosi metodai: atsitiktiniai miškai, atraminių vektorių mašinos ir neuroniniai tinklai. Iš išmėgintų modelių, su turima maža duomenų imtimi geriausi modelio įverčiai gauti naudojant atsitiktinių miškų metodą. Atraminių vektorių mašinų modeliui turimas duomenų rinkinys buvo turintis per daug triukšmo, o neuroninių tinklų metodams reikalingos didesnės duomenų įvestys siekiant sudaryti patikimą sukčiavimo rizikos vertinimo modelį.

Sudarytas sukčiavimo rizikos vertinimo modelis gali būti pritaikomas tiek tarpusavio skolinimosi platformų klientų, tiek įprastų kreditų gavėjų, kreditinių kortelių turėtojų ar bet kokių kitų finansinių paslaugų gavėjų sukčiavimo rizikai vertinti. Ateities tyrimams siūlytina išbandyti sudarytą tyrimo modelį su didesne duomenų imtimi bei tyrimo duomenų disbalanso šalinimui pasirenkant kitas technikas (sukčiavimo atvejų dublikavimas imtyje (angl. *Over-sampling*); dirbtinių mažumos įrašų sukūrimas (angl. *Synthetic Minority Over-sampling Technique*)) bei ieškoti mašininio mokymo metodų pritaikomų ir mažų duomenų rinkinių analizei.

## Išvados

1. Tarpusavio skolinimosi platformos tai sparčiai populiarėjantis, inovatyvus bei nusistovėjusią skolinimosi tvarką keičiantis būdas. Naujiems finansinių paslaugų teikimo modeliams yra reikalingas griežtas reguliavimas. Nesant pakankamai stipraus, apgalvoto reguliavimo, finansų sistema susiduria su galima sukčiavimo rizika, kuri glaudžiai susijusi su visos sistemos ekonominiu stabilumu.
2. Mokslinėje literatūroje dažniausiai analizuojami elgsenos veiksniai, kurie skatina sukčiavimą. Iš sukčiavimo rizikos vertinimo modelių sudarytų remiantis klientų demografinėmis, ekonominėmis savybėmis bei prašomos paskolos požymiais, mokslinėje literatūroje, kaip svariausi sukčiavimą skatinantys veiksniai išskirti šie: paskolos suma bei tikslas, kliento pajamos, įsipareigojimai, išsilavinimas, lytis, amžius, šeimyninė padėtis bei darbo stažas. Sudaryto tyrimo metu išskirti svariausi veiksniai: istorinė kredito vėlavimo informacija, įsipareigojimų bei pajamų santykis, paskolos suma, tarpusavio skolinimo platformoje pateikiamas nemokumo rizikos įvertis bei amžius.
3. Dažniausiai mokslinėje literatūroje sutinkami sukčiavimo rizikos vertinimo modeliai sudaryti naudojant atsitiktinių miškų, atraminių vektorių mašinų bei dirbtinių neuroninių tinklų metodus. Tyrimo metu atlikta pagrindinių komponentių analizė bei sudaryti atsitiktinių miškų, atraminių vektorių mašinų ir dirbtinių neuroninių tinklų modeliai naudojant tiek apdorotus duomenų failus pagrindinių komponentių analizės metu, tiek neapdorotą duomenų rinkinį. Tyrimo rezultatai parodė, jog tinkamiausias metodas UAB „Bendras finansavimas“ sukčiavimo rizikai nustatyti yra – atsitiktinių miškų modelis su aštuoniomis pirmomis pagrindinėmis komponentėmis bei kokybiniais tyrimo kintamaisiais, duomenų rinkinio disbalansą pašalinus naudojant atsitiktinio imties retinimo techniką.
4. Palyginus mokslinėje literatūroje pateikiamus sukčiavimo rizikos vertinimo modelius pastebėta, kad didžiausiu tikslumu pasižymi vienas iš dažniausiai sukčiavimo rizikai vertinti pasitelkiamų metodų – atsitiktiniai miškai. Dauguma mokslinės literatūros autorių nepateikia sudarytų modelių išsamumo ir preciziškumo įverčių, todėl sunku teisingai palyginti modelius. Vis dėlto, pagal turimus duomenis, tyrimo metu sudarytas modelis taikliausiai suklasifikuoja tiek sukčiavimo, tiek nesukčiavimo atvejus lyginant su kitais mokslinėje literatūroje pateikiamais modeliais. Tačiau, kadangi tyrimo imtis buvo nedidelė, rezultatus vertinti reikia atsargiai ir modelio tikslumą patikrinti su naujais duomenimis.

## Literatūros sąrašas

1. ABBASI, K.; ALAM, A.; DU, M.; HUYNH, T. L. D. 2021. FinTech, SME Efficiency and National Culture: Evidence from OECD Countries. *Technological Forecasting & Social Change*, vol. 163, 12045. Available at <https://doi.org/10.1016/j.techfore.2020.120454>
2. Abreu., P., H., Araujo, H., J., Santos, M. S. ir Soares, J., P, (2018) Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches. *IEEE Computational Intelligence Magazine* 13(4):59-76 DOI:10.1109/MCI.2018.2866730.
3. Agarwal, Nishant, and Meghna Sharma. 2014. "Fraud Risk Prediction in Merchant-Bank Relationship Using Regression Modeling." *Vikalpa: The Journal for Decision Makers* 39(3):67–76. doi: 10.1177/0256090920140305.
4. Alešiūnaitė, Ž., Malakaitė, D. (2019) SUTELKTINIO FINANSAVIMO GALIMYBĖS IR GRĖSMĖS Vilniaus kolegija, Ekonomikos fakultetas
5. Anderson, Edward J. 2013. *Business Risk Management: Models and Analysis*. Hoboken Wiley.
6. Association of Certified Fraud Examiners. (2014). *Fraud Examiners Manual: financial transactions and fraud schemes*. Austin Texas: ACFE.
7. Avortri, C. ir Agbanyo R. 2020. "Determinants of Management Fraud in the Banking Sector of Ghana: The Perspective of the Diamond Fraud Theory." *Journal of Financial Crime* 28(1):142–55. doi: 10.1108/JFC-06-2020-0102.
8. Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. 2017. "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis." Pp. 1–9 in 2017 International Conference on Computing Networking and Informatics (ICCNi). Lagos: IEEE.
9. Barua, Sujoy, Divya Gavandi, Pooja Sangle, Leena Shinde, and Jyoti Ramteke. 2021. "Swindle: Predicting the Probability of Loan Defaults Using CatBoost Algorithm." Pp. 1710–15 in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. Erode, India: IEEE.
10. Bendroz, J. (2020) The History of P2P Lending. Prieiga per internetą: <https://p2pmarketdata.com/blog/p2p-lending-history/>
11. Bernotas, E., (2009) Finansinių ir fizinių procesų prognozavimo modelių tyrimas ir taikymas
12. Bytautas, A. ir Morkūnas, P. (2023) Kokias didžiausias rizikas finansų sistemai kelia finansinių technologijų sektorius? Prieiga per internetą: <https://www.lb.lt/lt/naujienos/finansiniu-technologiju-sektorius-ir-jo-riziku-finansu-sistemai-apzvalga>
13. Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32.
14. Brenner, L., Meyll, T., Stolper, O., & Walter, A. (2020). Consumer fraud victimization and financial well-being. *Journal of Economic Psychology*, 76, 102243. <https://doi.org/10.1016/j.joep.2019.102243>
15. Date, S. (2023) Time Series Analysis, Regression, and Forecasting. Prieiga per internetą: <https://timeseriesreasoning.com/contents/effect-of-irrelevant-variables/>
16. Dhar, V., ir Stein, R. M. (2017). FinTech platforms and strategy. *Communications of the ACM*, 60(10), 32-35.
17. Dmitrievsky, M. (2018) Random Decision Forest in Reinforcement learning. Prieiga per internetą: <https://www.mql5.com/en/articles/3856>
18. Emekter, R., Tu, Y., Jirasakuldech, B., et al.: Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Appl. Econ.* 47(1), 54–70 (2015)
19. Grigaliūnas, A. (2017). Mašinių mokymosi taikymas kompiuterinio saugumo uždaviniuose.

20. Gu, Dingwei, Zhengqing Gui, and Yangguang Huang. 2022. "Fintech Market and Regulation: Lessons from China's Peer-to-Peer Lending Platforms." SSRN Electronic Journal. doi: 10.2139/ssrn.4109877.
21. Hao, W., (2018) Detection of fraudulent users in P2P financial market. MATEC Web of Conferences 189, 06004 (2018) <https://doi.org/10.1051/mateconf/201818906004>
22. Horning, N, 2010 Random Forests : An algorithm for image classification and generation of continuous fields data sets
23. Huang S. Y. 2022 Why did the Peer-to-peer Lending Market Fail in China?
24. Huang, Robin Hui, Christine Menglu Wang. (2021) . "The Fall of Online P2P Lending in China: A Critique of the Central-Local Co-Regulatory Regime."
25. Yingying , Ying ir Yan (2019) What factors are influencing credit card customer's default behavior in China? A study based on survival analysis
26. Iris H-Y Chiu, Fintech and Disruptive Business Models in Financial Products, Intermediation and Markets- Policy Implications for Financial Regulators, 21 J. TECH. L. & POL'Y 55, 64 (2016).
27. Yu, Tao, and Wei Shen. 2019. "Funds Sharing Regulation in the Context of the Sharing Economy: Understanding the Logic of China's P2P Lending Regulation." Computer Law & Security Review 35(1):42–58. doi: 10.1016/j.clsr.2018.10.001.
28. Yufeng Kou; Chang-Tien Lu; S. Sirwongwattana; Yo-Ping Huang (2004) Survey of fraud detection techniques.
29. J. Hembram, B. Patel, H. Patil, and S. Jaswal, "Loan Default Forecasting using Data Mining", 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 1-4 June 2020.
30. Jiang, J., L. Liao, Z. Wang, and X. Zhang (2021). Government affiliation and peer-to-peer lending platforms in china. Journal of Empirical Finance 62, 87–106.
31. Juodzevičiūtė, G., (2020) Klasifikavimo algoritmų tyrimas ir optimizavimas vertinant kredito riziką
32. Labunskiy, A. (2023) Is Peer-to-Peer (P2P) Lending Safe? Experts Explain How to Avoid Fraud. Prieiga per internetą: <https://www.mob.id/is-peer-to-peer-p2p-lending-safe-experts-explain-how-to-avoid-fraud/>
33. Lerner.,M., Howard, E.,(2022) 7 Bad Reasons to Refinance Your Mortgage. Prieiga per internetą: <https://www.investopedia.com/mortgage/refinance/7-bad-reasons-to-refinance-mortgage/>
34. Lietuvos bankas (2018) Prieiga per internetą: <https://www.lb.lt/lt/rizikos-valdymas>
35. Mandrijauskaitė, G.(2018). "FINANSINĖS ATSKAITOMYBĖS SUKČIAVIMO RIZIKOS VERTINIMO MODELIS"
36. Mano Credit info. Kas yra kredito reitingas ir nuo ko jis priklauso? (2022) <https://www.manocreditinfo.lt/klausimai>
37. Meržvinskaitė, I. (2016). Tiesiosios žarnos vėžio epigenetinių modifikacijų tyrimas naudojant mašininio mokymosi metodus.
38. Meržvinskaitė, I., 2016. TIESIOSIOS ŽARNOS VĖŽIO EPIGENETINIŲ MODIFIKACIJŲ TYRIMAS NAUDOJANT MAŠININIO MOKYMOŠI METODUS
39. Misius, A. 2019, Klientų lojalumo modeliavimas: detekcijos ir išlikimo analizės uždavinių palyginimas.
40. Nemoto, N., Huang, B., ir Storey, D. 2019. "Optimal Regulation of P2P Lending for Small and Medium-Sized Enterprises." SSRN Electronic Journal. doi: 10.2139/ssrn.3313999.
41. Óskarsdóttir, M. (2021) Social Network Analytics for Supervised Fraud Detection in Insurance

42. Płoński P. (2019) Random Forests® vs Neural Networks: Which is Better, and When? Prieiga per internetą: <https://www.kdnuggets.com/2019/06/random-forest-vs-neural-network.html>
43. Precedence Research, (2022) P2P Payment Market Size to Worth Around USD 9.87 Trillion by 2030 Prieiga per internetą: <https://www.precedenceresearch.com/p2p-payment-market>
44. Pukėnas K. (2009) KOKYBINIŲ DUOMENŲ ANALIZĖ SPSS PROGRAMA. Mokomoji knyga.
45. Putri Westyas Saykita, Zuliani Dalimunthe, and Rachmadi Agus Triono. (2019). “How Perceived Risk Affects Continuance Intention to Invest through Peer-to-Peer Lending Platforms: Indonesia Case.” Prieiga per internetą: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3780605](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3780605)
46. Qiyi Wang, Hongyan Liu, Jun He, and Xiaoyong Du (2020) A Graph Attentive Network Model for P2P
47. Rahman, R. A., and I. S. Khair Anwar. 2014. “Types of Fraud among Islamic Banks in Malaysia.” International Journal of Trade, Economics and Finance 5(2):176–79. doi: 10.7763/IJTEF.2014.V5.365.
48. Ray, S. (2023) Learn How to Use Support Vector Machines (SVM) for Data Science. Prieiga per internetą: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>
49. Rai., G. (2017) The Vulnerability to Fraud: Factors, Motivations, and Fraud Detection and Deterrence. The vulnerability to fraud. 5, 8 – 11
50. Rodenbaugh, R.(2020) China's P2P Bubble and Lufax. Poor financial controls, bubbles, and a \$40bn Ant Group competitor. Prieiga per internetą: <https://eastmeetswest.substack.com/p/lufax>
51. SAVY, (2023). Investuotojų fondas Prieiga per internetą: <https://gosavy.com/investuotoju-fondas/>
52. Schultebraucks, L., (2017) Introduction to Support Vector Machines. Prieiga per internetą: <https://medium.com/@LSchultebraucks/introduction-to-support-vector-machines-9f8161ae2fcb>
53. Schumacker, R. E., and Lomax, R. G. (2004). A beginner's guide to structural equation modeling, Second edition. Mahwah, NJ: Lawrence Erlbaum Associates.
54. Shrestha, N. (2021) Factor Analysis as a Tool for Survey Analysis. American Journal of Applied Mathematics and Statistics, 2021, Vol. 9, No. 1, 4-11. DOI:10.12691/ajams-9-1-2
55. Smart Invest. (2021) Pelningiausia P2P platforma 2021 metais Prieiga per internetą: <https://smartinvest.lt/pelningiausia-p2p-platforma-2021-metais/>
56. Stankūnaitė, L. (2009) TIRIANČIOSIOS IR PATVIRTINANČIOSIOS FAKTORINĖS ANALIZĖS TYRIMO METODAI. Magistro baigiamasis darbas, Kaunas
57. Suryanarayana, K. (2020) A study on Risk and Return and their relation in various investments 1Department of Management Sciences, R.V.R & J.C. College of Engineering,Chowdavaram Guntur-19, AP, India.
58. TICHONOV, J. (2018) KLASIFIKAVIMO METODAIS GRINDŽIAMI SKAITMENINIŲ VAIZDŲ GLAUDINIMO SPRENDIMAI. Daktaro disertacija.
59. Vadeika, M. (2018) Pirkinių krepšelio analizė ir vartotojų pasitenkinimo prognozavimas.
60. Varghese.,D. (2018) Comparative Study on Classic Machine learning Algorithms. Quick summary on various ML algorithms. Prieiga per internetą: <https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222>
61. Verified Market Research (2021) Europe Peer to Peer (P2P) Lending Market Size By Type (Consumer Lending, Business Lending), By End-User (onsumer Credit Loans, Small Business

- Loans, Student Loans, Real Estate Loans), And Forecast Prieiga per internetą: <https://www.verifiedmarketresearch.com/product/europe-peer-to-peer-p2p-lending-market/>
62. Verikas, A ir Gelžninis, A. (2008). Neuroniniai tinklai ir neuroniniai skaičiavimai. Vadovėlis.
  63. Vorobjova, M. ir Kartašova, J. 2015. "LIETUVOS FINANSŲ SISTEMOS YPATUMAI."
  64. Wainberg, Michael, Babak Alipanahi, and Brendan J. Frey. 2016. "Are Random Forests Truly the Best Classifiers?"
  65. Wang, Hao, Zonghu Wang, Bin Zhang, and Jun Zhou. 2018. "Information Collection for Fraud Detection in P2P Financial Market" edited by N. Asnafi. MATEC Web of Conferences 189:06006. doi: 10.1051/mateconf/201818906006.
  66. Woodruff, J., (2019) What Are the Advantages of Decision Trees? Prieiga per internetą: <https://smallbusiness.chron.com/advantages-decision-trees-75226.html>
  67. World Bank. 2021. *Consumer Risks in Fintech: New Manifestations of Consumer Risks and Emerging Regulatory Approaches*. World Bank.
  68. Xu, Jennifer J., Yong Lu, and Michael Chau. 2015. "P2P Lending Fraud Detection: A Big Data Approach." Pp. 71–81 in *Intelligence and Security Informatics*. Vol. 9074, Lecture Notes in Computer Science, edited by M. Chau, G. A. Wang, and H. Chen. Cham: Springer International Publishing.
  69. Zalatorius, M. (2017). Naujasis bankų asociacijos vadovas: kita krizė bus visiškai kitokia. Prieiga per internetą: <https://www.lrt.lt/naujienos/verslas/4/185468/naujasis-banku-asociacijos-vadovas-kita-krize-bus-visiskai-kitokia>
  70. Zhang, Zerui, and Shaoli Jin. 2020. "Risk Assessment and Measurement of P2P Industry." Pp. 4163–66 in *2020 Chinese Control And Decision Conference (CCDC)*. Hefei, China: IEEE
  71. Žiugžda, A., (2023). Kur investuoti verslui, o ko vengti ?



## Informacijos šaltinių sąrašas

1. Lietuvos Respublikos baudžiamasis kodeksas (2022) <https://www.infolex.lt/ta/66150:str182>
2. Lietuvos Respublikos Seimas (2022) LIETUVOS RESPUBLIKOS VARTOJIMO KREDITO ĮSTATYMAS. Prieiga per internetą: <https://e.seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.390016/asr>

## Priedai

### 1 priedas. Kintamųjų paaiškinimas

„SAVY“ kredito reitingas	„SAVY“ apskaičiuojamas ir suteikiamas kredito reitingas 1 - A klasė – aukštas patikimumas (0,00 – 1,40%) 2 - B klasė – aukštesnis nei vidutinis patikimumas (1,40 – 4,99%) 3 - C klasė – vidutinis patikimumas (4,99 – 16,11%) 4 - D klasė – žemas patikimumas (16,11 – 41,23%) 5 - E klasė – žemiausias patikimumas (41,23 – 100,00%)
Tikslas	Skolintų lėšų panaudojimo tikslas. 1 – Refinansavimas. 2 – Vartojimas. 3 – Medicininėms išlaidoms. 4 – Mokslo/Studijų finansavimas. 5 – Nekilnojamo turto vystymas. 6 – Nekilnojamo turto pirkimas. 7 – Transporto priemonės pirkimas. 8 – Namų remontas 9 - Namų apyvokos daiktams pirkti. 10 -Verslo finansavimas.
Šaltinis	1 – Darbo užmokestis. 2 - Savarankiškai dirbantis. 3 – Pensininkas. 4 - Socialinė pašalpa. 5 -Statutinis pareigūnas 6 - Kita 7 - Nuoma
Darbo sritis	1 -Paslaugų teikimas 2 - Žemės ūkis 3 - Prekyba 4 - Transportas 5 - Kita 6 - Gamyba 7 – Finansinės paslaugos 8 - Energetika 9 – Informacinės technologijos 10 - Statyba 11 – Statutinis pareigūnas 12 – Valstybės tarnautojas
Išsilavinimas	1 - Nėra 2 - Pagrindinis 3 - Vidurinis 4 - Profesinis 5 - Aukštasis neuniversitetinis 6 - Aukštasis universitetinis.
Šeimyninė padėtis	1 - Susituokęs

	2 - Nesusituokęs 3 - Išsiskyęs 4 - Našlys
Turtas	1 - Automobilis 2 - Butas 3 - Namas 4 – Nekilnojamasis turtas 5 – Žemės sklypas 6 – Neturi.