

VILNIAUS UNIVERSITETAS
EKONOMIKOS IR VERSLO ADMINISTRAVIMO FAKULTETAS

EKONOMINĖ ANALITIKA

Beatričė Raukšaitė
MAGISTRO BAIGIAMASIS DARBAS

KREDITO RIZIKOS VERTINIMO MODELIS FIZINIAMS ASMENIMS	CREDIT RISK ASSESSMENT MODEL FOR INDIVIDUALS
---	---

Darbo vadovė: Prof., Dr. (HP) Rasa Kanapickienė

Vilnius, 2026

TURINYS

SANTRUMPOS	4
LENTELIŲ SĄRAŠAS	5
PAVEIKSLŲ SĄRAŠAS	6
PRIEDŲ SĄRAŠAS	7
IVADAS	8
1. KREDITO RIZIKOS VERTINIMO TEORINIAI ASPEKTAI	11
1.1 Kredito rizikos koncepcija ir iššūkiai	11
1.1.1 Kredito rizikos koncepcija	11
1.1.2 Kredito rizikos iššūkiai ir jų ekonominė reikšmė	12
1.2 Kredito rizikos vertinimo modelio sudarymo koncepcija	14
1.2.1 Kredito rizikos vertinimo modelio kintamųjų pasirinkimo ir imties formavimo teoriniai aspektai	14
1.2.2 Kredito rizikos vertinimo modelio sudarymui taikytų klasifikavimo metodų teoriniai aspektai	25
1.2.3 Kredito rizikos vertinimo modelio įvertinimui taikytų metodų teoriniai aspektai	31
2. KREDITO RIZIKOS VERTINIMO MODELIO FIZINIAMS ASMENIMS SUDARYMO METODIKA	34
2.1 Priklausomo kintamojo apibrėžimas	36
2.2 Nepriklausomų kintamųjų pasirinkimas	37
2.3 Imties sudarymas	42
2.4 Klasifikavimo metodų pasirinkimas	44
2.5 Modelio įvertinimo metodų pasirinkimas	46
2.6 Rangų skalės sistemos sudarymas	49
3. KREDITO RIZIKOS VERTINIMO MODELIO FIZINIAMS ASMENIMS REZULTATAI IR JŲ INTERPETACIJA	50

3.1 Modelio kūrimo imties sudarymas	50
3.2 Modelių sudarymas ir atranka	58
3.3 Optimalaus lūžio taško nustatymas ir rangų skalės sistemos sudarymas	61
IŠVADOS	63
PASIŪLYMAI	65
SANTRAUKA	66
SUMMARY	67
LITERATŪROS SĄRAŠAS	68
PRIEDAI	72

SANTRUMPOS

α – I rūšies klaidos rodiklis;

β – II rūšies klaidos rodiklis;

AUC – plotas po gavėjų charakteristikų kreive (angl. area under curve);

Ar – bendro tikslumo rodiklis;

Bar – subalansuoto tikslumo rodiklis;

CCR – teisingo klasifikavimo rodiklis;

F1 – jautrumo ir įsipareigojimų nevykdymo prognozės tikslumo rodiklis;

IV – informacinės vertės rodiklis (angl. information value ratio);

KS – Kolmogorovo-Smirnov rodiklis;

MCR – klaidingo klasifikavimo rodiklis;

NPV – „gerų“ skolininkų klasifikavimo tikslumo rodiklis;

PPV – įsipareigojimų nevykdymo prognozės tikslumo rodiklis;

Pr – „blogų“ skolininkų tikslumo rodiklis (angl. precision);

PD – nemokumo tikimybė;

PRDB – paskolų rizikos duomenų bazė;

ROC – gavėjų charakteristikų kreivė (angl. receiver operating characteristic curve);

ROS – mažumos didinimo technika (angl. over sampling);

RUS – daugumos mažinimo technika (angl. under sampling);

Se – jautrumo rodiklis (angl. sensitivity);

SMOTE – sintetinė mažumos didinimo technika (angl. Synthetic minority over-sampling technique);

Sp – specifiskumo rodiklis (angl. specificity);

TP, TN, FP, FN – dvinarinio klasifikavimo sprendimo tipai: teisingas teigiamas (angl. true positive), teisingas neigiamas (angl. true negative), klaidingas teigiamas (angl. false positive), klaidingai neigiamas (angl. false negative);

VIF – dispersijos mažėjimo daugiklis (angl. variance inflation factor);

WOE – įtakos svoris (angl. weight of evidence).

LENTELIŲ SĄRAŠAS

- 1 lentelė. Kredito rizikos apibrėžimų palyginimas
- 2 lentelė. Imties proporcijų ir sprendimo būdų mokslinės literatūros apibendrinimas
- 3 lentelė. Klasifikavimo metodų privalumų ir trūkumų mokslinės literatūros apibendrinimas
- 4 lentelė. Modelio įvertinimo metodų pasiskirstymo mokslinėje literatūroje apibendrinimas
- 5 lentelė. Nepriklausomų kintamųjų sąvadas
- 6 lentelė. Duomenų transformavimo variantų apibendrinimas
- 7 lentelė. Metodai, pasirinkti atlikti modelių įvertinimą
- 8 lentelė. Nepriklausomų kintamųjų informacinės vertės (IV)
- 9 lentelė. Atliktos šalinamosios logistinės regresijos (LR) rezultatai ir jų interpretacija
- 10 lentelė. Galutinis nepriklausomų kintamųjų sąvadas
- 11 lentelė. Geriausio imties varianto (var3) ir subalansavimo technikos (SMOTE) modelių rezultatų vertinimo metodų reikšmės
- 12 lentelė. Rangų skalės sistema, pasirinkto modelio atžvilgiu

PAVEIKSLŲ SĄRAŠAS

- 1 paveikslas. Mokslinio tyrimo atlikimo procesas
- 2 paveikslas. Nepriklausomų kintamųjų sąvado koreliacijos matrica (Kendall metodas)
- 3 paveikslas. Kredito rizikos vertinimo modelio sudarymo procesas
- 4 paveikslas. Geriausio imties varianto (var3) ir subalansavimo technikos (SMOTE) modelių ROC ir AUC reikšmės

PRIEDŲ SĄRAŠAS

- 1 priedas. Nepriklausomų kintamųjų stačiakampės diagramos
- 2 priedas. Pirminio nepriklausomų kintamųjų sąvado koreliacijos matrica
- 3 priedas. Nepriklausomų kintamųjų WOE įverčiai
- 4 priedas. Pirmo varianto (var1) duomenų imties modelių rezultatų vertinimo reikšmės
- 5 priedas. Antro varianto (var2) duomenų imties modelių rezultatų vertinimo reikšmės
- 6 priedas. Trečio varianto (var3) duomenų imties modelių rezultatų vertinimo reikšmės
- 7 priedas. Skirtingais imties variantais ir balansavimo technikomis paremtų modelių ROC ir AUC reikšmių rezultatai

IVADAS

Darbo aktualumas. 2007–2008 metais įvykusi finansų krizė atskleidė kredito rizikos vertinimo svarbą finansų įstaigose. Šių priemonių integravimas ir tinkamas jų valdymas gali užtikrinti sveiką kredito riziką, priimtina atitinkamai įstaigai, dėl kurios galimi nuostoliai nesukeltų įstaigos egzistencinių klausimų. Todėl, tik efektyvus kredito rizikos vertinimas gali apibrėžti potencialią galimybę apie kreditą gaunančius asmenis, kure negalės įvykdyti prisiimtų finansinių įsipareigojimų.

Tinkamas kredito rizikos vertinimas yra susijęs ne tik su pačia kreditus teikiančia įstaiga, bet ir tokiais finansų sektoriaus ūkio subjektais, kurie atlieka tarpininko vaidmenį tarp skolininko ir skolintojo. Tokią funkciją atlieka ir pakankamai šviežia finansų sektoriaus šaka, susijusi su finansinių technologijų platformomis, teikiančiomis paskolų palyginimo paslaugas. Šias paslaugas teikiantys ūkio subjektai sukuria paskolų rinkos aplinką, kurioje susitinka paskolas gaminantys ir vartojantys ūkio subjektai, kurie palygindami, gali pasirinkti labiausiai jų preferencijas atitinkantį produktą. Todėl, paskolų palyginimo aplinką sudarantiems ūkio subjektams tenka atsakinga pozicija ne tik palaikyti sveiką konkurenciją finansų rinkoje, tačiau ir padėti suvaldyti kredito riziką paskolas teikiančioms finansų įstaigoms. Šios funkcijos atlikimo galimybių pagrindumas yra susijęs su gaunamais dideliais duomenų kiekiais, susijusiais apie kiekvieną klientą nuo jo elgesio iki kredito istoriją, demografinius, finansinius aspektus apimančius duomenis. Todėl, sukurtas fizinių asmenų kredito rizikos vertinimo modelis būtų pagrįstas šališkumo išvengimu, kuris gali atsirasti iš išorinių ūkio subjektų, teikiančių kredito rizikos vertinimo balais paslaugas.

Analizuojamos temos ištirimo lygis. Iki šiol mokslinėje literatūroje kredito rizikos vertinimo modelių sudarymo pagrindumas yra siejamas su mašininio mokymosi metodais, tokiais kaip klasikiniai statistiniai metodai, bei pažangesnėmis dirbtinio intelekto metodų klasifikavimo technikomis (Bello, 2023; Byanjankar ir kt., 2015; Dzidzevičiūtė, 2013; Kang ir kt., 2021; B. Li, 2022; Z. Li ir kt., 2017; Liu ir kt., 2024; Sanz-Guerrero & Arroyo, 2024; Shen ir kt., 2019, 2021; Song ir kt., 2023; Suhadolnik ir kt., 2023; Špicas, 2017; Špicas ir kt., 2023; Y. Tian ir kt., 2018; Z. Tian ir kt., 2020; Verbraeken ir kt., 2013; Xiao ir kt., 2021). Nors šiuolaikinėje mokslinėje literatūroje pagrindžiamas pažangesnių klasifikavimo metodų taikymo kredito rizikos vertinimui pranašumas, visgi klasikiniai statistiniai metodai atlieka lyginimo funkciją (angl. benchmark) pažangesniems metodams vertinti (Kang ir kt., 2021; Shen ir kt., 2019, 2021; Suhadolnik ir kt., 2023; Špicas ir kt., 2023; Z. Tian ir kt., 2020; Xiao ir kt., 2021). Šių dviejų grupių klasifikavimo metodų taikymas kredito rizikos vertinimui yra pagrįstas ir tuo, kad nepaisant pažangesnių

dirbtinio intelekto metodų aukštesnio vertinimo našumo, yra pabrėžiamas jų trūkumas, susijęs su žemomis paaiškinimo galimybėmis. Tačiau, mokslinėje literatūroje atskleidžiamos naujos technikos, padedančios padidinti pažangesnių dirbtinio intelekto metodų paaiškinamumą, o tai sudaro galimybes finansų įstaigoms šias technikas pritaikyti ir kredito rizikos vertinime (Bello, 2023a; Dastile ir kt., 2020; Gramegna & Giudici, 2021; Malakauskas, 2023). Taip pat dėl kredito rizikos svarbos ir tikslaus „gerų“ ir „blogų“ klientų atskyrimo modelio sudarymo, mokslinėje literatūroje pabrėžiamos problemos, susijusios su nepakankamai atlikta nepriklausomų kintamųjų atranka, duomenų bruožų nepakankamu ištyrimu, ypač nesubalansuotais duomenimis, duomenų transformavimo stoka bei atrankos šališkumu, kurių integravimas į kredito rizikos vertinimą atskleidžiamas kaip reikšmingą poveikį sukeliantis modelio tikslumui (Kang ir kt., 2021; Z. Li ir kt., 2017; Shen ir kt., 2019; Y. Tian ir kt., 2018; Xiao ir kt., 2021; Zhang & Yu, 2024). Todėl, atsižvelgiant į mokslinėje literatūroje iškeltas problemas, šiuo darbu siekiama užpildyti spragas ir sudaryti fizinių asmenų kredito rizikos vertinimo modelį, kuris būtų tinkamas integruoti į paskolų palyginimo paslaugas teikiančios finansų įstaigos kredito rizikos valdymo priemonių rinkinį, tuo pačiu užtikrinant ir paskolas teikiančioms finansų įstaigoms pagalbą tinkamai valdyti kredito riziką.

Darbo naujumas. Mokslinio darbo naujumas yra pagrįstas trimis pagrindiniais aspektais, kurie išskiria jį iš ankstesnių mokslinių tyrimų ir užtikrina pridėtinę vertę Lietuvos kredito rizikos vertinimo srityje. Pirma, fizinių asmenų kredito rizikos vertinimo modelis pirmą kartą sukuriamas moksliniame lygmenyje, atsižvelgiant į Lietuvos finansų rinkos specifiką ir sąlygas, kas iki šiol nebuvo išsamiai analizuota. Antra, skolininkų imtis formuojama ne tik naudojantis klientą apibūdinančiais palyginimo platformos turimus duomenis, bet ir pasitelkiant konkrečių paskolas teikiančių finansų įstaigų informacija apie klientų elgseną vykdant finansinius įsipareigojimus realiomis jų aplinkos sąlygomis, taip užtikrinant sukurto modelio praktinį pritaikomumą. Trečia, modelis remiasi mokslinėje literatūroje iškeltomis kredito rizikos vertinimo problemomis, pavyzdžiui, nepakankamai atlikta nepriklausomų kintamųjų atranka, nepakankamu duomenų požymių ištyrimu, nesubalansuotais duomenimis bei duomenų transformavimo stoka. Siekiant užtikrinti objektyvesnius vertinimo rezultatus ir padidinti modelio tikslumą bei patikimumą, šios problemos yra integruojamos į modelio sudarymo procesą. Tokiu būdu šis mokslinis darbas ne tik apjungia mokslininkų išskirtas problemas kredito rizikos vertinime, bet ir sukuria pagrindą praktiniam Lietuvos finansų sektoriaus kredito rizikos vertinimo tobulinimui.

Darbo problema. Ar fizinių asmenų kredito rizikos vertinimo modelis, sukurtas Lietuvos rinkai, gali efektyviai atskirti mokių ir nemokių skolininkus pagal sudarytą kredito reitingų sistemą ir tuo pačiu užtikrinti efektyvų kredito rizikos valdymą?

Darbo tikslas. Atlikus mokslinės literatūros analizę, apibrėžti kredito rizikos keliamus iššūkius, todėl ir jos svarbą ekonominėje aplinkoje, bei išnagrinėjus pagrindinius modelio sudarymo proceso etapus, sukurti praktinį fizinių asmenų kredito rizikos vertinimo modelį.

Darbo uždaviniai:

1. Atlikus mokslinės literatūros analizę, apibrėžti kredito rizikos koncepciją, iššūkius ir jų ekonominę reikšmę;
2. Atlikus mokslinės literatūros analizę, išanalizuoti kredito rizikos vertinimo modelio sudarymo koncepciją: apibendrinti kintamųjų ir imties formavimo teorinius aspektus; kritiškai įvertinti taikomus klasifikavimo metodus kredito rizikos vertinimo modelio sudarymui, atskleidžiant jų privalumus ir trūkumus; apibendrinti kredito rizikos vertinimo modelio įvertinimui taikomus metodus;
3. Remiantis atlikta mokslinės literatūros analize, pagrįsti modelio sudarymo metodologijos pasirinkimą;
4. Pateikti fizinių asmenų kredito rizikos vertinimo modelio rezultatus bei jų vertinimą.

Darbo metodai. Darbe naudojami metodai skaidomi į teorinės analizės ir analitinio tyrimo metodus:

- *Teorinės analizės metodai.* Lyginamoji mokslinės literatūros analizė, sintezė, aprašomoji statistinė analizė.
- *Analitinio tyrimo metodai.* Duomenų transformavimo technikos, imties subalansavimo technikos, kiekybiniai duomenų analizės metodai, mašininio mokymosi klasifikavimo metodai, vertinimo metodai, reitingavimo sistema, teisės aktų analizė.

Darbo struktūra. Mokslinio darbo dėstomoji dalis išskiriama į tris skyrius. Pirmajame skyriuje mokslinės literatūros analizės pagrindu apibrėžiami kredito rizikos koncepcija, iššūkiai ir jų ekonominė reikšmė, išnagrinėjamas kredito rizikos vertinimo modelio sudarymo procesas. Antrajame skyriuje, kredito rizikos vertinimo modelio fiziniams asmenims sudarymo metodika, pristatomas kredito rizikos vertinimo modelio sudarymo techninis procesas, pagrįstas autoriaus atliktomis išvadomis iš pirmajame skyriuje atliktos mokslinės literatūros analizės. Trečiajame skyriuje, kredito rizikos vertinimo modelio fiziniams asmenims rezultatai ir jų interpretacija, atliekama tyrimų rezultatų interpretacija. Galiausiai mokslinis darbas yra užbaigiamas autoriaus išskirtomis pagrindinėmis išvadomis, pasiūlymais.

1. KREDITO RIZIKOS VERTINIMO TEORINIAI ASPEKTAI

1.1 Kredito rizikos koncepcija ir iššūkiai

Kredito rizikos valdymas yra esminė iš kreditą teikiančių finansų įstaigų veiklų, susijusių su atitinkamai finansų įstaigai priimtinu šios rizikos lygiu. Netinkamas kredito rizikos lygio prisiėmimas gali sukelti reikšmingų nuostolių ne tik atitinkamos finansinės įstaigos mikroaplinkoje bet ir bendrame finansų sektoriuje. Todėl, dėl plitimo savybe pasižyminčios kredito rizikos, svarbu apibrėžti jos sampratą, galimus iššūkius ir reikšmę ekonomikai.

1.1.1 Kredito rizikos koncepcija

Rizikos samprata finansų teorijoje yra glaudžiai susijusi su ateities neapibrėžtumu ir galimų rezultatų tikimybe (Bello, 2023; Kang ir kt., 2021; B. Li, 2022; Liu ir kt., 2024; Markov ir kt., 2022; Song ir kt., 2023; Suhadolnik ir kt., 2023; Špicas, 2017; Z. Tian ir kt., 2020; Xiao ir kt., 2021). Finansų sektoriuje rizika pasireiškia įvairiomis formomis, iš kurių dažniausiai išskiriamos rinkos, likvidumo, operacinė ir kredito rizika (B. Li, 2022; Špicas, 2017). Pastaroji laikoma viena reikšmingiausių, kadangi ji tiesiogiai susijusi su pagrindine finansų įstaigų veikla – skolinimu.

Mokslinėje literatūroje kredito rizikos sąvoka interpretuojama įvairiai, priklausomai nuo tyrimo konteksto, analizės lygmens ir nagrinėjamo subjekto. Siekiant sistemiskai palyginti skirtingų autorių požiūrius, 1 lentelėje pateikiami kredito rizikos apibrėžimai.

1 lentelė

Kredito rizikos apibrėžimų palyginimas

Šaltinis	Kredito rizikos apibrėžimas
B. Li, 2022	Kredito rizika apibrėžiama kaip galimi finansiniai nuostoliai, kylantys dėl skolininko nemokumo ar mokėjimų vėlavimo.
Bello, 2023	Kredito rizika apima ne tik nemokumo tikimybę, bet ir nuostolių dydį bei jų pasiskirstymą laike.
Kang, 2021	Kredito rizika apibrėžiama kaip negebėjimas išlaikyti pelningos veiklos, kai skolininkų nemokumas viršija apibrėžtą ribą.
Liu ir kt., 2024	Kredito rizika siejama su galimu finansinio stabilumo praradimu kredito įstaigose ir ekonomikoje.
Markov ir kt., 2022	Kredito rizika laikoma sisteminė finansų sektoriaus rizika, galinčia sukelti neigiamus kredito įstaigos rezultatus bei turėti grandininį poveikį kitoms įstaigos valdymo sritims.
Song, 2022	Kredito rizika apibrėžiama kaip skolininko nemokumo tikimybė.

Šaltinis	Kredito rizikos apibrėžimas
Suhadolnik ir kt., 2023	Kredito rizika siejama su skolininko nemokumo tikimybe, kurią lemia veiksniai, keičiantys skolininko elgseną.
Špicas, 2017	Kredito rizika apibrėžiama kaip tikimybė, kad faktiniai kredito įstaigos rezultatai skirsis nuo planinių.
Tian ir kt., 2020	Kredito rizika – tai neapibrėžtumas, susijęs su skolininko gebėjimu ateityje grąžinti skolą pagal sutarties sąlygas.
Xiao ir kt., 2021	Kredito rizika apibrėžiama kaip veiklos rezultato suprastėjimo ar bankroto tikimybė dėl skolininkų nemokumo.

Šaltinis: parengta autorės, remiantis nagrinėta moksline literatūra

Apibendrinant 1 lentelėje pateiktus apibrėžimus, galima teigti, kad dauguma autorių kredito riziką sieja su trimis pagrindiniais aspektais: nemokumo tikimybe, galimų nuostolių dydžiu ir neapibrėžtumu laike. Skirtumai tarp apibrėžimų dažniausiai kyla dėl analizuojamo lygmens – mikroekonominio (individualus skolininkas) arba makroekonominio (finansų sistema, tuo pačiu ir ekonomika).

Kredito rizikos vertinimą lemia tiek objektyvūs, tiek subjektyvūs veiksniai. Objektyvūs veiksniai apima skolininko finansinius rodiklius, kredito istoriją, makroekonominės sąlygas bei rinkos ciklišumą. Tuo tarpu subjektyvūs veiksniai siejami su skolininko elgsena, finansine drausme ir sprendimų priėmimo ypatumais (Tran ir kt., 2021).

Dėl šių daugialypių aspektų kredito rizika laikoma viena esminių finansų sektoriaus rizikų, turinčių reikšmingą poveikį ne tik pavieniams finansų rinkos dalyviams, bet ir visos ekonomikos stabilumui (Bello, 2023; Liu ir kt., 2024; Markov ir kt., 2022; Špicas, 2017; Tran ir kt., 2021; Xiao ir kt., 2021; Zhang & Yu, 2024).

1.1.2 Kredito rizikos iššūkiai ir jų ekonominė reikšmė

Kredito rizika pasižymi dideliu jautrumu makroekonominiams veiksniams, todėl jos materializacija dažnai sukelia tarpusavyje susijusių ūkio subjektų nuostolius, kurie ilgainiui paveikia viso finansų sektoriaus stabilumą. Ekonominės aplinkos pokyčiai, tokie kaip ekonomikos nuosmukis, infliacijos šuoliai ar monetarinės politikos griežtinimas, tiesiogiai veikia kredito gavėjų gebėjimą vykdyti finansinius įsipareigojimus, o netiesiogiai – kredito davėjų finansinę būklę (Z. Tian ir kt., 2020).

Kredito rizikos problematika pirmiausia pasireiškia kredito gavėjo lygmeniu, tačiau jos pobūdis skiriasi priklausomai nuo to, ar kredito gavėjas yra juridinis, ar fizinis asmuo. Juridinių asmenų atveju kredito rizika dažniausiai siejama su įmonių veiklos cikliškumu, pajamų svyravimais, investicinių projektų neapibrėžtumu bei priklausomybe nuo makroekonominių

sąlygų. Tuo tarpu fizinių asmenų kredito rizika yra labiau susijusi su asmens pajamų stabilumu, užimtumu, finansiniais įsipareigojimais ir individualia finansine drausme. Ekonominių sukrėtimų laikotarpiais, tokiais kaip recesijos ar restriktinės monetarinės politikos ciklai, fiziniai asmenys susiduria su didesne kredito rizika dėl galimo darbo praradimo, sumažėjusių pajamų ar išaugusios paskolų aptarnavimo naštos. Didėjant palūkanų normoms, ypač kintamų palūkanų paskolų atveju, ženkliai padidėja mėnesinių įmokų dydis, o tai padidina tikimybę, kad kredito gavėjas nebegalės laiku vykdyti prisiimtų finansinių įsipareigojimų. Dėl šių priežasčių fizinių asmenų kredito rizika pasižymi dideliu jautrumu tiek makroekonominiams, tiek individualiems socialiniams ir elgsenos veiksniams.

Kredito gavėjų įsipareigojimų nevykdymas tiesiogiai veikia kreditą teikiančių įstaigų veiklą. Viena iš pagrindinių kredito rizikos materializacijos pasekmių kredito davėjui yra neveiksnių paskolų (angl. non-performing loans, NPL) augimas. Didėjanti neveiksnių paskolų dalis pablogina finansų įstaigų turto kokybę, mažina pelningumą dėl didesnių atidėjinių formavimo ir gali neigiamai paveikti kapitalo pakankamumo rodiklius (Shen ir kt., 2019). Ilgainiui neefektyvus kredito rizikos valdymas gali lemti finansų įstaigų finansinio stabilumo silpnėjimą, mažesnes galimybes finansuoti ekonomiką bei sumažėjusį investuotojų ir indėlininkų pasitikėjimą. Be to, didėjanti kredito rizika verčia finansų įstaigas griežtinti skolinimo sąlygas, o tai dar labiau apriboja paskolų prieinamumą fiziniams asmenims ir gilina ekonomikos ciklinius svyravimus.

Efektyvus kredito rizikos valdymas paskolos davėjo lygmeniu yra glaudžiai susijęs su gebėjimu tiksliai atskirti „gerus“ ir „blogus“ skolininkus. Kredito vertinimo klaidos, atsirandančios dėl netikslaus klasifikavimo, gali sukelti reikšmingų finansinių nuostolių, todėl kredito rizikos vertinimo modelių tikslumas yra esminis veiksnys kreditą teikiančių finansų įstaigų veiklos tvarumui (Dastile ir kt., 2020).

Kai kredito rizika nėra tinkamai valdoma pavienių finansų įstaigų lygmeniu, jos pasekmės gali išplisti į visą finansų sistemą. Finansų įstaigų bankrotai ar reikšmingi finansiniai sunkumai gali sukelti vadinamąjį „drugelio efektą“, kai pradiniai nuostoliai virsta grandinine reakcija, paveikiančia kitus finansų rinkos dalyvius ir realųjį ekonomikos sektorių (Z. Tian ir kt., 2020). Makroekonominio lygmeniu didėjanti kredito rizika lemia kreditavimo apimčių mažėjimą, investicijų sulėtėjimą, vartojimo mažėjimą bei ekonomikos augimo tempo lėtėjimą. Esant sisteminiams sukrėtimams, kredito rizikos materializacija gali peraugti į finansines krizes, turinčias ilgalaikių neigiamų pasekmių tiek nacionalinei, tiek globaliai ekonomikai. Dėl šios priežasties kredito rizikos vertinimas ir valdymas yra laikomi vienu iš svarbiausių finansinio stabilumo užtikrinimo elementų.

1.2 Kredito rizikos vertinimo modelio sudarymo koncepcija

Kredito rizikos vertinimas yra esminis procesas, leidžiantis finansų įstaigoms palaikyti veiklos tęstinumą ir finansinį stabilumą tiek mikroekonominiu, tiek makroekonominiu lygmeniu. Mokslinėje literatūroje kredito rizikos vertinimas taikomas tiek fiziniams, tiek juridiniams asmenims, tačiau šių ekonominių vienetų kreditingumo vertinimas iš esmės skiriasi dėl nevienodos pajamų struktūros, veiklos tęstinumo pobūdžio ir elgsenos veiksnių. Dėl šių skirtumų vieningo, universalaus kredito rizikos vertinimo modelio, tinkamo tiek fiziniams, tiek juridiniams asmenims, sudarymas nėra metodologiškai pagrįstas. Vis dėlto, nepaisant vertinamų subjektų skirtumų, kredito rizikos vertinimo modelių kūrimas remiasi bendrais metodologiniais principais, susijusiais su gebėjimo vykdyti prisiimtus finansinius įsipareigojimus prognozavimu ir patikimu rizikos klasifikavimu.

Kredito rizikos vertinimo modelių sudarymas, nepriklausomai nuo vertinamo subjekto tipo, apima pamatinius etapus: priklausomojo kintamojo (įsipareigojimų nevykdymo fakto) apibrėžimą, nepriklausomųjų kintamųjų parinkimą, imties formavimą atsižvelgiant į duomenų struktūrą, tinkamo klasifikavimo metodo pasirinkimą bei sudaryto modelio patikimumo ir prognozavimo tikslumo įvertinimą. Šių etapų nuoseklus įgyvendinimas ir kritinis vertinimas yra būtini siekiant sukurti reprezentatyvų ir empiriškai pagrįstą kredito rizikos vertinimo modelį.

Atsižvelgiant į tai, kad šis darbas orientuotas į fizinių asmenų kredito rizikos vertinimą, tolesnėje literatūros analizėje pagrindinis dėmesys skiriamas būtent fizinių asmenų kredito rizikos vertinimo modelių sudarymo koncepcijai, jų struktūrai bei taikomiems metodams.

1.2.1 Kredito rizikos vertinimo modelio kintamųjų pasirinkimo ir imties formavimo teoriniai aspektai

Kredito rizikos vertinimo modelio efektyvumas tiesiogiai priklauso nuo teisingai pasirinktų kintamųjų (tiek įvesties, tiek išvesties) ir imties formavimo proceso. Šioje dalyje nagrinėjami teoriniai aspektai, apibūdinantys priklausomo kintamojo apibrėžimą, nepriklausomų kintamųjų pasirinkimą bei imties formavimo proceso apibrėžimą.

Priklausomo kintamojo apibrėžimas. Priklausomo kintamojo apibrėžimas yra paremtas „blogo“ kliento apibrėžimu, t.y. nustatant finansinių įsipareigojimų įvykdymo pradžios terminą, bei stebėjimo laikotarpio apibrėžimą, per kurį būtų atliekamas vertinimas „gerų“ ir „blogų“ klientų priskyrimui.

„Blogo“ kliento apibrėžimas. Dažniausiai kredito rizikos vertinimo modelio priklausomas kintamasis yra apibrėžiamas kaip dvinaris, atitinkantis „gero“ ir „blogo“ kliento reikšmes.

Priklausomo kintamojo apibrėžimo tikslumas yra tiesiogiai susijęs su kuriamo modelio efektyvumu. Jeigu „blogas“ kliento apibrėžimas yra per daug griežtas – išauga tikimybė, kad geras klientas bus priskiriamas prie blogų klientų (antros rūšies klaidos tikimybė). Ši situacija dažnai apibrėžiama kaip „techninis nemokumas“ (Špicas, 2017). Priešingu atveju, kai „blogas“ kliento apibrėžimas yra daugiau nuosaikesnis – blogo kliento tikimybė būti priskirtam prie gerų klientų išauga (pirmos rūšies klaidos tikimybė). Kredito riziką nagrinėjantys mokslininkai pastebi, kad pirmos rūšies klaidos tikimybės išaugimas yra susijęs su didesniu nuostoliu nei antros rūšies klaidos atžvilgiu. Todėl, „blogas“ kliento apibrėžimas yra svarbus veiksnys modelio sudarymo procese, o modelio efektyvumui išlaikyti yra svarbu, kad priklausomo kintamojo apibrėžimas tarp „gero“ ir „blogas“ kliento būtų subalansuotas. Šio modelio sudarymo proceso etapo užtikrinimui, verta atlikti analizę, susijusią su pasirinktu „blogas“ kliento apibrėžimu paremtu klientų priskyrimo prie „blogas“ atvejų reikšmingumo įvertinimu.

„Blogas“ kliento apibrėžimo požymių rinkinio formavimui pabrėžiama tikslinė apibrėžimo taikymo grupės svarba, kuri turi būti susijusi su homogenine paskolų grupe (Špicas, 2017). Taip pat pastebima, kad fizinių asmenų kredito rizikos vertinimo modeliuose „blogas“ kliento apibrėžimas yra siejamas su prisiimtų įsipareigojimų nevykdymu (Dzidzevičiūtė, 2013). Kliento prisiimtų įsipareigojimų nevykdymo požymis dažniausiai nustatomas kaip 90 dienų ar ilgesnį laikotarpį pradelstas įsipareigojimų įvykdymas, susijęs su mokėjimo terminu, kreditą suteikusiai finansų įstaigai. Taip pat 2024 metais išleistas Europos Parlamento ir Tarybos Reglamentas (ES) 2024/1623 keičiantis Reglamentą (ES) 575/2013 pagrindžia šio mokėjimo vėlavimo termino nustatymo reikšmingumą, kai 178 st., susijęs su įsipareigojančio asmens įsipareigojimo nevykdymu, apibrėžia įsipareigojimų nevykdantį asmenį, jei asmuo daugiau kaip 90 dienų vėluoja įvykdyti bet kokią reikšmingą kreditinį įsipareigojimą įstaigai (Europos Parlamento ir Tarybos reglamentas (ES) 2024/1623, 2024). Mokslinėje literatūroje pastebima, kad „blogas“ kliento apibrėžimas nėra paremtas tik teisės aktų rekomendacijomis, ekspertų požiūriu, t.y. greta šių kokybinių metodų, yra taikomi ir kiekybiniai metodai. Vienas iš jų yra Markovo migracijos matricos, kuriomis vertinamas kiekvienos paskolos pasiektas didžiausias mokėjimo pradelsimas (Špicas, 2017). Remiantis šiuo metodu būtų galima atlikti labiau pagrįstą „blogas“ kliento požymių nustatymą, susijusį su modelio sudarymui taikoma intimi.

Stebėjimo laikotarpio apibrėžimas. Stebėjimo laikotarpio nustatymu yra siekiama įvertinti laikotarpį, kuris apibrėžtų pakankamą terminą, skirtą paskoloms subręsti (Špicas, 2017). Taip pat pabrėžiama optimalaus stebėjimo laikotarpio nustatymo svarba, susijusi su tiesiogine įtaka modeliui prognozavimui, diskriminacinei galiai. Minima, kad esant per trumpam stebėjimo laikotarpio nustatymui, išauga tikimybė, kad galimas nuostolis bus ne pilnai įvertintas, o tokiu

atveju pirmos rūšies tikimybė bus linkusi išaugti. Per ilgas stebėjimo laikotarpis yra nesuderinamas su skolininkų savybėmis, kurios gali būti reikšmingai besiskiriančios, lyginant paskolos suteikimo pradžioje buvusiomis su stebėjimo laikotarpio pabaigoje esančiomis (Špicas, 2017). Todėl, optimaliam stebėjimo laikotarpio apibrėžimui dažnai taikomi kiekybiniai metodai, kaip kohortų analizė, kuria remiantis yra nustatomas toks laikotarpis, per kurį vidutinis vėlavimo dienų skaičius sparčiai auga, t.y. turi eksponentinio augimo požymių (Špicas, 2017). Remiantis šiuo kiekybiniu metodu, suformuojamos skirtingais mėnesiais suteiktos paskolų kohortų grupės, pagal kurias yra pavaizduojami bent kartą suvėlavusių paskolų rodikliai. Špicas (2017) disertacijoje naudoja skirtingus vėlavimo paskolų rodiklius, išskirtus į 30, 60, 90, 120, 180 dienų vėlavimus.

Nepriklausomų kintamųjų pasirinkimas. Kredito rizikos vertinimo modelių sudarymui ir priklausomo kintamojo vertinimui taikomi nepriklausomi kintamieji gali būti suskirstyti į elgsenos, demografinės, finansinės, paskolos charakteristikas bei makroekonominis veiksniai apibūdinančius kintamuosius.

Mokslinėje literatūroje išskirti kredito rizikos vertinimo modeliams reikšmingi fizinio asmens elgesį apibūdinantys kintamieji yra susiję su paraiškos pildymo trukme, grįžtančių klientų per trumpesnę nei 30 dienų laikotarpį įvertinimu, paraiškos pateikimo laiką darbo valandų atžvilgiu (Špicas ir kt., 2023). Demografinius apibūdinančius aspektus, autoriai išskiria amžių, vedybinę padėtį, lytį, gyvenamąją vietą, darbinę patirtį, būsto nuosavybės tipą, išsilavinimą (Byanjankar ir kt., 2015; Croux ir kt., 2020; Liu ir kt., 2024; Sanz-Guerrero & Arroyo, 2024; Suhadolnik ir kt., 2023; Špicas ir kt., 2023). Prie finansinę asmens padėtį apibūdinančių kintamųjų yra priskiriamos tvarios pajamos, skirtumas nuo paraiškoje nurodytų pajamų ir tvarių pajamų, įsipareigojimų ir pajamų rodiklis (angl. debt service to income) (toliau – DSTI), turimų aktyvių paskolų skaičius, su skirtingomis finansinėmis įstaigomis saistančių ryšių, pavyzdžiui, atsidarytų banko sąskaitų skirtinguose bankuose ar prisiimtų finansinių įsipareigojimų iš skirtingų finansinių įstaigų, skaičius, finansinė patirtis metais nuo pirmos atidarytos kredito linijos iki pasiimtos paskolos, ar istoriškai asmuo turėjo problemų su mokesčių mokėjimu, prieš pasiimant paskolą turimų paskolų suma, pajamų šaltinis, bendrų pajamų suma (Croux ir kt., 2020; Liu ir kt., 2024; Sanz-Guerrero & Arroyo, 2024; Suhadolnik ir kt., 2023; Špicas ir kt., 2023). Su paskolos charakteristikomis susiję kintamieji apibūdina paskolos trukmę, paraiškoje nurodytos paskolos sumą, paskolos paskirtį, paskolos metinės palūkanas ir, ar klientas yra naujas (Croux ir kt., 2020; Liu ir kt., 2024; Sanz-Guerrero & Arroyo, 2024; Suhadolnik ir kt., 2023; Špicas ir kt., 2023). Taip pat pastebima, kad vertinant kredito riziką, autoriai naudoja ir makroekonominis kintamuosius, kurie pasižymi reikšmingumu (Croux ir kt., 2020; Liu ir kt., 2024). Tokie kintamieji yra susiję su VIX indekso

(kitais dar laikomas jautrumo matu) vidurkiu, atitinkantį išduotos paskolos mėnesiui, taip pat ir bendro vidaus produkto augimo tempu.

Praktikoje dažnai nepriklausomų kintamųjų pasirinkimas yra susijęs su galimų įvesties kintamųjų sąvado sudarymu, o vėliau iš jų formuojamas nepriklausomų kintamųjų poaibis, kuris yra susijęs su atrankos kriterijais (Špicas, 2017). Atrankos kriterijus būtų galima išskirti į duomenų kokybės požymių ištyrimą, nepriklausomiems kintamiesiems įtakos svorių (toliau – WOE) apskaičiavimą, individualių kintamųjų diskriminacinės galios nustatymą, multikolinearumo ištyrimą ir šalinamosios regresijos atlikimą. *Pirma*, vertinama, ar yra reikšmingai trūkstamos dalies reikšmių nepriklausomajame kintamajame. *Antra*, apskaičiuojami WOE, kuriuos pritaikoma individualių nepriklausomų kintamųjų diskriminacinei galiai nustatyti dažniausiai literatūroje taikomų metodų – informacinės vertės (toliau – IV) nustatymui. WOE metodo pasirinkimas yra pagrįstas tuo, kad šio metodo priskirti kintamieji pasižymi tiksliau konkrečios reikšmių grupės rizikingumo atspindėjimu (Dzidzevičiūtė, 2013). WOE yra atvirkščiai susiję su kredito rizika, t.y. kuo reikšmė yra mažesnė, tuo fizinio asmens kredito rizika yra didesnė (Dzidzevičiūtė, 2013). Apskaičiuotos IV vertės pasižymi tiesioginiu ryšiu su nepriklausomo kintamojo skiriamąja galia. Mokslinėje literatūroje išskiriama, kad galutinio modelio nepriklausomų kintamųjų pasirinkimas yra susijęs su IV vertėmis, kurios yra didesnės nei 0.02 (Y. Tian ir kt., 2018). Šio sprendimo pagrindimas yra susijęs su IV verčių išskaidymu į grupes: 1) <0.02 – neturintys diskriminacinės galios kintamieji; 2) $0.02-0.1$ – silpną prognozinę galią talpinantys kintamieji; 3) $0.1-0.3$ – vidutinę prognozinę galią talpinantys kintamieji; 4) $0.3-0.5$ – stiprią prognozinę galią talpinantys kintamieji; 5) ≥ 0.5 – labai stiprią, dažnai neįtikėtiną prognozinę galią talpinantys kintamieji (Dastile & Celik, 2021). *Trečia*, individualių kintamųjų reikšmingumo nustatymui, mokslinėje literatūroje pastebimas ANOVA metodo pritaikymas, kuriuo siekiama įvertinti, ar nepriklausomų kintamųjų vidurkiai skirtingose priklausomo kintamojo grupėse reikšmingai skiriasi (Boguslauskas ir kt., 2011). *Ketvirta*, siekiant išvengti multikolinearumo problemos ir naudoti nepriklausomus kintamuosius, kurie talpina savyje skirtingą informaciją, pastebima, kad yra apskaičiuojamos koreliacijos matricos tarp kintamųjų (Špicas ir kt., 2023). Jeigu atitinkamos nepriklausomų kintamųjų poros koreliacijos koeficientas viršija 0.7, mažiau reikšmingas kintamasis yra šalinamas iš sudaryto nepriklausomo kintamojo sąvado poaibio (Špicas ir kt., 2023). Multikolinearumo problemai spręsti literatūros analizėje pastebimas ir dispersijos mažėjimo daugiklio (toliau – VIF) apskaičiavimas, kai daugiklio reikšmė siekia 4 ir didesnę įvertį, indikuoja apie multikolinearumo problemos egzistavimą (Dzidzevičiūtė, 2013). *Penkta*, pagal atrankos kriterijus sudaromas nepriklausomų kintamųjų sąvado poaibis, pašalinant atrankos kriterijų neatitikusius nepriklausomus kintamuosius ir atliekama šalinamoji regresija, kai

didžiausią nereikšmingą galią turintys kintamieji šalinami palaipsniui, kol modelį sudaro tik reikšmingą prognozinę galią turintys įvesties kintamieji.

Imties formavimas. Nagrinėjant mokslinę literatūrą, pastebima, kad imties formavimo procesas gali būti suskirstytas į 3 etapus, kurių integravimas į modelio kūrimo procesą sudarytų reikšmingą įtaką modelio veiksmingumui: 1) atmestų paraiškų (angl. rejected inference) įtraukimas; 2) duomenų požymių ištyrimas; 3) duomenų normalizavimas.

Atmestų paraiškų įtraukimas. Atmesta paraiška yra užklausa, kuri buvo vertinama, tačiau, remiantis nustatytais kriterijais, nebuvo priimta. Markov ir kt. (2022) atlikta sisteminė literatūros analizė, pagrįsta 2016-2021 metų moksliniais straipsniais, išskiria atmestų paraiškų neįtraukimo problemą efektyvumui, kaip vieną iš esminių, sudarant kredito rizikos vertinimo modelius. O atmestų paraiškų įtraukimas prisideda prie modelio efektyvumo didinimo (Markov ir kt., 2022). Tai pagrindžia ir Zhang & Yu (2024) atliktos sisteminės literatūros analizės išvados, kad duomenų kokybės vienas iš šaltinių, turinčių įtaką modeliavimo tikslumui, yra susijęs su sprendimu, ar įtraukti atmestas paraiškas, ar ne. Taip pat modelio sudarymas, remiantis tik priimtomis paraiškomis, nėra reprezentatyvus didžiausią riziką keliančių paraiškų atžvilgiu (Dzidzevičiūtė, 2013). Visgi, atmestų paraiškų įtraukimas į modelio sudarymo procesą ne visada pastebimas moksliniuose darbuose, susijusiuose su kredito rizikos vertinimu. Atmestų paraiškų neįtraukimas pagrindžiamas tuo, kad kredito rizikos vertinimo modeliuose priklausomas kintamasis yra pagrįstas faktu, ar per numatytą laikotarpį skolininkas įgavo „blogą“ skolininko požymių (Špicas, 2017a).

Visgi, šią problemą nagrinėjantys mokslininkai kredito rizikos vertinimo srityje, išskiria metodus, kaip būtų galima integruoti atmestas paraiškas (Dastile ir kt., 2020; Dzidzevičiūtė, 2013; Kang ir kt., 2021; Z. Li ir kt., 2017; Y. Tian ir kt., 2018). Dastile ir kt. (2020) atliktoje sisteminės mokslinės literatūros analizėje pateikia Kolmogorovo-Smirnovo statistikos pritaikomumą atmestų paraiškų įtraukimui, kai yra nustatomas ribinis balas, pagrįstas „gerų“ ir „blogų“ skolininkų pasiskirstymo maksimaliu atstumu. Nustatytas ribinis balas yra pritaikomas „blogų“ skolininkų prognozavimui atmestų paraiškų atžvilgiu, t.y. jei skolininko balas yra mažesnis už ribinį balą, skolininkas priskiriamas prie „blogų“ skolininkų grupės (Dastile ir kt., 2020).

Kitais atvejais, išskirti atmestų paraiškų įtraukimo metodai yra paremti trūkstamų priklausomo kintamojo reikšmių numatytomis prielaidomis, kaip reikšmių trūkumas visiškai atsitiktinis, reikšmių trūkumas atsitiktinis ir reikšmių trūkumas neatsitiktinis (Dzidzevičiūtė, 2013; Y. Tian ir kt., 2018). Kai trūkstamos priklausomo kintamojo reikšmės yra visiškai atsitiktinės (angl. missing completely at random) (toliau – MCAR), tokiu atveju trūkstamos reikšmės bus nepriklausomos nei nuo modelio įvesties kintamųjų, nei nuo išvesties. Tokia prielaida yra

pagrindžiama tokiais atvejais, kai paraiškų atmetimo procesas yra atsitiktinis, o trūkstamų reikšmių papildymas taip pat būtų atsitiktinis (Dzidzevičiūtė, 2013). Esant atsitiktinai trūkstamoms priklausomo kintamojo reikšmėms (angl. missing at random) (toliau – MAR), prielaidos pagrįstumas būtų susijęs, kai trūkstamos reikšmės būtų priklausomos nuo modelio nepriklausomų kintamųjų. O šiai grupei priskiriami metodai yra didinimas, mišrūs skirstiniai, daugiareikšmis įrašymas. Laikantis neatsitiktinai trūkstamų priklausomo kintamojo reikšmių (angl. missing not at random) (toliau – MNAR) prielaida, priklausomo kintamojo reikšmės yra priklausomos tiek nuo modelio nepriklausomų kintamųjų, tiek nuo priklausomo kintamojo reikšmių. Pasak Dzidzevičiūtės (2013), šia prielaida pagrįstų atmestų paraiškų įtraukimo metodų pasirinkimas yra sietinas su pakankamai gerai skirstančių paraiškas modeliu į priimtas ir atmestas, o galimi metodai yra išorinių duomenų įsigijimas, vidinių duomenų naudojimas, priskyrimas prie „blogų“ skolininkų, ekstrapoliacija, klasterizacija, dvifaktorinės išvados.

Mokslinė literatūra, nagrinėjanti atmestų paraiškų įtraukimo problemą pagrindžia padidėjusį kredito rizikos vertinimo modelių efektyvumą (Kang ir kt., 2021; Z. Li ir kt., 2017; Y. Tian ir kt., 2018). O kadangi kredito rizikos vertinimo modeliuose trūkstamų priklausomų kintamųjų reikšmių prielaida yra pagrįsta MAR, kai priėmimas ir atmetimas yra pagrįstas ir ekspertinio vertinimo pagal skolininko savybių rinkinį, galimas nežinomų reikšmių priskyrimas pagal priimtų skolininkų savybių ir priklausomo kintamojo reikšmių rinkinį. Kang ir kt. (2021) atmestų paraiškų įtraukimo problemą nagrinėjančiame moksliniame darbe, atmestoms paraiškoms yra pritaikomas priimtų paraiškų paskirstymo algoritmas bei parenkamas optimalus atmestų paraiškų imties dydis, kuris užtikrina maksimalų prognozuojamo modelio našumą. Kituose moksliniuose darbuose atmestų paraiškų įtraukimas yra paremtas daugiau mašininio mokymosi algoritmais, kaip įvairių tipų atraminių vektorių mašinos (toliau – SVM) – pusiau prižiūrimas SVM, branduolio neturinčiu kvadratinio paviršiaus SVM, kurių apmokymui priklausomos kintamojo reikšmės nėra būtinos (Z. Li ir kt., 2017; Y. Tian ir kt., 2018). Šių mokslinių darbų rezultatai atskleidžia padidintą modelio tikslumą, prognozuojant „gerus“ ir „blogus“ skolininkus, kai atmestos paraiškos yra integruojamos į modelio imtį.

Taigi, atmestų paraiškų neįtraukimas į kredito rizikos vertinimo modelį yra opi problema, dėl egzistuojančios imties šališkumo problemos, kuomet modelio skirstymas į „gerus“ ir „blogus“ skolininkus gali būti mažiau tikslus. Tokiu atveju nėra žinoma visa informacija apie skolininką, o tai sukelia kredito rizikos valdymo problemas, ypač tais atvejais, kai didelė dalis paraiškų yra atmetamos (Z. Li ir kt., 2017).

Duomenų požymių ištyrimas. Modelio sudarymo tikslumo užtikrinimui yra svarbus duomenų požymių ištyrimas. Zhang & Yu (2024) moksliniame darbe, nagrinėjančiame naujausią

mokslinę literatūrą, susijusią su kredito rizikos vertinimu, pateikiama kredito rizikos vertinimui skirta modeliavimo sistema, apimanti: 1) išorinius struktūros požymius, 2) duomenų kokybės požymius, 3) duomenų kiekio požymius ir 4) vidinius informacijos požymius. Šios sistemos integravimas į modelio sudarymas procesą yra kertinis prognozavimo tikslumui.

Išorinių struktūrų požymių ištirimas yra susijęs su duomenų šaltiniais, kuriuos būtų galima išskirti į pagalbinius duomenis, trečiųjų šalių duomenis ir kelių platformų duomenis (Zhang & Yu, 2024). Pagalbinių duomenų požymio egzistavimas duomenyse yra būdingas tuomet, kai yra nepakankamas duomenų kiekis, kuris užtikrintų geresnį esamų duomenų modeliavimo galimybes. Trečiųjų šalių duomenų požymio egzistavimas yra susijęs su siekiu padidinti modelio efektyvumą, naudojant duomenis, susijusius su skolininko elgseną apibūdinančiais duomenimis, kaip, pavyzdžiui, įrenginio naudojimo duomenys, bei su makroekonominiais kintamaisiais, socialine žiniasklaida bei operacijomis susijusiais duomenimis (Zhang & Yu, 2024). Šių požymių egzistavimas yra susijęs su modelio prognozavimo galimybių didinimu.

Duomenų kokybės požymių išyrimas yra susijęs su trūkstamų duomenų ir triukšmo duomenų analize (Zhang & Yu, 2024). Trūkstamų duomenų požymius galima išskirti į tris tipus: MCAR, MAR ir MNAR. Nuo šio požymio tipo nustatymo, priklauso, kokia bus taikoma elgsena trūkstamų duomenų apdorojimui. Pavyzdžiui, esant nustatytam MCAR tipui, duomenų eilutės gali būti pašalintos tiesiogiai. Triukšmo duomenys yra susiję su išskirtimis, kurios pasižymi klaidomis ar anomalijomis, kurių įtraukimas į modelį turėtų tiesioginę įtaką prognozavimo tikslumui. Triukšmo duomenų analizės atlikimui yra patartina taikyti kelis statistinius metodus, kurie padėtų užtikrinti išsamų triukšmo duomenų ištirimą, kaip vidurkis, standartinis nuokrypis, apribojimo intervalas, klasterizacija (Zhang & Yu, 2024).

Duomenų kiekio ištirimas apima imties dydžio ir dimensijų kiekio požymių nagrinėjimą, apdorojant duomenis (Zhang & Yu, 2024). Nagrinėjant imties dydžio savybes, mokslinėje literatūroje aiškus kriterijus nėra nustatomas, kada imtis yra laikoma kaip didelė, tačiau mažos imties dydžio atžvilgiu, jei imties dydžio ir nepriklausomų kintamųjų skaičiaus santykis yra mažesnis nei 10, laikoma, kad imtis yra maža (Zhang & Yu, 2024). Visgi, kiti autoriai mėgina nurodyti tikslų imties dydį, kuris laikomas pakankamu, kuriant kredito rizikos vertinimo modelį, kai imtį sudaro po 1500–5000 „gerų“ ir „blogų“ skolininkų (Dzidzevičiūtė, 2013). Siekiant padidinti modelio tikslumą, pastebima, kad mokslinėje literatūroje siekiama optimizuoti priklausomo kintamojo grupių dydį. Vienas iš tokių algoritmų, leidžiantis nustatyti optimalų skolininkų skaičių, yra Neymano ir Pearsono optimalus pasiskirstymas (Dzidzevičiūtė, 2013). Gautas optimalus skolininkų skaičius, būdingas atitinkamai grupei, tampa siektinu skaičiumi, kuriuo remiantis, faktinis atitinkamos grupės skaičius yra padidinamas ar sumažinamas

atitinkamai. Dimensijų kiekio požymio nagrinėjimas gali būti išskirtas į didelio ir mažo kiekio dimensijas. Didelio dimensijų kiekio požymio egzistavimui yra būdingas didelis tarpusavyje nesusijusių nepriklausomų kintamųjų skaičiui, kurie turi įtakos modelio tikslumui. Mažo dimensijų kiekio požymis yra susijęs, kai skolininką apibūdinančių kintamųjų kiekis yra ribotas, todėl tokio požymio egzistavimui spręsti yra naudojami funkcijų konstravimo metodai (Zhang & Yu, 2024).

Vidinių informacijos požymių ištirimas apima imties disbalanso ir retumo požymių egzistavimą duomenyse (Zhang & Yu, 2024). Imties disbalanso požymio egzistavimas duomenyse gali lemti imties šališkumą, todėl ir didesnę polinkį prognozuoti į daugumą sudarančią imties grupę. Tačiau, pastebima, kad kredito rizikos vertinimui taikytų klasifikavimo metodų jautrumas nesubalansuotai imčiai išsiskiria – atsitiktinio miško, gradiento didinimo mašininio mokymo algoritmai nepasižymi jautrumu, naudojant nesubalansuotą imtį, tačiau, sprendimo medžių algoritmas, kvadratinė diskriminacinė analizė, k-artimiausių kaimynų klasifikatoriai pasižymi jautrumu ir tuo pačiu gauti įverčiai nėra efektyvūs (Dastile ir kt., 2020). Siekiant išvengti imties šališkumo ir sudaryti reprezentatyvią imtį modelio sudarymui, yra naudojamos klasikinės imties balansavimo technikos, kaip daugumos mažinimo (angl. under sampling) (toliau – RUS), mažumos didinimo (angl. over sampling) (toliau – ROS). Tačiau, pastebima, kad RUS metodo pritaikymo atveju, kai daugumos imties grupės kiekis yra sumažinamas, gali lemti tam tikros reikšmingos informacijos apie skolininką praradimą, o ROS metodo pritaikymas, susijęs su mažumos imtį sudarančių duomenų eilučių dubliavimu, gali prisidėti prie triukšmo duomenyse atsiradimo (Kang ir kt., 2021; Shen ir kt., 2021; Xiao ir kt., 2021; Zhang & Yu, 2024). Taip pat šie tradiciniai atrankos metodai pasižymi savybe, kuri iš naujo perbalansuoja pradinis duomenis (Shen ir kt., 2021). Todėl, šių trūkumų eliminavimui, mokslinėje literatūroje, sprendžiančią imties disbalanso problemą, pastebimas mišrių atrankos metodų pritaikymas. Retumo požymio egzistavimas duomenyse yra susijęs su atitinkamo kintamojo atžvilgiu egzistuojančia reta informacija, apibūdinančia skolininką (Zhang & Yu, 2024). Tokių požymių egzistavimui duomenyse spręsti yra naudojami kintamųjų pasirinkimo metodas (angl. feature selection method) ir kintamųjų mažinimo technika (angl. dimension reduction technique) (Zhang & Yu, 2024). Modelio pranašumą didina ir tam tikri klasikiniai atrankos metodais paremti jų praplėtimai, kaip 2002 metais Chawla ir kt. pasiūlytas sintetinė mažumos didinimo technika (angl. Synthetic minority over-sampling technique) (toliau – SMOTE) (Dastile ir kt., 2020). Visgi, mokslinėje literatūroje pabrėžiama, kad vieno ar kito atrankos metodo pasirinkimas yra priklausomas nuo to, koks klasifikatorius yra naudojamas. Iš atliktos literatūros analizės, kurioje buvo sprendžiama duomenų disbalanso problema, pastebima, kad SMOTE ir ansamblinio klasifikavimo metodų

derinys turi didesnę pranašumą nei pastarojo atrankos metodo derinys su klasikiniiais, mašininio mokymosi klasifikavimo metodais (žr. 2 lentelę) (Shen ir kt., 2019; Špicas ir kt., 2023; Z. Tian ir kt., 2020). Visgi, vienas iš klasikinių metodų, RUS, pastebima, kad turi didesnę pranašumą, pasižymintį didesniu modelio rezultatų tikslumu, nei išmanieji atrankos metodai, kai taikoma kartu su ansamblinio klasifikavimo metodais (Xiao ir kt., 2021). Kiti autoriai, taip pat patvirtina, kad klasikinių metodų taikymas, kaip RUS, kartu su ansamblinio klasifikavimo metodais padidina modelio tikslumą, tačiau bendrame kontekste su klasikiniiais ir mašininio mokymosi klasifikavimo metodais, ansamblinio klasifikavimo metodai yra pranašesni, net ir nesubalansuotos imties atžvilgiu (Suhadolnik ir kt., 2023). Visgi, SMOTE metodas daugelio literatūros atžvilgiu pripažįstamas pranašesniu, o taip pat šios technikos išplėtimas, kaip Borderline-SMOTE, ir taikymas kartu su ansamblinio klasifikavimo metodais (XGB, GBDT), turi didesnę pranašumą nei nesubalansuotos imties atžvilgiu (Kang ir kt., 2021). Tačiau, Borderline-SMOTE metodas yra mažiau ištirtas, o autoriai pabrėžia, kad šis technikos patobulinimas gali sukelti papildomo triukšmo duomenyse, dėl sintetinių pavyzdžių persidengiančiose srityse bei pastebima, kad metodas nepasižymi efektyvumu, taikant didelių dimensijų duomenims (Kang ir kt., 2021). Taip pat pabrėžiama, kad patobulinti pradiniai SMOTE nepasižymi reikšmingu modelio efektyvumo padidinimu, lyginant su RUS, kuris savo pritaikymu yra paprastesnis nei kiti SMOTE plėtiniai (Markov ir kt., 2022).

2 lentelė

Imties proporcijų ir sprendimo būdų mokslinės literatūros apibendrinimas

Šaltinis	Imties proporcijos	Sprendimo būdas	Klasifikatorius
Byanjankar ir kt., 2015	16/84	-	Neuroniniai tinklai
Boguslauskas ir kt., 2011	25/75	-	Logistinė regresija
Suhadolnik ir kt., 2023	20/80	- RUS	Logistinė regresija
			Sprendimų medis
			K-artimiausi kaimynai
			Atraminių vektorių mašinos
			Nuroniniai tinklai
			Atsitiktiniai miškai
			Papildomi medžiai
			AdaBoost
			Gradient boosting
			XGB
Špicas ir kt., 2023	37/63	– RUS ROS SMOTE	Logistinė regresija
			Rekurentiniai neuroniniai tinklai
			Atraminių vektorių mašinos
			XGB
Xiao ir kt., 2021	16/84	– RUS ROS OSS CBOS SMOTE SMOTE + ENN ADASYN SL-SMOTE MWMOTE	NB
			Sprendimų medis
			Neuroniniai tinklai
			Atraminių vektorių mašinos
			Logistinė regresija
			Bagging
			AdaBoost
			RSS
			Atsitiktiniai miškai
Kang ir kt., 2021	5/95	– RI-DAB RI-EXP RI-LS RI-BSLS BM-BS Borderline-SMOTE	LR
			Atraminių vektorių mašinos
			Atsitiktiniai miškai
			GBDT
			XGB
			LightGBM
Shen ir kt., 2019	45/55 ir 30/70	SMOTE	AdaBoost ir atgalinio sklidimo neuroniniai tinklai

Šaltinis	Imties proporcijos	Sprendimo būdas	Klasifikatorius
			Tiesinė diskriminantinė analizė
			Logistinė regresija
			K-artimiausias kaimynas
			Atraminų vektorių mašinos
			Klasifikavimo medis
			NB
			BP
			Logistinė regresija
			Atraminų vektorių mašinos
			AdaBoost
			Atsitiktinis miškas
			MLP neuroniniai tinklai
			Sprendimų medis
			GBDT
			Neuroniniai tinklai
			Tiesinė diskriminantinė analizė
			Logistinė regresija
			K-artimiausias kaimynas
			Atraminų vektorių mašinos
			NB
			CART
			LSTM
			Kvadratinė diskriminantinė analizė

Šaltinis: sudaryta autorės, remiantis nagrinėta mokslinė literatūra

Šių požymių egzistavimas kredito rizikos vertinimo modelio sudarymui taikomuose duomenyse, turi įtakos modelio prognozavimo tikslumui. Todėl, siekiant sudaryti efektyvų ir populiarią reprezentuojantį modelį, yra svarbu užtikrinti aptartą duomenų požymių ištyrimą.

Normalizavimas. Sudarytų modelių rezultatų našumui padidinti, yra svarbus ir kintamųjų normalizavimas, taikant įvairias transformacijas. Taip pat normalizavimo pagrindumas yra susijęs ir su galimybe palyginti skirtingomis technikomis pagrįstų klasifikavimo metodų rezultatus (Z. Tian ir kt., 2020). Markov ir kt. (2022) atliktoje sisteminėje mokslinės literatūros apžvalgoje, išskiria kelias kintamųjų normalizavimo technikas:

1) Kintamųjų standartizavimas – kiekybinių kintamųjų atveju pastebimas Z balo normalizavimo technikos pritaikomumas, o kokybiniams taikoma WOE metodika (Song ir kt., 2023).

2) Kiekvieną kiekybinį kintamąjį reprezentuojančių duomenų susiejimas su intervalu [0,1]:

$$x_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}.$$

3) Kiekvienos kiekybinio kintamojo reikšmės padalijimas iš didžiausios atitinkamo kintamojo reikšmės.

Apžvelgtus imties formavimo aspektus apibūdinančius mokslinėje literatūroje, svarbus atmestų paraiškų įtraukimo, duomenų požymių ištyrimo, susijusio su išoriniais struktūros, duomenų kokybės, duomenų kiekio ir vidiniais informacijos požymiais, bei kintamųjų normalizavimo etapai, kurių įtraukimas į modelio sudarymą yra esminis, siekiant gauti populiaciją reprezentuojančius ir efektyvius rezultatus.

1.2.2 Kredito rizikos vertinimo modelio sudarymui taikytų klasifikavimo metodų teoriniai aspektai

Šių dienų mokslinėje literatūroje pastebimas platus klasifikavimo metodų taikymas kredito rizikos vertinimo srityje. Remiantis naujausia literatūra, kredito rizikos vertinimui taikomus klasifikavimo metodus pagal klasifikatorių skaičių ir modelio mechanizmą autoriai skiria į tradicinius pavienius klasifikatoriaus, dirbtinio intelekto pavienio klasifikatoriaus ir ansamblinio (grupinių) kelių klasifikatorių metodus (Zhang & Yu, 2024).

Tradiciniai klasifikavimo metodai. Šiai klasifikatorių grupei priskiriami metodai apima logistinę regresiją (toliau – LR), tiesinę diskriminantinę analizę (toliau – LDA), k artimiausius kaimynus (toliau – KNN), naivų Bayes (toliau – NB), sprendimų medžius (toliau – DT). Šiai klasifikatorių grupei priskiriami metodai pasižymi nesudėtingu pritaikymu ir aukštomis paaiškinimo galimybėmis (Zhang & Yu, 2024). Dėl šių priežasčių, dalis mokslinės literatūros, susijusios su kredito rizikos vertinimu, autorių priimdavo sprendimus taikyti šiuos klasifikavimo metodus (Boguslauskas ir kt., 2011; Dzidzevičiūtė, 2013; Špicas, 2017). Taip pat ankstesnių metų mokslinėje literatūroje sprendimą lėmė ir tai, kad tuo laikotarpiu nebuvo atliktas sudėtingesnių klasifikavimo metodų pakankamas ištyrimas, todėl autoriai rėmėsi aukštesnį patikimumą turinčiais metodais, kaip LR (Boguslauskas ir kt., 2011; Dzidzevičiūtė, 2013; Špicas, 2017). Visgi, senesnės mokslinės literatūros dalis autorių, siekiančių integruoti naujas technikas į kredito rizikos vertinimą ir mažiau ištirtų metodų aplinkoje, atskleidžia šiai klasifikavimo grupei nepriskiriamų metodų pranašumą prieš tradicinius klasifikavimo metodus (Byanjankar ir kt., 2015). Byanjankar

ir kt. (2015) taikytų dirbtinio intelekto neuroninių tinklų (toliau – ANN) ir LR metodų prognozavimo galimybes vertinimo kredito rizikos kontekste, atskleidė ANN nuoseklumą ir neapmokytos imties atžvilgiu, ko buvo pasigendama, taikant LR metodą. Taigi, šių dienų literatūroje šiai grupei priskiriami metodai yra dažnai kritikuojami, dėl šališkumo, todėl ir mažesnio efektyvumo, tikslumo (Dastile ir kt., 2020; Zhang & Yu, 2024). Taip pat ir trūkumo aptinkant netiesinius ryšius tarp kintamųjų (Zhang & Yu, 2024). Metodų tikslumo problema kyla ir iš taikomų prielaidų tokiems modeliams, kaip LR, LDA, kai laikoma, kad nepriklausomi kintamieji yra nepriklausomi tarpusavyje (Shen ir kt., 2021). Taip pat pabrėžiamas šių metodų jautrumas nesubalansuotoms priklausomo kintamojo grupėms (Kang ir kt., 2021). Visgi, nepaisant tradicinių klasifikavimo metodų trūkumų, dažnai autoriai renkasi LR, DT metodus, kaip lyginamuosius metodus (Dastile ir kt., 2020). LR metodo išskyrimas yra pagrįstas, kadangi daugelio empirinių tyrimų atžvilgiu, kai yra lyginami skirtingų metodų tikslumas, pastarasis metodas pasižymi pranašumu prieš tradicinių klasifikavimo metodų grupei priskiriamus likusius metodus (Y. Tian ir kt., 2018). Visgi, kitų klasifikavimo grupių atžvilgiu, tokie klasifikavimo metodai, kaip LR, DT, tampantys atsvara kredito rizikos vertinimui skirtų modelių rezultatų vertinimui, nusileidžia sudėtingesniu mechanizmu pagrįstiems metodams (Shen ir kt., 2019; Suhadolnik ir kt., 2023).

Dirbtinio intelekto klasifikavimo metodai. Šiai klasifikatorių grupei yra priskiriami tokie metodai, kaip ANN, konvoliuciniai neuroniniai tinklai (toliau – CNN), atraminių vektorių mašinos (toliau – SVM), atsitiktiniai miškai (toliau – RF), ilgos trumpo laikotarpio atminties metodas (toliau – LSTM), klasifikacijos ir regresijos medžių metodas (toliau – CART). Šie klasifikavimo metodai pasižymi aukštu tikslumu, tačiau pasižymi jautriais parametrais, todėl siekiant optimizuoti modelį, kyla ir per didelio pritaikymo (angl. overfitting) rizika (Zhang & Yu, 2024). Šią problemą atskleidė ir Boguslauskas & Mileris (2009) empiriniame kredito rizikos vertinimo tyrime, kuriame nuo tam tikros duomenų eilutės ribos modelio tikslumas pasižymėjo neigiamu ribiniu poveikiu, kai papildomas informacijos kiekis mažino modelio tikslumą. Visgi, atsižvelgiant į pranašumus, šiai grupei priskiriami klasifikavimo metodai geba apimti sudėtingus ryšius duomenų rinkiniuose taip pat ir besikeičiančias sąlygas, todėl yra labiau pritaikomi prie realios aplinkos sąlygų (Bello, 2023). SVM naudojama branduolio funkcija padeda apimti netiesiškumą duomenų rinkiniuose, tačiau, tai apsunkina modelio sudarymą, dėl tinkamos branduolio funkcijos ir atitinkamų parametrų paieškos (Y. Tian ir kt., 2018). O tai yra atsvara LR klasifikavimo metodams, kai ryšiai tarp kintamųjų yra netiesiniai (Dastile ir kt., 2020). Mokslinė literatūra, pagrįsta empiriniais tyrimais, atskleidžia SVM metodu įvertintų modelių aukštesnį našumą nei LR atveju, kai vertinama kredito rizika (Y. Tian ir kt., 2018). Tačiau, pastebima ir prieštaringų rezultatų, kai

didesnių žmogiškųjų resursų reikalaujantis SVM metodas ne visai atvejais pasižymės pranašumu prieš tradicinį LR klasifikavimo metodą. Tai atskleidžia sudarytas internetinės paskolų nemokumo prognozės modelis, kai LR pasižymėjo aukštesniu našumu nei pritaikyto SVM metodo atžvilgiu (Baodong Li, 2022). Visgi, atliktas tyrimas atskleidžia, kad sudėtingesnių technikų metodai, pagrįsti neuroninių tinklų šakos gilaus mokymo metodais, pasižymi aukščiausiu našumu lyginamųjų metodų grupėje. O taip pat ir mokslinės literatūros autorių pasiūlyti SVM metodo plėtiniai, kaip be branduolio funkcijos SVM (toliau – FQSSVM) ar pusiau prižiūrimo SVM (toliau – SSVM), turi reikšmingą pranašumą nemokumo tikimybės rezultatų tikslumo atžvilgiu (Y. Tian ir kt., 2018).

Ansamblinio klasifikavimo metodai. Šiai klasifikatorių grupei priskiriami metodai atitinka stiprinimo (angl. boosting) metodus, kaip AdaBoost, XGB, gradiento didinimo sprendimo medžius (toliau – GBDT), bei maišymo (angl. bagging) metodus. Stiprinimo metodui yra priskiriamos mažiau veikiančios klasifikatorių grupės, kurios yra sujungiamos (Suhadolnik ir kt., 2023). Stiprinimo metodas yra pagrįstas silpnų klasifikatorių pavertimu stipriais, kai silpni klasifikatoriai yra nuosekliai mokomi, kad atskirų klasifikatorių klaidingi atpažinimai sulauktų didesnio dėmesio (Zhang & Yu, 2024). Šiai grupei priskiriamų metodų veikimas yra priklausomas nuo naudojamų klasifikatorių savybių, kaip tikslumo, ir pačių sudedamųjų klasifikatorių įvairovės (Shen ir kt., 2021). Mokslinėje literatūroje pabrėžiama, kad šiai grupei priskiriami metodai pasižymi aukštu tikslumu (Zhang & Yu, 2024). Bei pastebima, kad ansambliniai klasifikatoriai veikia geriau nei atskiri klasifikatoriai, tiek ir tradicinių, tiek ir dirbtinio intelekto klasifikavimo metodų atžvilgiu (Shen ir kt., 2021). Tai pagrindžia ir atlikti empiriniai tyrimai kredito rizikos vertinimo srityje, kai lyginamųjų tradicinių ir dirbtinio intelekto klasifikatorių kontekste, ansamblinio klasifikavimo metodai, kaip XGB pasižymi aukštesniu našumu (Špicas ir kt., 2023). XGB ansamblinio klasifikavimo metodas taip pat pasižymi aukštesniu našumu ir kitų šiai grupei priskiriamų metodų atžvilgiu (Kang ir kt., 2021; Suhadolnik ir kt., 2023). Pavyzdžiui, XGB pranašumas prieš RF klasifikavimo metodą yra pagrįstas metodų technikos skirtumais, kai RF naudoja maišymą, kad vienu metu sukurtų visus sprendimų medžius, o XGB sukuria sprendimų medžių grupes, naudodamas gradiento didinimo algoritmą, kad sukurti medžiai nuosekliai mažintų neatitikimus (Špicas ir kt., 2023). Taip pat autoriai išskiria, kad stiprinimo metodai pasižymi ir aukštesniu našumu ir nesubalansuotų duomenų atžvilgiu (Shen ir kt., 2019). Shen ir kt. (2019) atliktas tyrimas nesubalansuotiems duomenims atskleidžia ansamblinio klasifikavimo metodo, tiksliau AdaBoost plėtinio, paremtu atgalinio sklidimo neuroninių tinklų algoritmu, aukštesnį našumą, lyginant su kitoms klasifikavimo grupėms priskiriamų metodų rezultatais. Tokį patį rezultatą atskleidžia ir Shen ir kt. (2021) atliktas empirinis tyrimas, kai sukurtas naujas

ansamblinio klasifikavimo metodas, pagrįstas AdaBoost ir LSTM metodais, pasižymi aukštesniu našumu, lyginant su kitais klasifikavimo metodais, priskirtais kitoms grupėms. Visgi, iš atliktos mokslinės literatūros analizės, įprastas AdaBoost klasifikavimo metodas pasižymi žemesniu našumu nei kiti stiprinimo metodui priskiriami klasifikatoriai (Suhadolnik ir kt., 2023; Z. Tian ir kt., 2020). Tian ir kt. (2020) atliktas empirinis tyrimas kredito rizikos vertinimui, pagrindžia GBDT metodo aukštesnį našumą bei pačio modelio pranašumą prieš AdaBoost ir kitoms grupėms priskiriamus klasifikavimo metodus. Visgi, autorių išskiriamas esminis trūkumas šiai klasifikavimo grupei priskiriamų metodų atžvilgiu, yra susijęs su žemu paaiškinamumu ir sudėtingesne technine integracija (Zhang & Yu, 2024).

Kiekvienas iš aptartų klasifikavimo metodų pasižymi privalumais ir trūkumais (žr. 3 lentelę). Visgi, žemas klasifikavimo metodų paaiškinamas yra esminė problema, apribojanti autorius taikyti sudėtingesnėmis metodikomis pagrįstus klasifikavimo metodus kredito rizikos vertinimo srityje. Žemas paaiškinamumas, kuris yra būdingas ne tik ansamblinio klasifikavimo metodų grupei priskiriamiems metodams, bet tuo pačiu ir dirbtinio intelekto klasifikavimo metodų grupei, sukelia iššūkių šiuos metodus taikyti praktikoje. (Dėl prudencinių reikalavimų kredito įstaigoms ir investicinėms įmonėms ir kuriuo iš dalies keičiamas Reglamentas (ES) Nr. 648/2012, 2013) atskleidžiama, kad įstaigoms, vertinančioms kredito riziką, metodų pasirinkimo pagrįstumas turėtų būti susijęs su aukštomis paaiškinimo galimybėmis. Visgi, sudėtingesnėmis technikomis paremtų mašininio mokymo metodai pasižymi ribotumu, siekiant atskleisti atskirų veiksnių ar jų sąveikos svarbą. Šiai problemai eliminuoti ir suteikti galimybę įmonėms vertinti kredito riziką aukštesniu tikslumu pasižyminčiais metodais, mokslinėje literatūroje atsirado XAI (angl. eXplainable AI) koncepcija, pristatanti mašininio mokymosi metodų rinkinį, kuriuo remiantis yra užtikrinamas modelių paaiškinamumas (Gramegna & Giudici, 2021). XAI koncepcija pagrįsti metodai yra tinkami atskirų veiksnių ar jų sąveikų svarbai nustatyti (Malakauskas, 2023). Mokslinėje literatūroje plačiai taikomos ir pripažįstamos dvi paaiškinimo technikos, kaip SHAP (angl. SHapley Additive ExPlanations) ir LIME (angl. Local Interpretable Model-Agnostic Explanations) (Dastile & Celik, 2021; Gramegna & Giudici, 2021; Malakauskas, 2023). Iš atliktos mokslinės literatūros analizės, susijusios su XAI koncepcija, SHAP metodo pritaikomumas pasižymėjo aukštesniu našumu ir tuo pačiu didesne diskriminantine galia nei LIME metodas, tiek ir XGB, tiek ir RF klasifikavimo metodų atžvilgiu (Dastile & Celik, 2021; Gramegna & Giudici, 2021). Šių metodų integravimas į kredito rizikos vertinimą, užtikrintų ne tik vertinimo tikslumą, bet ir netiesinės priklausomybės ir kintamųjų sąveikos nustatymą, kuris yra pagrindas teisingą sistemą formuojančių institucijų rekomendacijose kreditą teikiančioms įstaigoms.

3 lentelė

Klasifikavimo metodų privalumų ir trūkumų mokslinės literatūros apibendrinimas

Grupė	Metodas	Privalumai	Trūkumai
Tradiciniai	LR	Aukštas pritaikomumo paprastumas ir paaiškinamumas	Jautrumas multikolinearumui, Tiesiniai ryšys tarp nepriklausomų kintamųjų ir logaritminių šansų Jautrumas duomenų kokybei, pokyčiams duomenyse Jautrumas nesubalansuotoms grupėms
	LDA	Aukštas pritaikomumo paprastumas ir paaiškinamumas	Normalus pasiskirstymas
	k-NN	Aukštas pritaikomumo paprastumas ir paaiškinamumas	Didelių sąnaudų reikalaujantis Jautrumas sudėtingų duomenų rinkinių atžvilgiu Jautrumas nesubalansuotoms grupėms
	NB	Aukštas pritaikomumo paprastumas ir paaiškinamumas	Kintamieji yra nepriklausomi Kokybiniais kintamiesiems būdingas normalusis skirstinys
	DT	Aukštas paaiškinamumas	Nestabilus, pakeitus duomenis (GBDT tai išsprendžia) Jautrumas nesubalansuotoms grupėms
Dirbtinio intelekto	ANN	Aukštas našumas, Aukštas imlumas sudėtingų sąveikų atžvilgiu	„juodoji dėžė“ – žemas paaiškinamumas

Grupė	Metodas	Privalumai	Trūkumai
	CNN	Aukštas našumas, Aukštas imlumas sudėtingų sąveikų atžvilgiu	„juodoji dėžė“ – žemas paaiškinamumas
	SVM	Aukštas našumas, Aukštas imlumas tvarkytis su didelio matmens duomenų rinkiniais, Nejautrus persimokymui	Jautrumas mokymosi parametrų nustatymui Jautrumas sudėtingų duomenų rinkinių atžvilgiu
	RF	Aukštas našumas	Įvesties kintamųjų koeficientai nėra pastovūs Galimi daug panašių sprendimų medžių, kiek gali uždengti tikslus rezultatus
	CART	Aukštas našumas	Polinkis į persimokymą, Jautrumas duomenų pasikeitimams, Jautrumas duomenų kokybei
	LSTM (ilgalaikės trumpalaikės atminties tinklas)		Žemas paaiškinamumas
Ansamblinio	Ada Boost	Aukštas našumas	Jautrus išoriniams duomenims, todėl ir nestabilus prognozavimo rezultatų tikslumas
	XGB	Aukštas našumas, efektyvumas, esant ir sudėtingoms duomenų priklausomybėms	Žemas paaiškinamumas, jautrumas persimokymui, jei duomenys nėra kokybiški ir imtis yra nedidelė, Reikalingos transformuotos kintamųjų formas,
	GBDT	Aukštas tikslumas ir efektyvumas	Reikalinga tinkamų metodo parametrų paieška Žemas paaiškinamumas, Didelės laiko sąnaudos

Grupė	Metodas	Privalumai	Trūkumai
	Bagging		

Šaltinis: sudaryta autorės, remiantis nagrinėta mokslinė literatūra (Bello, 2023; Dastile ir kt., 2020; Dzidzevičiūtė, 2013; Kang ir kt., 2021a; Shen ir kt., 2021; Suhadolnik ir kt., 2023; Špicas ir kt., 2023; Zhang & Yu, 2024)

Taigi, šių dienų mokslinėje literatūroje kredito rizikos vertinimui taikomi klasifikavimo metodai yra skirstomi į tradicinių, dirbtinio intelekto ir ansamblinio klasifikavimo metodų grupes. Atliktų empirinių tyrimų rezultatai atskleidžia ansamblinio klasifikavimo metodų, ypač stiprinimo metodu grupei priskiriamų, aukštesnį tikslumą nei kitų klasifikavimo metodų grupių atžvilgiu. Nors tradiciniai klasifikavimo metodai yra mažiau tikslūs, pastarieji atlieka lyginamąją funkciją empiriniuose tyrimuose. Tradicinių klasifikavimo metodų įtraukimas yra siejamas ir su sudėtingesnėmis metodikomis pagrįstų klasifikavimo metodų žemomis paaiškinimo galimybėmis. Visgi, XAI koncepcijos atsiradimas mokslinėje literatūroje sprendžia žemą metodų paaiškinimo galimybių spragą, o ši mokslinė plėtra turėtų užtikrinti aukštesnėmis paaiškinimo galimybėmis pagrįstų klasifikavimo metodų įtraukimą į kredito rizikos vertinimą realioje aplinkoje.

1.2.3 Kredito rizikos vertinimo modelio įvertinimui taikytų metodų teoriniai aspektai

Kredito rizikos vertinimo modelio sudarymui, jo vertinimui ir rangų skalės sudarymui, reikšmingą įtaką turi nustatytas lūžio taškas (angl. threshold), kuriuo remiantis yra atmetamos skolininkų paraiškos. Kitaip, lūžio taškas atspindi nemokumo tikimybės toleranciją. Šiam taškui nustatyti, mokslinėje literatūroje, nagrinėjančioje kredito rizikos vertinimo sritį, pateikiami metodai, susiję su: 1) analizuojamu faktinių „blogų“ skolininkų dažnio priklausomybe nuo paraiškų atmetimo lygio; 2) kaupiamojo faktinių „gerų“ skolininkų skaičiaus pokyčio ir kaupiamojo faktinių „blogų“ skolininkų skaičiaus pokyčio santykiu, kai šis yra nuo 5:1 iki 3:1; 3) portfelio grynosios dabartinės vertės maksimizavimu; 4) maksimaliu tikėtinu pelnu (angl. expected maximum profit) (toliau – EMP) (Dzidzevičiūtė, 2013; Špicas, 2017; Verbraken ir kt., 2013). EMP metodas yra AUC kreivės alternatyva, kuriuo remiantis yra nustatoma optimali skolininkų atmetimo proporcija, kuri maksimizuotų galimą pelną (Špicas, 2017).

Lūžio taško nustatymas yra susijęs su gauto kredito rizikos vertinimo modelio įvertinimu, t.y. galimi du modelio įvertinimo būdai: 1) atsižvelgiant į lūžio tašką; 2) neatsižvelgiant į lūžio tašką (Špicas, 2017). Pagal atliktos mokslinės literatūros analizę ir išnagrinėtus taikomus modelio įvertinimo metodus buvo sudaryta vertinimo metodų pasiskirstymo lentelė (žr. 4 lentelę). Iš atliktos mokslinės literatūros analizės pastebima, kad į lūžio tašką atsižvelgiančių metodų

pasirinkimas yra susijęs su teisingo klasifikavimo matrica (angl. classification confusion matrix) ir iš jos išplaukiančiais klasifikavimo tikslumo rodikliais: bendro tikslumo rodiklis (toliau – Ar), teisingo klasifikavimo rodiklis (toliau – CCR), klaidingo klasifikavimo rodiklis (toliau – MCR), I rūšies klaida (toliau – α), II rūšies klaida (toliau – β), modelio jautrumas (toliau – Se), modelio specifiškumas (toliau – Sp), įsipareigojimų nevykdymo prognozės tikslumas (toliau – PPV), „gerų“ skolininkų klasifikavimo tikslumas (toliau – NPV), F1 įvertis, atspindintis jautrumo ir įsipareigojimų nevykdymo prognozės tikslumą, G vidurkis, atskleidžiantis modelio jautrumo ir specifiškumo subalansuotumą, bei apytikslės koreliacijos koeficientas (toliau – ACP). Taip pat į lūžio tašką atsižvelgiantys metodai yra IBA, kuriuo atsižvelgiama į nesubalansuotų klasių poveikį modelio našumui bei tikslumo ir atšaukimo kreivė (angl. precision recall curve) ir esantis plotas po ja (toliau – AP) atspindi modelio našumą (Malakauskas, 2023; Xiao ir kt., 2021). Nagrinėtoje mokslinėje literatūroje modelio įvertinimui dažniausiai naudojamas metodas, nesusijęs su lūžio tašku yra ROC (angl. receiver operator characteristics) kreivė, kuri taikoma rezultatų tikslumui įvertinti, sprendžiant dvinarę klasifikavimo problemą (Mileris, 2009). Esant visiškai teisingam klasių klasifikavimui, ROC kreivė turėtų pasiekti viršutinį kairįjį kampą ir tokiu atveju modelio jautrumas būtų maksimalus (Mileris, 2009). AUC (angl. area under curve) yra ROC kreivės skaitinė reikšmė, t.y. plotas, esantis po ROC kreive. Modelio efektyvumas vertinamas AUC reikšmės intervale nuo 0.5 (atitinkančio naivų modelį) iki 1 (atitinkančio idealų modelį) (Mileris, 2009). Kitas metodas, kuris yra rečiau taikomas, tačiau taip pat yra reikšmingas – Kolmogorovo-Smirnov (toliau – KS) – kuriuo siekiama įvertinti modelio gebėjimą atskirti įsipareigojimus vykdančius nuo nevykdančių skolininkų (Shen ir kt., 2021).

4 lentelė

Modelio įvertinimo metodų pasiskirstymo mokslinėje literatūroje apibendrinimas

Metodas	Tyrimai
<i>Metodai, atsižvelgiant į lūžio tašką</i>	
Teisingo klasifikavimo matrica	Byanjankar ir kt., 2015; Boguslauskas ir kt., 2011; Boguslauskas & Mileris, 2009; Dastile & Celik, 2021; Dzidzevičiūtė, 2013; Kang ir kt., 2021; Z. Li ir kt., 2017; Malakauskas, 2023; Mileris, 2009; Shen ir kt., 2019; Song ir kt., 2023; Suhadolnik ir kt., 2023; Špicas, 2017; Špicas ir kt., 2023; Y. Tian ir kt., 2018; Z. Tian ir kt., 2020; Xiao ir kt., 2021
Klasifikavimo tikslumo rodikliai: Ar, CCR, MCR, Se, Sp, BAC, MCC, PPV, NPV, F, G-vidurkis, ACP	Boguslauskas ir kt., 2011; Boguslauskas & Mileris, 2009; Dastile & Celik, 2021; Kang ir kt., 2021; B. Li, 2022; Malakauskas, 2023; Mileris, 2009; Shen ir kt., 2019; Song

Metodas	Tyrimai
CAP ir GINI indeksas	ir kt., 2023; Suhadolnik ir kt., 2023; Špicas, 2017; Špicas ir kt., 2023; Z. Tian ir kt., 2020; Xiao ir kt., 2021
IBA	Dzidzevičiūtė, 2013; Špicas, 2017
<i>Metodai, neatsižvelgiant į lūžio tašką</i>	Xiao ir kt., 2021
AUC/ROC	Byanjankar ir kt., 2015; Boguslauskas & Mileris, 2009; Dastile & Celik, 2021; Dzidzevičiūtė, 2013; Kang ir kt., 2021; Z. Li ir kt., 2017; Malakauskas, 2023; Mileris, 2009; Shen ir kt., 2019, 2021; Suhadolnik ir kt., 2023; Špicas, 2017; Špicas ir kt., 2023; Y. Tian ir kt., 2018; Z. Tian ir kt., 2020; Verbraken ir kt., 2013; Xiao ir kt., 2021
Rrecision-Recall kreivė	Malakauskas, 2023
KS	Kang ir kt., 2021; Shen ir kt., 2021
Šaltinis: sudaryta autorės, remiantis atlikta mokslinės literatūros analize	

Pastebima, kad dauguma mokslinių tyrimų autorių renkasi daugiau nei vieną vertinimo metodą iš skirtingų vertinimo grupių lūžio taško atžvilgiu. Svarba taikyti tiek ir į lūžio tašką atsižvelgiančius vertinimo metodus, tiek ir neatsižvelgiančius, dėl ribinių verčių ir duomenų pasiskirstymo tarp klasių priklausymo atitinkamai (Markov ir kt., 2022).

2. KREDITO RIZIKOS VERTINIMO MODELIO FIZINIAMS ASMENIMS SUDARYMO METODIKA

Ižanga. Finansų rinkose kredito rizikos vertinimas yra neatsiejama veiklos dalis, ypačingai kredito įstaigoms, kurioms šis procesas yra privalomas pagal reguliavimo reikalavimus. Siekiant priimti pagrįstus sprendimus dėl paskolos suteikimo, reguliavimo reikalavimuose yra apibrėžiamas kiekybiškai pagrįstų metodų taikymas kredito rizikos vertinimui. Šiuo atžvilgiu aktualus tampa internetinių paskolų palyginimo ir tarpininkavimo (toliau – IPPT) platformų vaidmuo, kurios siekis padėti tiek atskiroms kredito įstaigoms, tiek visai rinkai efektyviai valdyti kredito riziką. Atsižvelgus į tai, šiame moksliniame tyrime sprendžiama problema yra mašininio mokymu paremtu fizinių asmenų kredito rizikos vertinimo modelio sukūrimu, kuris pasižymėtų aukštu našumu bei galėtų būti pritaikomas kredito rizikos vertinimui rinkos lygmeniu.

Tyrimo objektas – fizinių asmenų vartojimo paskolų kredito rizikos vertinimas, taikant mašininio mokymo kredito rizikos vertinimo metodus.

Tyrimo tikslas – išanalizavus kredito rizikos problematiką nagrinėjančią mokslinę literatūrą, sudaryti fizinių asmenų vartojimo paskolų kredito rizikos vertinimo modelį, remiantis mašininio mokymo metodais.

Tyrimo tikslui pasiekti keliami tokie **uždaviniai**:

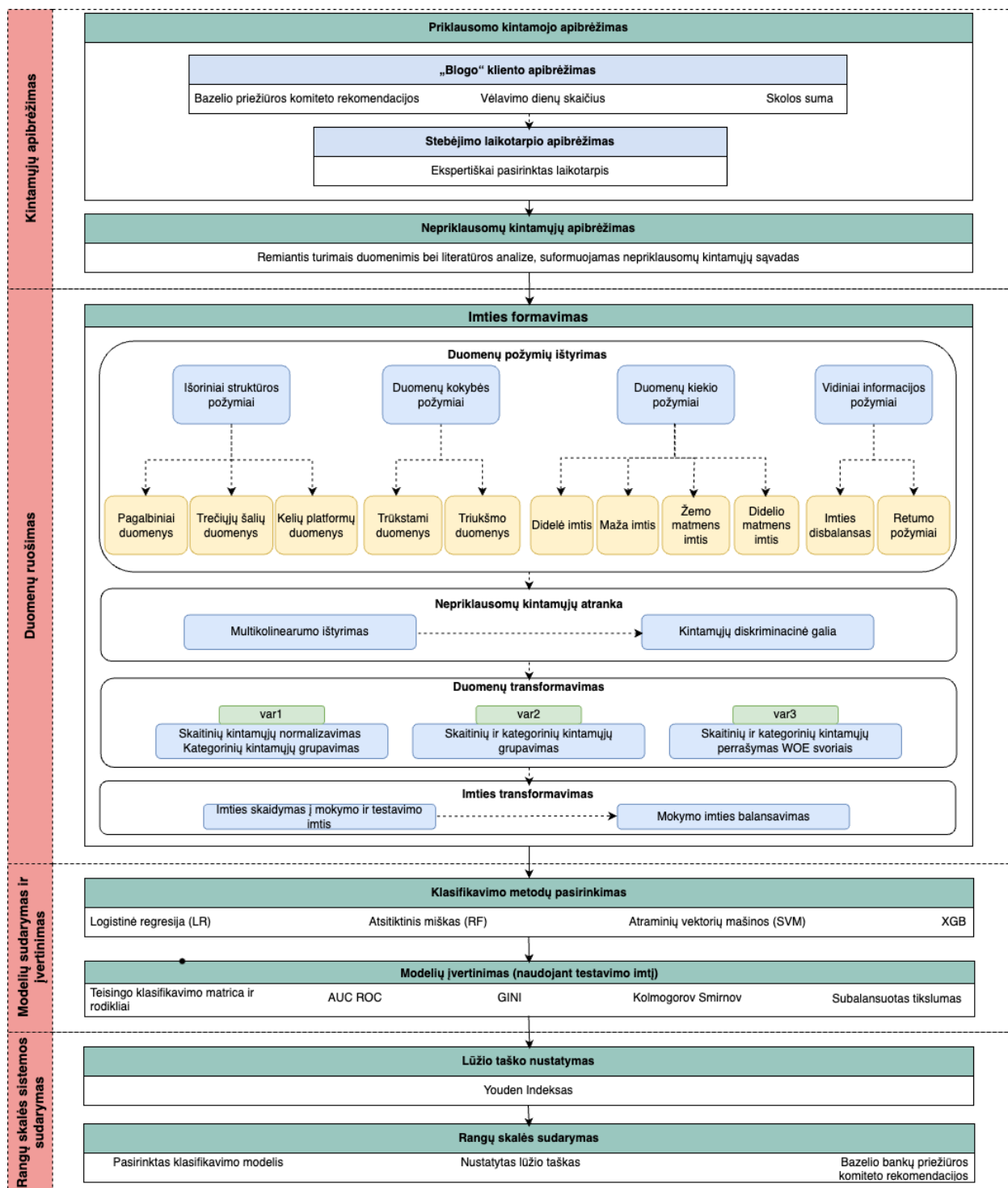
1. Apibrėžti priklausomą kintamąjį;
2. Sudaryti nepriklausomų kintamųjų sąvadą;
3. Atlikti imties formavimą, atsižvelgiant į duomenų požymių ištyrimą, nepriklausomų kintamųjų atranką bei duomenų transformavimą;
4. Remiantis atlikta mokslinės literatūros analize, pasirinkti klasifikavimo metodus, kurie bus naudojami tyrimui atlikti;
5. Atlikti modelių rezultatų įvertinimą;
6. Nustatyti lūžio tašką ir pagal gautą lūžio tašką sudaryti rangų skalės sistemą.

Apibendrinta tikslui pasiekti išsikeltų uždavinių įgyvendinimo seka pateikiama 1 paveiksle.

Tyrimo metodai. Iškeliamam tyrimo tikslui pasiekti naudojamos duomenų transformavimo technikos, imties subalansavimo technikos, duomenų analizės metodai, mašininio mokymosi klasifikavimo metodai, vertinimo metodai, reitingavimo sistema, teisės aktų analizė.

1 paveikslas

Mokslinio tyrimo atlikimo procesas



Šaltinis: sudaryta autorės

2.1 Priklausomo kintamojo apibrėžimas

Formuojant kredito rizikos vertinimo modelį, dažniausiai priklausomas kintamasis yra apibrėžiamas kaip dvinaris. Pastarojo reikšmės apima skolininko klasifikavimą į „gerą“ bei „blogą“. Kredito įstaigos atžvilgiu, „blogas“ skolininkas atitinka patiriamus kaštus, o „geras“ – naudą, todėl „blogo“ skolininko tikslus apibrėžimas yra vienas svarbiausių kredito įstaigos tikslų kredito rizikos valdyme. Be to, nuo apibrėžto „blogo“ skolininko priklauso ir kredito rizikos vertinimo modelio prognozinė galia praktikoje. Per daug konservatyvus apibrėžimas gali sumažinti kredito įstaigos potencialias pajamas, atmetant ir „gerus“ skolininkus. Priešingai, per daug nuosaikus apibrėžimas gali turėti reikšmingos įtakos kredito įstaigos rizikos išaugimui. Todėl, tarp konservatyvumo ir nuosaikumo „blogo“ skolininko apibrėžimas yra siektinas tikslas formuojant kredito rizikos vertinimo modelį.

Reikšmingą įtaką kredito įstaigos rizikos vertinimui bei valdymui turėjo įsteigta Bazelio sistema. „Bazelio III“ sistemoje nurodyta, kad kiekybiškai vertinant rizikos parametrus, įsipareigojimų nevykdantis asmuo yra laikomas tas, kuris vėlavo atlikti mokėjimą 90 dienų ir ilgiau. Šį terminą pagrindžia ir nagrinėta mokslinė kredito rizikos vertinimo literatūra. Todėl, remiantis teisės aktuose nurodyta ir mokslinėje literatūroje pritaikoma vėlavimo dienų trukme, šiame moksliniame darbe „blogas“ skolininkas yra apibrėžiamas kaip 90 dienų vėluojantis atlikti mokėjimą kredito įstaigai.

Formuojant „blogo“ skolininko apibrėžimą, yra svarbu įvertinti turimos skolos dydį. Autorės ekspertiniu vertinimu, skolos sumą būtų galima įvertinti kaip vidutiniu vieno mėnesio įsipareigojimams dengti skirtą mokėjimo sumos dydžiu. Remiantis tyrimui atlikti naudota duomenų imtimi, vidutinė vieno mėnesio paskolos įmoka siekia 136 Eur. Todėl, remiantis šiuo įverčiu bei „blogų“ skolininkų sąsaja su polinkiu skolintis mažesnėmis sumomis, skolos suma yra apibrėžiama 100 Eur dydžiu.

Kadangi kliento elgsena ir atitikimas apibrėžtiems kriterijams yra vertinimas po kredito išdavimo, yra svarbu nustatyti skolininkų stebėjimo laikotarpį. Autorės tikslas yra stebėti skolininkus bent vienerius metus, kadangi kredito rizikos vertinime vienerių metų stebėjimo horizontas yra dažniausiai taikomas laikotarpis, leidžiantis identifikuoti vėluojančius ar nemokius klientus (Markov ir kt., 2022; Tran ir kt., 2021). Atsižvelgiant į šį siekį, yra formuojamas duomenų rinkinys, kuris aptariamas šio darbo tiriamojoje dalyje.

Taigi, remiantis teisės aktais, mokslinė literatūra bei autorės vertinimu, „blogas“ skolininkas yra apibrėžiamas kaip asmuo, vėlavęs 90 dienų ir ilgiau bent 100 Eur skolos suma, atsižvelgiant bent į vienerių metų laikotarpį.

2.2 Nepriklausomų kintamųjų pasirinkimas

Kredito rizikos vertinimo modelio kokybei ir tikslumui turi reikšmingos įtakos ir pasirinktas klientą apibūrinančių veiksnių sąvadas. Remiantis atlikta mokslinės literatūros analize, išskiriamos šios pagrindinės skolininką apibūrinančių bruožų grupės, kaip: demografinės, elgsenos, finansinės bei paskolos charakteristikos. Šios charakteristikos kredito rizikos vertinimą apimančiuose mokslinės literatūros rezultatuose pasižymi kaip reikšmingą įtaką turinčios kredito rizikos vertinimo modeliavimui. Todėl, remiantis šia informacija bei turimais duomenimis, buvo suformuotas nepriklausomų kintamųjų sąvadas, apimantis skolininko demografinės, elgsenos, finansinės bei paskolos charakteristikas (žr. 5 lentelę).

Demografinės charakteristikos. Atsižvelgiant į reikšmingą informaciją apie skolininko kredito riziką savyje laikančiose demografinėse charakteristiose, į sąvadą buvo įtraukti tokie nepriklausomi kintamieji, kaip skolininko lytis (Gender), turimų nepilnamečių vaikų skaičius (Children), tapatybės patvirtinimo dokumento tipas (Doc_type), deklaruota gyvenamoji vieta (App_Applicant_City_fixed). Taip pat buvo sudaryti sudurtinių kintamųjų komponentai, kaip skolininko tipas, atsižvelgiant į paskolos paskirtį bei šeimyninę padėtį (App_type) bei amžiaus projekcija (Age_proj). Amžiaus kintamojo modifikacija buvo sudaryta siekiant apimti skolininką apibūdinančią situaciją, galinčią pakeisti jo pajėgumą atlikti mokėjimus laiku.

Elgsenos charakteristikos. Apie skolininko kredito riziką gali apibūdinti ir jo elgseną apibūrinantys veiksniai. Tokie veiksniai, kaip paraiškos pildymas darbo valandomis (Working_hours), įvairaus laikotarpio atžvilgiu skolininko pildytų paraiškų skaičius gali atskleisti skolininko polinkį į didesnę riziką. Priešingai, tai, kad klientas yra naujas (New_client) arba jam ankstesnės paraiškos pildymo metu buvo išduota paskola (Reissuance), gali atskleisti skolininko polinkį į mažesnę riziką. O skolininkai, kurių tapatybės patvirtinimo metodas (Auth_meth) atitinka laikinąjį leidimą gyventi Lietuvoje, gali turėti polinkį į didesnę riziką.

Finansinės charakteristikos. Finansinėmis charakteristikomis galima numatyti skolininko pajėgumą atlikti prisiimtus įsipareigojimus. Šią informaciją atskleidžia ne tik jo turima kredito istorija (esamų, buvusių skolų suma bei skolų apmokėjimo trukmė įvairiais laikotarpiais), bet ir jo jau turimų finansinių įsipareigojimų santykis su tvariomis pajamomis (DSTI), kur tvarios pajamos atitinka reguliarių, prognozuojamų ir ilgalaikių pajamų apibrėžtį, kaip, pavyzdžiui, pajamos gaunamos iš nuolatinių darbo santykių. Skolininko turimus įsipareigojimus (Oblig) galima vertinti pagal skolininko nurodytą informaciją arba remiantis paskolų rizikos duomenų baze (toliau – PRDB). Kadangi ne visiems duomenų imtį sudarantiems skolininkams buvo tikrinami PRDB duomenys, įsipareigojimus atspindintis kintamasis yra priklausomas nuo to, kuris sumos dydis yra didesnis, jei abu duomenų šaltiniai buvo prieinami, kitais atvejais – ką nurodė

skolininkas. Priešingai nei įsipareigojimų atveju, informacija apie kliento pajamas buvo prieinama tiek iš jo nurodytų duomenų, tiek iš Sodros registro. Skolininko pajamas atspindintis kintamasis (Income) yra apibrėžiamas kaip pajamų sumos dydis, atsižvelgiant į abu prieš tai išvardintus duomenų šaltinius, priklausomai kuris sumos dydis yra mažesnis. Skolininko įsipareigojimų ir pajamų vertinimas yra pagrįstas Lietuvos banko parengtomis vartojimo kredito teikimo gairėmis (Vartojimo kredito teikimo gairės, 2023). Kadangi skolininko pajamas galima apibrėžti tiek oficialiais, tiek ne oficialiais duomenimis, buvo sudarytas kintamasis, kuris apibrėžia skirtumą tarp kliento nurodytų pajamų ir Sodroje esamų duomenų arba kliento nurodytų pajamų, priklausomai kuris pajamų dydis yra mažesnis (Income_diff). Autorės manymu, tai svarbus skolininko elgseną apibrėžiantis kintamasis, todėl pastarasis buvo įtrauktas į nepriklausomų kintamųjų sąvadą.

Paskolos charakteristikos. Skolininko kredito rizikos pasiskirstymas gali būti reikšmingai besiskiriantis tarp skirtingų produkto tipų (Product_name) ir paskolos sumos dydžių (Loan_Granted_Amount), todėl kintamieji yra įtraukti ir į tolimesnį kredito rizikos vertinimo modeliavimą. Duomenų imtis yra suformuota remiantis vartojimo paskolomis, kurios apima tokius produkto tipus, kaip: vartojimo paskola, vartojimo kredito linija, paskola automobiliui, autolizingas, paskola būsto remontui bei refinansavimas. Šių produktų paskolos sumos rėžis duomenų imtį formuojančio laikotarpio atžvilgiu yra 100 Eur – 25 tūkst. Eur.

5 lentelė

Nepriklausomų kintamųjų sąvadas

Kintamųjų grupė	Kintamojo trumpinys	Duomenų tipas	Aprašymas
Demografinė	Age_proj	Numeric	Pareiškėjo prognozuojamas amžius, atsižvelgiant į paskolos terminą.
	Doc_type	String	Pareiškėjo tapatybės patvirtinimo dokumento tipas (pasas, kortelė, laikinas leidimas, nuolatinis leidimas).
	App_type	String	Paraiškos tipas pagal paskolos paskirtį (asmeninėms ar šeimos reikmėms) ir pareiškėjo šeimyninę padėtį (vedęs, išsiskyręs, našlys ar nevedęs).
	Gender	String	Pareiškėjo lytis.
	Children	Integer	Pareiškėjo turimų vaikų iki 18 metų amžiaus skaičius.
	App_Applicant_City_fixed	String	Pareiškėjo deklaruota gyvenamoji vieta.

Kintamųjų grupė	Kintamojo trumpinys	Duomenų tipas	Aprašymas
Finansinė	Main_income_dur	Numeric	Pareiškėjo darbo trukmė mėnesiais pagrindinėje darbovietėje. Pagrindinė darbovietė yra ta, kurioje dirbama ilgiausiai.
	Income_source	Integer	Pareiškėjo nurodytas pagrindinis pajamų šaltinis.
	Workplaces_count	Numeric	Pareiškėjo esamų darboviečių skaičius.
	Avg_working_period	Numeric	Vidutinė darbo trukmė mėnesiais buvusiose darbovietėse.
	Job_search_dur	Numeric	Dienų skaičius nuo buvusios darbovietės pabaigos iki esamos darbovietės pradžios datų.
	Income	Integer	Pareiškėjo pajamų dydis, paremtas tvariomis arba kliento nurodytomis pajamomis, priklausomai kuris dydis buvo mažesnis.
	Income_diff	Integer	Skirtumas tarp tvarių pajamų arba kliento nurodytų pajamų (priklausomai, kuris dydis buvo mažesnis) ir kliento nurodytų pajamų.
	Oblig	Integer	Mėn. įmokos dydis įsipareigojimams dengti, paremtas PRDB duomenimis arba kliento nurodytu dydžiu. Esant duomenims abiejų šaltinių atveju, remiamasi didesniu dydžiu.
	DSTI	Integer	Pareiškėjo DSTI rodiklis.
	Has_open_debt	Boolean	Ar pareiškėjas turi nepadengtų skolų?
	Open_debt_count	Integer	Pareiškėjo nepadengtų skolų skaičius.
	Open_debt_sum	Numeric	Pareiškėjo nepadengtų skolų suma.
	Open_debt_dur	Numeric	Nepadengtos skolos trukmė nuo anksčiausiai atsiradusios nepadengtos skolos.
	Debts_Count	Integer	Pareiškėjo padengtų ir nepadengtų skolų skaičius.
	Debts_sum	Float	Padengtų skolų suma paraiškos pateikimo datai.
	Debts_sum_filter	Float	Padengtų skolų suma paraiškos pateikimo datai, kai skolos tipas susijęs su priimtais finansiniais įsipareigojimais: Paskola/Kreditas, Finansinės institucijos ar Lizingas.
	Debts_Sum_3y	Float	3 metų padengtų skolų suma paraiškos pateikimo datai.
	Debts_Sum_3y_filter	Float	3 metų padengtų skolų suma paraiškos pateikimo datai, kai skolos tipas susijęs su priimtais finansiniais įsipareigojimais:

Kintamųjų grupė	Kintamojo trumpinys	Duomenų tipas	Aprašymas
			Paskola/Kreditas, Finansinės institucijos ar Lizingas.
	Debts_Sum_5y	Float	5 metų padengtų skolų suma paraiškos pateikimo datai.
	Debts_Sum_5y_filter	Float	5 metų padengtų skolų suma paraiškos pateikimo datai, kai skolos tipas susijęs su prisiimtais finansiniais įsipareigojimais: Paskola/Kreditas, Finansinės institucijos ar Lizingas.
	Max_closed_debt_dur	Numeric	Maksimali gražintos skolos padengimo trukmė.
	Max_closed_debt_dur_filter	Numeric	Maksimali skolos padengimo trukmė, kai skolos tipas susijęs su prisiimtais finansiniais įsipareigojimais: Paskola/Kreditas, Finansinės institucijos ar Lizingas.
	Max_closed_debt_3y_dur	Numeric	Maksimali skolos padengimo trukmė per 3 metus paraiškos pateikimo datai.
	Max_closed_debt_3y_dur_filter	Numeric	Maksimali skolos padengimo trukmė per 3 metus paraiškos pateikimo datai, kai skolos tipas susijęs su prisiimtais finansiniais įsipareigojimais: Paskola/Kreditas, Finansinės institucijos ar Lizingas.
	Max_closed_debt_5y_dur	Numeric	Maksimali skolos padengimo trukmė per 5 metus paraiškos pateikimo datai.
	Max_closed_debt_5y_dur_filter	Numeric	Maksimali skolos padengimo trukmė per 5 metus paraiškos pateikimo datai, kai skolos tipas susijęs su prisiimtais finansiniais įsipareigojimais: Paskola/Kreditas, Finansinės institucijos ar Lizingas.
Paskolos	Product_name	String	Produkto klasifikatoriaus reikšmė. Galimos klasifikatoriaus reikšmės: V – vartojimo paskola, VK – vartojimo kredito linija, PA – paskola automobiliui, PB – paskola būsto remontui, A – autolizingas, R – refinansavimas.
	Loan_granted_amount	Numeric	Paskolos suma.
Elgsenos	Working_hours	Boolean	Ar pareiškėjas pateikė paraišką darbo valandomis?
	Reissuance	Boolean	Ar pareiškėjui buvo išduota paskola anksčiau?
	New_client	Boolean	Ar pareiškėjas pirmą kartą užpildė paraišką?

Kintamųjų grupė	Kintamojo trumpinys	Duomenų tipas	Aprašymas
	Auth_meth	String	Tapatybės patvirtinimo metodo grupė (Smart-ID, el.parašas, 1ct pavedimas).
	Total_apps_count	Numeric	Pareiškėjo užpildytų paraiškų skaičius nuo 2020-01-01.
	Apps_count_last_30d	Numeric	Pareiškėjo užpildytų paraiškų skaičius per 30 dienų.
	Apps_count_last_60d	Numeric	Pareiškėjo užpildytų paraiškų skaičius per 60 dienų.
	Apps_count_last_90d	Numeric	Pareiškėjo užpildytų paraiškų skaičius per 90 dienų.
	Apps_count_last_180d	Numeric	Pareiškėjo užpildytų paraiškų skaičius per pusę metų.
	Apps_count_last_360d	Numeric	Pareiškėjo užpildytų paraiškų skaičius per metus.
	Days_after_return_last	Numeric	Dienų skaičius nuo paskutinės pareiškėjo pildytos paraiškos.
	Days_after_return_min	Numeric	Minimalus dienų skaičius tarp pareiškėjo pildytų paraiškų.
	Days_after_return_max	Numeric	Maksimalus dienų skaičius tarp pareiškėjo pildytų paraiškų.
	Days_after_return_avg	Numeric	Vidutinis dienų skaičius tarp pareiškėjo pildytų paraiškų.
	CI_Quer_by_30d	Numeric	Užklausų skaičius apie asmens kreditingumą per 30 dienų.
	CI_Quer_by_90d	Numeric	Užklausų skaičius apie asmens kreditingumą per 90 dienų.
	CI_Quer_by_180d	Numeric	Užklausų skaičius apie asmens kreditingumą per 180 dienų.

Šaltinis: sudaryta autorės, remiantis mokslinės literatūros analize ir turimais duomenimis

Taigi, remiantis atlikta mokslinės literatūros analize bei ekspertiniu vertinimu, buvo suformuotas nepriklausomų kintamųjų sąvadas, kuriuo buvo siekiama apimti kuo didesnę informacijos kiekį dėl kredito rizikos vertinimo modelio rezultatų kokybiškumo. Tolimesniuose modelio sudarymo etapuose nepriklausomų kintamųjų sąvadas bus vertinimas skirtingais metodais, todėl galutinis nepriklausomų kintamųjų skaičius sąvade bus mažesnis.

2.3 Imties sudarymas

Suformavus pradinį duomenų rinkinį iš priklausomo ir nepriklausomų kintamųjų, svarbu atlikti galutinio duomenų rinkinio išgryninimą. Remiantis atlikta mokslinės literatūros analize, galutinis imties suformavimas turi reikšmingos įtakos modelio rezultatams. Todėl, pirmu etapu yra atliekamas duomenų požymių ištyrimas, tuomet seka nepriklausomų kintamųjų atranka ir galiausiai paskutiniame etape duomenys yra transformuojami. Dažnai mokslinės literatūros autoriai į kredito rizikos modeliavimą įtraukia ir atmetas paraiškas. Tačiau, dėl aiškaus atskaitos stebėjimo taško nebuvimo, kaip, kad išduotų paskolų atveju, ekspertiniu požiūriu buvo nuspręsta naudoti tik išduotas paskolas imties formavimui. Todėl, suformuotas imties formavimo procesas atitinka anksčiau įvardintą veiksmų seką.

Duomenų požymių ištyrimas. Remiantis (Zhang & Yu, 2024) apibendrinta duomenų požymių ištyrimo kryptimis, formuojant imtį kredito rizikos vertinimo modeliavimui, yra taikomas duomenų kokybės, duomenų kiekio, išorinių struktūros ir vidinių informacijos požymių ištyrimas.

Siekiant įvertinti, ar turimi duomenys yra kokybiški, atsižvelgiama, ar duomenyse yra trūkstančių reikšmių bei išskirčių. Išskirčių įvertinimui yra naudojamos duomenų vizualizavimo technikos, kaip stačiakampių diagramų (angl. box plot) formavimas kiekvieno kintamojo atžvilgiu. Imties formavimui remiamasi ir duomenų kiekio požymių atitikimo vertinimu. Šio magistrinio darbo 1.2.1 dalyje įvardijamas siektinas teorinis imties dydis, kuris atitiktų 1500-5000 „gerų“ ir „blogų“ skolininkų, tampa ir šio darbo imties formavimo atskaitos tašku.

Kredito rizikos vertinimo modelio sudarymui pasitelkiami ne tik vidiniai platformos duomenys, bet ir išoriniai duomenys. Išorinių duomenų integravimas į modeliavimą atitinka mokslinėje literatūroje apibrėžiamą išorinių struktūrinių požymių analizę. Šio tyrimo lygmenyje naudojami išoriniai duomenys, gaunami iš Sodros, PRDB, Nepadengtų skolų ir mokėjimų registro ir Creditinfo duomenų bazių.

Mokslinė literatūra, nagrinėjanti kredito rizikos vertinimo modeliavimą, išskiria vidinių informacijos požymių ištyrimo svarbą. Vienas iš tokių požymių – imties disbalansas. Siekiant eliminuoti imties šališkumo atsiradimą, yra naudojamos imties balansavimo technikos – ROS, SMOTE, RUS. Pasirinkimas tarp didinimo ir mažinimo technikų, ekspertiniu vertinimu, turėtų būti pagrįstas svarbios informacijos apie skolininkus išlaikymu. Todėl, sprendimas turi būti priimtas remiantis mažumos imties dydžio atžvilgiu. Jeigu imties mažuma atitinka minimalų įvardintą teorinį imties dydį, turi būti naudojamos tik didinimo technikos, siekiant neprarasti

svarbios informacijos apie imties daugumą. Priešingu atveju, yra priimtinos tiek didinimo, tiek mažinimo balansavimo technikos.

Nepriklausomų kintamųjų atranka. Atlikus duomenų požymių ištyrimą, vykdoma nepriklausomų kintamųjų atranka iš sudaryto galimų nepriklausomų kintamųjų sąvado. Remiantis atlikta mokslinės literatūros analize, atrankos įvykdymui pirmiausiai taikomas multikolinearumo problemos ištyrimas, naudojant duomenų vizualizavimo technikas, kaip koreliacijos matrica. Esant 0.7 ir didesniai tarpusavio nepriklausomų kintamųjų koreliacijos įverčiui, šalinamas mažiau reikšmingas nepriklausomas kintamasis, priklausomo kintamojo atžvilgiu (sprendimas pagrįstas moksline literatūra, aptarta šio magistrinio darbo 1.2.1 dalyje). Taip pat multikolinearumo problemos ištyrimo atveju yra naudojamas VIF įvertis. Esant didesniai nei 4 VIF įverčiui, nepriklausomas kintamasis turėtų būti pašalintas iš imties (sprendimas pagrįstas moksline literatūra, aptarta šio magistrinio darbo 1.2.1 dalyje). Atlikus šalinimą ir sekančiu etapu formuojant nepriklausomų kintamųjų sąvadą, vertinama ir kintamųjų diskriminacinė galia. Šiuo atveju yra pasitelkiamas IV metodas. Esant žemesnei nei 0.02 IV kintamojo atžvilgiu, pastarąjį rekomenduojama šalinti iš imties (sprendimas pagrįstas moksline literatūra, aptarta šio magistrinio darbo 1.2.1 dalyje). Galutinis nepriklausomų kintamųjų sąvadas yra pagrindžiamas šalinamąja logistine regresija.

Duomenų transformavimas. Atlikus duomenų požymių ištyrimą ir nepriklausomų kintamųjų atranką, suformuotam galutiniam nepriklausomų kintamųjų rinkiniui yra pritaikomos duomenų transformavimo technikos. Remiantis atlikta moksline literatūros analize ir ekspertiniu vertinimu, buvo sudaryti trimis skirtingais transformavimo metodais pagrįsti imties variantai (žr. 6 lentelė). Šis sprendimas buvo priimtas ir siekiant sudaryti našiausią kredito rizikos vertinimo modelį turimų duomenų atžvilgiu.

Pirmasis imties transformavimo variantas. Pirmuoju imties transformavimo (toliau – var1) atveju skaitiniai nepriklausomi kintamieji yra normalizuojami, naudojant min-max normalizavimo funkciją, taip susiejant kintamojo reikšmes su intervalu [0,1]. Daugelio kategorinių nepriklausomų kintamųjų atveju nebuvo vykdomos transformacijos, išskyrus Auth_seg ir City_seg kintamųjų atžvilgiu, kuriuose esančių kategorijų skaičius yra didesnis nei 10. Šių kintamųjų atveju kategorijos buvo sugrupuojamos į klasterius, pagal turimus didžiausius poveikio panašumus nepriklausomam kintamajam.

Antrasis imties transformavimo variantas. Antrasis imties transformavimo variantas (toliau – var2) yra besiskiriantis nuo var1 tuo, kad klasterizavimo technika naudojama ir skaitiniams kintamiesiems, taip sugrupuojant pastaruosius į intervalus. Šio varianto atveju klasterizavimas naudojamas ir visiems kategoriniams nepriklausomiems kintamiesiems.

Trečiasis imties transformavimo variantas. Formuojant trečiąją imties transformavimo variantą (toliau – var3) , buvo taikoma WOE normalizavimo technika. Nepriklausomų kintamųjų perkodavimui WOE įtakos svoriais buvo naudojami antruoju imties transformavimo variantu suformuotos nepriklausomų kintamųjų reikšmės (klasterizuoti skaitiniai ir kategoriniai kintamieji). Remiantis atlikta transformacija, buvo apskaičiuojami WOE įtakos svoriai pagal šią formulę:

$$WOE_i = \ln \left(\frac{\frac{G_i}{\sum_{i=1}^n G_i}}{\frac{B_i}{\sum_{i=1}^n B_i}} \right),$$

kur G – „gerų“ skolininkų kiekis, B – „blogų“ skolininkų kiekis, o n – atitinkamą grupę atitinkančių skolininkų kiekis.

6 lentelė

Duomenų transformavimo variantų apibendrinimas

Duomenų transformavimo variantas	Skaitinių kintamųjų transformacija	Kategorinių kintamųjų transformacija
var1	Normalizavimas min-max funkcija	Sudaryti klasteriai kintamiesiems, kurie turi daugiau nei 10 kategorijų
var2	Sudaryti klasteriai, taip suskirstant kintamuosius į intervalus	Sudaryti klasteriai visiems kintamiesiems
var3	var2 sudaryti klasteriai perkoduojami į WOE	var2 sudaryti klasteriai perkoduojami į WOE

Šaltinis: sudaryta autorės, remiantis atlikta literatūros analize ir ekspertiniu vertinimu

Atlikus duomenų požymių ištyrimą, nepriklausomų kintamųjų atranką bei duomenų transformavimą, laikytina, kad imtis yra paruošta modeliavimui. Toliau darbe yra aptariamas klasifikavimo metodų pasirinkimas, kuris yra svarbus kaip sekantis kredito rizikos vertinimo modelio sudarymo etapas.

2.4 Klasifikavimo metodų pasirinkimas

Klasifikavimo metodo pasirinkimas turi reikšmingą įtaką modelio prognozinėms galimybėms. Kredito įstaigoms, kurios privalo kiekybiškai vertinti kredito riziką, klasifikavimo metodo pasirinkimas yra svarbus ir dėl pasirinkto metodo rezultatų paaiškinimo galimybių. Šiame moksliniame darbe kuriamas kredito rizikos vertinimo modelis internetinei IPPT platformai, kuri nėra įpareigota kiekybiškai vertinti kliento kredito riziką. Todėl, klasifikavimo metodo pasirinkimas yra pagrįstas atliktų mokslinių tyrimų rezultatais nagrinėjama tema.

Pirmoje šio darbo dalyje aptartos klasifikavimo metodų grupės yra naudojamos formuojant našiausią kredito rizikos vertinimo modelį. Šios grupės apima tradicinio, dirbtinio intelekto bei ansamblinio klasifikavimo metodus. Nors iš nagrinėtos mokslinės literatūros buvo pastebėtas sudėtingesnių metodų rezultatų (kaip dirbtinio intelekto, ansamblinio) pranašumas prieš techniškai paprastesnius metodus, atitinkančius tradicinius klasifikavimo metodus. Visgi, kaip dažnai atliekama ir mokslinėje praktikoje, LR klasifikavimo metodas pasirenkamas kaip lyginamasis. Toliau, pasirenkami metodai, priklausantys dirbtinio intelekto grupei, kaip SVM, RF, dėl jų aukšto našumo, pagrįsto mokslinių tyrimų rezultatais nagrinėjama tema. Ansamblinio klasifikavimo metodai taip pat pasižymi aukštu našumu, ypatingai mokslinėje literatūroje yra išskiriamas XGB metodas, kuris pasirenkamas ir šio tyrimo remuose.

LR. Dažnai LR metodo pasirinkimas yra pagrįstas dėl pritaikymo paprastumo bei gautų rezultatų interpretavimo galimybių. Visgi, kaip buvo aptarta pirmoje dalyje, mokslinių darbų autoriai įvardina ir metodo silpnąsias vietas, ypatingai susijusias su dideliais duomenų rinkiniais bei multikolinearumo problema. Šis metodas yra panašus į paprastą regresiją dėl savo funkcinės formos bei pačio ryšio modeliavimo tarp priklausomo kintamojo bei kelių nepriklausomų kintamųjų. Visgi, LR metodu modeliuojamas ryšys tarp priklausomo kintamojo ir nepriklausomo kintamojo nėra tiesinis. Ryšys tarp pastarųjų kintamųjų yra apibrėžiamas per netiesinę funkciją, kaip logistinę ar sigmoidinę. Logistinės funkcijos išraiška yra apibrėžiama, kaip:

$$P(Y = 1|X) = \frac{1}{1 + e^{-z}},$$

kur $P(Y = 1|X)$ apibūdina tikimybę, kad skolininkas taps „blogas“ turimo nepriklausomų kintamųjų rinkinio atžvilgiu, o z yra tiesinis nepriklausomų ir jų koeficientų sąveikos darinys, apibrėžiamas, kaip:

$$z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n,$$

kur β_0 yra laisvasis narys, β_n , kur $n > 0$ yra koeficientas, X_n – i-tasis nepriklausomas kintamasis, kai n yra nepriklausomų kintamųjų skaičius.

SVM. SVM metodas priklauso dirbtinio intelekto klasifikavimo metodų grupei. Šiuo metodu maksimizuojamas atstumas tarp dviejų artimiausių stebėjimų, kurie priklauso skirtingai priklausomo kintamojo grupei, per nubrėžtą ribą, vadinamą hiperplokštumą. SVM metodas pasižymi tuo, kad turi branduolio funkciją, apibrėžiančią hiperplokštumą. Todėl, vienas iš SVM klasifikavimo metodo pranašumų yra susijęs su jo gebėjimu įvertinti netiesinį ryšį tarp priklausomo ir nepriklausomų kintamųjų per netiesinę branduolio funkciją.

RF. RF metodas yra vienas iš dirbtinio intelekto grupei priskiriamų klasifikavimo metodų ir yra pagrįstas sprendimų medžiais. Kaip jau buvo minėta pirmoje šio mokslinio darbo 1.2.2

dalyje, RF metodas pasižymi mažesniu jautrumu išskirtims, lyginant su tradiciniais metodais, taip pat sudėtingesnė technika padidina metodo atsparumą multikolinearumo problemai bei užtikrina aukštesnį tikslumą apdorojant ir didelius duomenų rinkinius. RF metodo sudarymo technika pasižymi tuo, kad mokymo metu yra suformuojamas pasirinktas (ar automatiškai parinktas) medžių skaičius, kurie sudaromi iš mokymo duomenų imties ir nepriklausomų kintamųjų poaibių. Kiekvieno medžio rezultatas yra priklausomo kintamojo dvinarė reikšmė, o jų agregacija yra RF metodo galutinis rezultatas, atitinkantis tikimybę, kad skolininkas taps „blogas“.

XGB. XGB metodas yra vienas iš ansamblinio klasifikavimo metodų, kuris atliktos mokslinės literatūros analizės rėmuose lenkia tikslumu tradicinio ir dirbtinio intelekto grupėms priskirtų klasifikavimo metodus. XGB metodas, kaip ir RF yra pagrįstas sprendimų medžiais. Esminis skirtumas yra tas, kad yra formuojama vienas po kitą sekančių medžių grandinė su klaidų optimizacija, t.y. kiekvieno kuriamo medžio metu yra įvertinamos ankstesnio medžio klasifikavimo klaidos.

Nors sudėtingesnės technikos dirbtinio intelekto ir ansamblinio klasifikavimo metodai dažnai mokslinių tyrimų rezultatuose yra pranašesni nei tradiciniai klasifikavimo metodai, visgi, tai reikalauja ir daugiau laiko sąnaudų. Siekiant maksimizuoti klasifikavimo metodų našumą, reikia atlikti metodų parametrų optimizaciją. Šiam tikslui pasiekti, buvo naudojama atviro kodo „Optuna“ optimizavimo sistema, skirta automatizuotai skirtingų klasifikavimo metodų parametrų paieškai.

Atlikus klasifikavimo metodų pasirinkimą bei jų parametrų optimizaciją mokymo imties atžvilgiu, apmokami modeliai klasifikuoti skolininkus į „gerus“ ir „blogus“. Įgyvendinus modelių apmokymą, pereinama prie sukurtų klasifikavimo modelių įvertinimo. O tam atlikti, toliau aptariamas įvertinimo metodų pasirinkimas.

2.5 Modelio įvertinimo metodų pasirinkimas

Modelių įvertinimo metodus būtų galima išskirti į dvi grupes: metodai, kurie priklauso nuo lūžio taško, ir metodai, kurie nepriklauso nuo lūžio taško. Lūžio taškas yra riba, pagal kurią gautas modelio rezultatas yra paverčiamas į klasę. Metodai, kurie nepriklauso nuo lūžio taško, apima modelio įvertinimą pagal visus įmanomus lūžio taškus. Priešingai, metodai, kurie priklauso nuo lūžio taško, įvertina modelį tik pasirinkto lūžio taško atžvilgiu.

Metodai, kurie nepriklauso nuo lūžio taško. Apžvelgtų mokslinių tyrimų, nagrinėjančių kredito riziką kontekste, vienas iš pagrindinių modelių vertinimo metodų yra *ROC*. Ši kreivė yra sudaroma kiekvieno galimo lūžio taško atžvilgiu. Gauta kreivė atvaizduojama grafike, kurio X ašyje pateikiama „gerų“ skolininkų dalis, kurią modelis teisingai klasifikavo, Y ašyje – „blogų“

skolininkų dalis, kurią modelis teisingai klasifikavo. Taip pat šis metodas atskleidžia ir vertinamo modelio diskriminacinę galią, t.y. kuo gauta kreivė yra arčiau diagonaliosios tiesės (kuri atitinka visiškai atsitiktinį klasifikavimą), tuo modelio diskriminacinė galia yra mažesnė. Priešingai, kuo gauta kreivė yra arčiau kairiojo viršutinio kampo, tuo modelio diskriminacinė galia yra didesnė. ROC metodas yra grafinė vizualizacija, todėl kiekybinio įverčio šis metodas nesuteikia. Tačiau, šį metodą papildė kitas modelių rezultatams vertinti skirtas metodas – *AUC*. Šis metodas kiekybiškai įvertina modelio diskriminacinę galią per plotą po ROC kreive. *AUC* įvertis lygus 1 atitinka tobulą modelio diskriminavimą klasifikavimo grupių atžvilgiu, o 0.5 atitinka visiškai atsitiktinį modelio klasifikavimą. Todėl, siekiama, kad modelio *AUC* įvertis būtų viršijantis 0.5 ir kuo artimesnis 1. *GINI indeksas*, naudojamas modelių rezultatų vertinimui, yra *AUC*, tuo pačiu ir ROC metodų plėtinys. Kaip ir pastarieji metodai, *GINI* indeksas vertina, kaip modelis gerai diskriminuoja klasifikavimo grupes. *KS* metodas taip pat vertina modelio klasifikavimo grupių diskriminavimo galią. Techninis šio modelio įgyvendinimas yra gaunamas išrūšiuojant modelio rezultatų įverčius nuo didžiausio iki mažiausio, tuomet apskaičiuojamas maksimalus skirtumas tarp kiekvieno teisingai klasifikuotų „blogų“ skolininkų lygio ir neteisingai klasifikuotų „gerų“ skolininkų lygio įsipareigojimų nevykdymo lygio atžvilgiu, kuris ir yra *KS* įvertis. Taigi, šiuo metodu yra įvertinamas modelio gebėjimas atskirti „blogus“ skolininkus nuo „gerų“ skolininkų.

Metodai, kurie priklauso nuo lūžio taško. Modelių vertinimo metodai, priklausomi nuo lūžio taško, taip pat yra įtraukiami į modelių vertinimą. Nagrinėtos mokslinės literatūros tyrimų rėmuose klasifikavimo tikslumo matrica ir su ja susiję rodikliai yra vieni iš pagrindinių metodų, vertinant modelių rezultatus. Klasifikavimo matrica atskleidžia modelio gebėjimą teisingai klasifikuoti skolininkus bei daromas klaidas pasirinkto lūžio taško atžvilgiu. Pagrindiniai matricos elementai skirstomi į: teisingai identifikuotus „blogus“ skolininkus (angl. true positive) (toliau – *TP*), teisingai identifikuotus „gerus“ skolininkus (angl. true negative) (toliau – *TN*), neteisingai identifikuotus „gerus“ skolininkus (angl. false positive) (toliau – *FP*), ir neteisingai identifikuotus „blogus“ skolininkus (angl. false negative) (toliau – *FN*). Šios matricos elementais yra paremtas toks metodas, kaip *Ar* kuriuo įvertinama, kiek modelis teisingai klasifikavo tiek „gerų“, tiek „blogų“ skolininkų. Visgi, nagrinėjant kredito riziką, dažnai susiduriama su imties disbalansu, todėl, tikslumo metodo taikymas gali būti kiek netikslingas, dėl daugumos grupės turimo didesnio svorio testavimo imtyje. Todėl, esant imties disbalansui, būtų pravartu naudoti ir *subalansuotą tikslumo metodą* (angl. balanced accuracy) (toliau – *Bar*). Remiantis tikslumo matricoje esančiais elementais, yra suformuojami metodai, kurie individualiai vertina modelio gebėjimą klasifikuoti „gerus“ ir „blogus“ skolininkus. Pavyzdžiui, „gerų“ skolininkų tikslumą vertinantis metodas yra *Sp* o „blogų“ skolininkų tikslumą – *Se*. Su „blogų“ skolininkų tikslumo vertinimu yra susijęs ir

tikslumo (angl. precision) (toliau – *Pr*) metodas, kuriuo matuojama, kiek iš imtį sudarančių „blogų“ skolininkų, modelis gebėjo klasifikuoti juos teisingai. Mokslinėje literatūroje išskiriamas ir jautrumo ir tikslumo metodų kombinaciją apibrėžiantis *F1* metodas, kuriuo naudojimas yra pagrįstas balansuoto rodiklio tarp jautrumo ir tikslumo įvertinimu.

Iš atliktos mokslinės literatūros analizės buvo pastebėtas poreikis vertinti modelio rezultatus tiek atsižvelgiant į lūžio tašką, tiek neatsižvelgiant. Taip pat pastebima, kad mokslinių tyrimų autoriai renkasi daugiau nei vieną vertinimo metodą. O tai būtų galima pagrįsti tuo, kad skirtingų metodų įvertinimas gali būti besiskiriantis vienas nuo kito, todėl svarbu įvertinti modelį kompleksiškai. Remiantis atlikta mokslinės literatūros analize, buvo pasirinktas vertinimo metodų rinkinys (žr. 7 lentelę).

7 lentelė

Metodai, pasirinkti atlikti modelių įvertinimą

Metodas	Apskaičiavimas
<i>Metodai, kurie nepriklauso nuo lūžio taško</i>	
AUC	$\frac{1}{2} \left(1 + \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \right)$
KS	$\max [P(s G) - P(s B)]$
GINI	$2 \times AUC - 1$
<i>Metodai, kurie priklauso nuo lūžio taško</i>	
Ar	$\frac{TP + FN}{TP + TN + FP + FN}$
BAr	$\frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$
Se	$\frac{TP}{TP + FN}$
Sp	$\frac{TN}{TN + FP}$
Pr	$\frac{TP}{TP + FP}$
F1	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

Šaltinis: sudaryta autorės, remiantis atlikta mokslinės literatūros analize

Remiantis gautais modelių vertinimo rezultatais, yra pasirenkamas vienas iš aukščiausią našumą bei naudotojo tikslus atitinkantis modelis. Atrinktas modelis yra taikomas sudarant rangų skalės sistemą, kuri aptariama toliau.

2.6 Rangų skalės sistemos sudarymas

Gautas modelio rezultatas yra absoliuti reikšmė, atskleidžianti skolininko tikimybę tapti „blogu“ skolininku. Visgi, absoliutus skaičius nesuteikia pilnos galimybės atsakyti, kokios kredito rizikos yra tiriamasis skolininkas. Šiam trūkumui eliminuoti, yra sudaroma rangų skalės sistema, kuri klasifikuoja skolininkus į rizikos grupes. Rangų skalės sistemos formavimui buvo taikomos Bazelio bankų priežiūros komiteto rekomendacijos bei Youden indekso metodas optimaliam lūžio taškui nustatyti.

Bazelio bankų priežiūros komiteto rekomendacijos. Remiantis Bazelio bankų priežiūros komiteto rekomendacijomis, skirtomis bankų kredito rizikos valdymui, patariama sudaryti skalę bent iš 8 rangų, iš kurių 6 apibrėžtų „gerus“ skolininkus, o likę 2 – „blogus“ skolininkus (Basel Committee on Banking Supervision, 2020). Taip pat, sudarant rangų skalę, svarbu atsižvelgti ir į „gerų“, „blogų“ skolininkų dalį, esančią kiekviename range. Kiekvienas rangas turėtų apimti rango viduje esančius kuo panašesnius skolininkus, o pastarieji turėtų maksimaliai skirtis nuo kituose ranguose esančių skolininkų.

Youden indeksas. Siekiant nustatyti lūžio tašką tarp „gerų“ ir „blogų“ skolininkų rangų skalėje, buvo naudojamas Youden indekso metodas. Šiuo metodu optimalus lūžio taškas yra grindžiamas specifiškumo ir jautrumo vertinimo metodų teisingu klasifikavimu, t.y. siekiama rasti tokią optimalią ribą, kuri maksimizuotų „gerų“ ir „blogų“ skolininkų klasifikavimo tikslumą (Coolen-Maturi & Coolen, 2019).

Taigi, remiantis Bazelio bankų priežiūros komiteto rekomendacijomis, Youden indekso metodu nustatytu lūžio tašku yra sudaroma rangų skalės sistema. Gautą modelio rezultatą ir pačią rangų skalės sistemą galima integruoti į produkcinę aplinką, kuri palengvintų sprendimų priėmimą ne tik internetinės IPPT platformos veiklos organizavimo, bet ir bendradarbiaujančių kredito įstaigų kredito rizikos valdyme.

3. KREDITO RIZIKOS VERTINIMO MODELIO FIZINIAMS ASMENIMS REZULTATAI IR JŲ INTERPETACIJA

Šioje dalyje atliekamas mašininio mokymo metodais sudarytų fizinių asmenų kredito rizikos vertinimo modelių vertinimas. Mašininio mokymo metodai buvo atrinkti remiantis atlikta mokslinės literatūros analize, todėl taikytų metodų rinkinys apėmė: LR, SVM, RF, XGB metodus. Pastaraisiais metodais sudaryti modeliai buvo vertinami ir geriausiais rezultatais pasižymėjęs modelis buvo taikytas rangų skalės sistemai sudaryti. Kredito rizikos vertinimo modelio sudarymo seka, vertinimas ir rangų skalės sistemos sudarymas pateikiami šioje dalyje.

3.1 Modelio kūrimo imties sudarymas

Priklausomo kintamojo apibrėžimas, formuojant modelio kūrimo imtį. Fizinių asmenų kredito rizikos vertinimo modelio sudarymas buvo paremtas išduotomis paskolomis per IPPT platformą 2021-01-01 – 2024-01-28 laikotarpiu. Imties formavimas, atsižvelgiant į šį laikotarpį yra pagrįstas dviem priežastimis. Pirma, ekspertiniu požiūriu Covid-19 pandemijos pradžia galėjo turėti įtakos klientų elgsenai, kuri būtų neatitinkanti realios klientų elgsenos. Tokiu atveju modelis būtų šališkas, todėl ir pasižymintis prastesniu pritaikomumu prie realių aplinkos sąlygų. Antra, remiantis atlikta mokslinės literatūros analize ir ekspertiniu požiūriu, buvo pasirinktas bent vienerių metų fizinių asmenų elgsenos stebėjimo laikotarpis. Per šį laikotarpį siekiama įvertinti fizinio asmens apibrėžimą. Modelio sudarymo etape fizinio asmens apibrėžimas atitinka priklausomą kintamąjį. Todėl, per bent per vienerių metų laikotarpį siekiama įvertinti, ar fizinis asmuo atitiko „blogą“ skolininko bruožus.

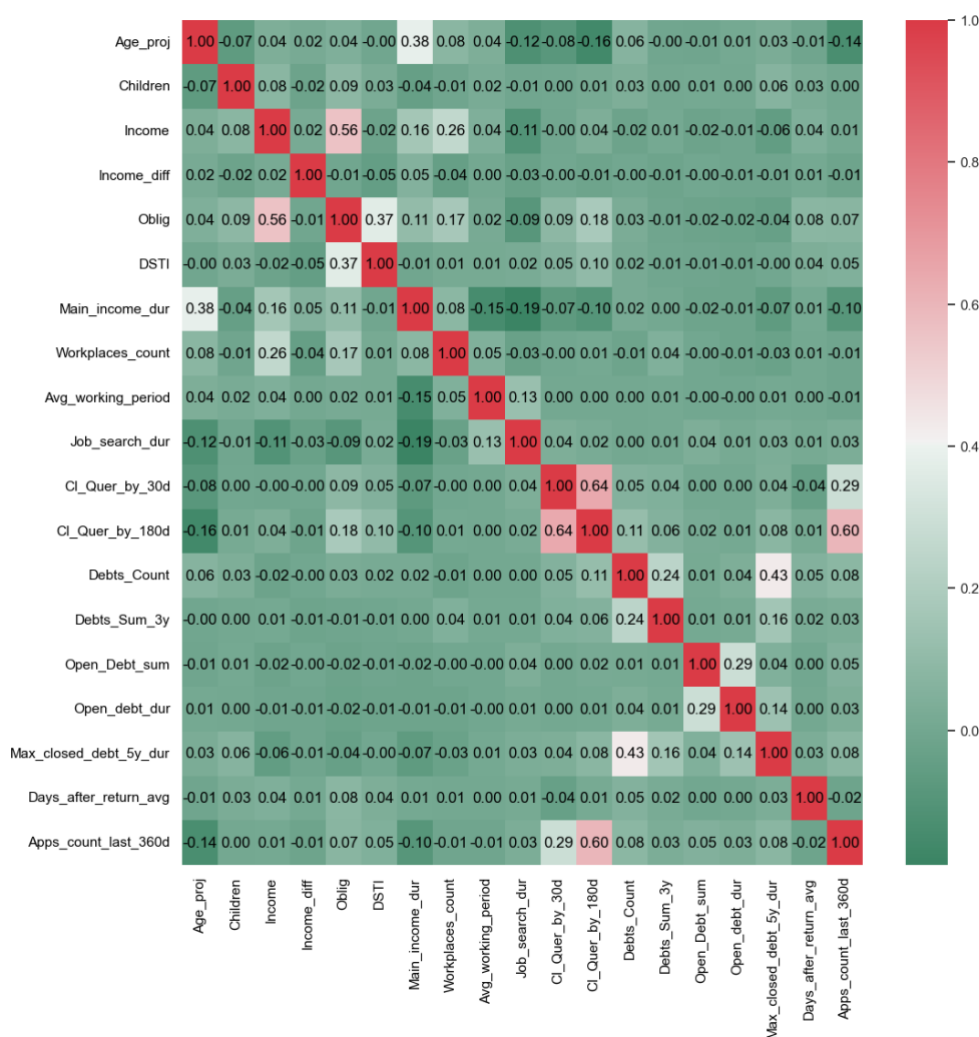
Vienas iš kertinių kredito riziką vertinimo modelių sudarymo etapų – „blogą“ skolininko apibrėžimas. Šiame tyrime „blogas“ skolininkas yra apibrėžiamas kaip turėjęs bent 90 dienų pradelimą atlikti mokėjimą už prisiimtą įsipareigojimą per stebėjimo laikotarpį. Šio vėlavimo termino pasirinkimas yra pagrįstas atsakingo skolinimo nuostatais (Lietuvos Bankas, 2025). Taip pat, kadangi imtis buvo suformuota iš fizinių asmenų, kuriems buvo išduota vartojimo paskola, buvo vertinama ir pradelsto mokėjimo suma. „Blogą“ skolininko apibrėžimą atitinkantis fizinis asmuo – kurio pradelsto mokėjimo suma atitiko bent 100 Eur. *Taigi, kredito rizikos vertinimo modelio sudarymui priklausomas kintamasis, atitinkantis dvinarę formą, kurio 1 reikšmė atitinka „blogą“ skolininko apibrėžimą, kai fizinis asmuo per stebėjimo laikotarpį turėjo bent 90 dienų mokėjimo pradelimą, ne mažesnę bent 100 Eur.*

Nepriklausomų kintamųjų atranka. Galutiniam nepriklausomų kintamųjų sąvadui suformuoti iš metodologinėje dalyje pateikto pradinio sąvado, buvo naudojami nepriklausomų kintamųjų atrankos metodai: 1) multikolinearumo problemos ištyrimas, 2) IV apskaičiavimas ir 3) WOE įtakos svorių monotoniskumo vertinimas kiekvieno nepriklausomo kintamojo atžvilgiu.

Pirma, buvo atliekamas nepriklausomų kintamųjų multikolinearumo problemos ištyrimas. Šiai problemai ištirti, buvo naudojamos koreliacijų matricos, tiesiniams (*pearson*) ir netiesiniams (*kendall*) ryšiams įvertinti (žr. 2 priedą). Nepriklausomų kintamųjų poroje identifikuota multikolinearumo problema buvo šalinta, jei koreliacijos reikšmės viršijo ± 70 ir galutiniame nepriklausomų kintamųjų sąvade buvo paliekamas tas kintamasis, kurio IV buvo aukštesnė (žr. 2 pav.).

2 paveikslas

Nepriklausomų kintamųjų sąvado koreliacijos matrica (Kendall metodu)



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Antra, buvo apskaičiuotos IV, atspindinčios kiekvieno nepriklausomo kintamojo diskriminacinę galią priklausomo kintamojo atžvilgiu. Iš galutinio nepriklausomų kintamųjų rinkinio buvo pašalinti tie kintamieji, kurių IV buvo mažesnė už 0.02 reikšmę. Šias vertes atitinkantys kintamieji yra už raudonai pažymėtos linijos (žr. 8 lentelę). Remiantis šiuo nepriklausomų kintamųjų atrankos metodu, naudojamų fizinių asmenų duomenų kontekste, kuriuo dienos metu buvo teikta paraiška, esamų darboviečių skaičius, gaunamų pajamų šaltinis, nepadengtų skolų suma ir trukmė, turimų prisiimtų įsipareigojimų suma bei turimų nepilnamečių vaikų skaičius neturi reikšmingos diskriminacinės galios priklausomo kintamojo atžvilgiu. Šio gauto rezultato priežastis – įvardinti kintamieji neturi didelės reikšmių sklaidos stebėjimų atžvilgiu, todėl jų reikšminga diskriminacinė galia priklausomo kintamojo atžvilgiu nebuvo identifikuota (žr. 1 priedą). Priešingai, turimų skolų suma per paskutinius 3-jus metus nuo paraiškos teikimo, per paskutinius 5-erius metus turėta maksimali skolos padengimo trukmė, užklausų pildymo skaičius per paskutinius pusę metų, skolų kiekis paraiškos teikimo metu identifikuoti kaip aukščiausią diskriminacinę galią priklausomo kintamojo atžvilgiu turintys nepriklausomi kintamieji. Rezultatas yra pagrįstas, kadangi įvardinti nepriklausomi kintamieji yra labiausiai susiję su priklausomo kintamojo, t.y. „bloga“ skolininko apibrėžimu.

8 lentelė

Nepriklausomų kintamųjų informacinės vertės (IV)

Kintamasis	IV
Debts_Sum_3y	0.5104
Max_closed_debt_5y_dur	0.466
CI_Quer_by_180d	0.4198
Debts_Count	0.4136
App_type	0.3404
Loan_Granted_Amount	0.3249
Apps_count_last_360d	0.2718
Main_income_dur	0.2385
Age_proj	0.2295
Days_after_return_avg	0.2191
CI_Quer_by_30d	0.166
Income_diff	0.1591
New_client	0.1353
Product_name	0.1312
Avg_working_period	0.1266
Job_search_dur	0.1159
Auth_meth	0.1015
Has_open_debt	0.097
Reissuance	0.0724

Kintamasis	IV
App_Applicant_City_fixed	0.0652
Doc_type	0.061
Gender	0.0507
Income	0.0465
DSTI	0.038
Working_hours	0.0194
Children	0.0145
Oblig	0.014
Workplaces_count	0.0055
Open_Debt_sum	0.0
Open_debt_dur	0.0
Income_source	0.0

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Trečia, buvo atliekama nepriklausomų kintamųjų analizė WOE įtakos svorių išraiškoje. Šioje dalyje buvo siekiama įvertinti iki šio nepriklausomų kintamųjų atrankos etapo atrinktų kintamųjų monotoniškumą, jų WOE įtakos svorių išraiškoje priklausomojo kintamojo atžvilgiu. Kintamieji, kurie nepasižymėjo monotoniškumu, buvo neįtraukti į galutinį nepriklausomų kintamųjų rinkinį, dėl galimo triukšmo sukėlimo. Monotoniškumo sąlygos neatitikę kintamieji – vidutinis fizinio asmens grįžimo į IPPT platformą dažnumas (Days_after_return_avg) ir vidutinė darbo stažo trukmė vienoje darbovietėje (Avg_working_period) (žr. 3 priedą). Gautos WOE vertės įvardintų kintamųjų atžvilgiu atskleidžia, kad fizinis asmuo, kuris dažniau grįžta į IPPT platformą (iki 3-jų dienų periodiškumu), teikdamas naujos paskolos prašymą, bei neturintis darbo stažo, turi mažesnę tikimybę būti priskirtas prie „blogo“ skolininko. Visgi, monotoniškumas akivaizdus nuo minėtų kintamųjų verčių, kada abiejų kintamųjų vertėms augant, fizinio asmens priskyrimas prie „blogo“ skolininko – mažėja.

Ketvirta, buvo taikyta šalinamoji LR, kurios pagrindu buvo suformuotas galutinis nepriklausomų kintamųjų sąvadas. Šalinamosios LR rezultatai parodė, kad didėjant tam tikrų kintamųjų reikšmėms – amžiui (atsižvelgiant į paskolos terminą), paskolos sumai bei darbo trukmei esamoje darbovietėje – mažėja fizinio asmens priskyrimo prie „blogo“ skolininko tikimybė (žr. 9 lentelę). Tuo tarpu didesnis skolų skaičius, bent vienos aktyvios skolos turėjimas, didesnė bendra skolų suma per pastaruosius trejus metus, ilgesnė maksimali skolos padengimo trukmė per pastaruosius penkerius metus, taip pat didesnis skirtumas tarp fizinio asmens tvarių pajamų ir paraiškoje nurodytų (atsižvelgiant į mažesnę reikšmę) bei kliento nurodytų pajamų, yra reikšmingai susiję su didesne tikimybe tapti „blogo“ skolininku. Be to, nustatyta, kad fizinio asmens turėtas ilgesnis nei 13 dienų laikotarpis tarp buvusios ir esamos darbovietės taip pat didina šią tikimybę. Rezultatai taip pat atskleidė, kad fiziniams asmenims, kurie paskolą ima kartu su

bendraskoliu, turi galimybę nuolat gyventi Lietuvoje, anksčiau buvo gavę paskolą per IPPT platformą, taip pat moterims, rizika tapti „blogu“ skolininku yra mažesnė. Visgi nustatyta, kad tuo atveju, kai DSTI rodiklis paraiškos teikimo metu siekė 37,63% ir daugiau, fizinio asmens tikimybė tapti „blogu“ skolininku reikšmingai padidėja.

9 lentelė

Atliktos šalinamosios logistinės regresijos (LR) rezultatai ir jų interpretacija

Kintamųjų grupė	Kintamojo pavadinimas	Pilnas kintamojo pavadinimas	Reikšmės	Koeficientas	P-vertė	Aprašymas
Demografinė	Constant	Laisvasis narys	(inf,31.42)	-2,7261	0	
			(31.42, 35.96)	-0.7142	0	
			(35.96, 39.46)	-0.9223	0	
	Age proj	Pareikėjo prognozuojamas amžius, atsižvelgiant į paskolos terminą	(39.46, 41.96)	-0.9622	0	Tikimybė tapti „blogu“ skolininku sumažėja vyresnio amžiaus klientams
			(41.96, 50.46)	-0.8162	0	
			(50.46, 53.62)	-1.0649	0	
			(53.62, 61.21)	-1.0287	0	
			(61.21, inf)	-0.8403	0	
	App_type	Paraiškos tipas pagal paskolos paskirtį (asmensinėms ar šeimos reikmėms) ir pareikėjo gyvenimo padėtį (vedęs, išsiskyręs, našlys ar nevedęs)	Married_Family_Loan	0.4842	<0.01	Ne šeimos reikmėms imantiems paskolą, tikimybė tapti „blogu“ skolininkais išauga.
			Not_Married	0.8516	0	
			Separated	0.9182	0	
			Alytus, Lazdijai, Utena, Jonava, Naujoji Akmenė	0	0	
			Barstonas, Šiauliai	0	>0.1	
			Druskininkai, Švenčionys, Pagėgiai, Kiti, Šalčininkai, Kelme, Karšių Ruda, Šilutė	0.5848	0	
Paraiška	App_Applicant_City_fixed	Pareikėjo deklaruota gyvenamoji vieta	Ignalina, Ukmergė, Klaipėda, Tauragė	0	>0.1	Deklaravusiems savo gyvenamąją vietą Vilniuje, Panevėžyje, Marijampolėje, tikimybė tapti „blogais“ skolininkais išauga, lyginant su kitais didžiausiais miestais, kaip Kaunas, Alytus, Utena, Jonava, Šiauliai, Mažeikiai, Klaipėda.
			Mažeikiai, Skuodas, Prienai, Kaimadorys, Telšiai	0	>0.1	
			Rietavas, Kupiškis, Zarasai, Neringa, Visaginas, Gargždai, Plungė, Kretinga	-0.3944	<0.1	
			Rokiškis, Marijampolė, Panevėžys, Vilkaviškis, Elektrėnai, Kalvarijos, Joniškis, Pakruojis	0.3191	<0.01	
			Varėna, Radviliškis, Palanga, Jurbarkas, Kaunas, Pasvalys, Šilalė	0	>0.1	
			Vilnius, Raseiniai	0.3209	<0.001	
			Šakiai, Moletai, Širvintai, Anykščiai, Kedainiai	0	>0.1	
	Gender	Pareikėjo lytis	Female	0	0	Tikimybė vyrams tapti „blogais“ skolininkais yra didesnė, lyginant su moterimis.
			Male	0.3290	0	
			AE	0	0	
			AR	0	>0.1	
			AT	0	>0.1	
	Doc_type	Pareikėjo tapatybės patvirtinimo dokumento tipas	I	0	>0.1	Skolininkai, kurių tapatybę patvirtinantis dokumentas atitinka laikiną leidimą, turi didesnę tikimybę tapti „blogais“ skolininkais.
			LI	2.1030	0	
			P	0	>0.1	
Pareikėjas			PI	0	>0.1	
			(inf, 737.5)	0	0	
			(737.5, 1437.5)	0	>0.1	
			(1437.5, 2512.7)	0	>0.1	
			(2512.7, 3476.1)	0	>0.1	
	Loan_Granted_Amount	Paskolos suma	(3476.1, 4580.5)	0	>0.1	Esant didesnei paskolos sumai (apytiksliai nuo 6000), tikimybė tapti „blogu“ skolininku sumažėja.
			(4580.5, 5990.5)	0	>0.1	
			(5990.5, 9990.65)	-0.2854	<0.01	
			(9990.65, 14680.0)	-0.3555	<0.01	
			(14680.0, inf)	-0.4488	<0.01	
			A	0	0	
			PA	-0.6307	0	
			PB	-0.3474	<0.1	Skolininkai, kurie ima paskolą autolizingo, vartojimo kredito linijos ir refinansavimo produktams, tikimybė tapti „blogais“ skolininkais išauga.
			YK	0	>0.1	
			R	0	>0.1	
Elgesys			V	-0.3425	<0.01	
			1et_vb2	0	0	
			BANKIDINK	0	>0.1	
	Auth_meth	Pasirinktas tapatybės tvirtinimo metodas	1et_mb, 1et_parox, 1et_wallet, SMARTID, 1et_ab,	0	>0.1	Skolininkai, kurie patvirtino tapatybę Swedbank ID būdu, turi mažesnę tikimybę tapti „blogais“ skolininkais.
			1et_banza, Instartor, 1et_nord	0	>0.1	
			Swedbank ID	-0.2837	<0.1	
			MPAR	0	>0.1	
			(inf,0.5)	0	0	
			(0.5, 2.5)	0.2385	<0.01	Skolininkai, kuriems paraiškos teikimo datai per 30 paskutinių dienų buvo užpildytos bent 1-6 užklauso apie kreditingumą, tikimybė tapti „blogais“ skolininkais yra didesnė.
			(2.5, 6.5)	0.2197	<0.1	
			(6.5, 9.5)	0	>0.1	
			(9.5, inf)	0	>0.1	
			(inf,0.5)	0	0	
			(0.5, 2.5)	0	>0.1	
			(2.5, 5.5)	0	>0.1	
Finansinė	CI_Quer_by_30d	Užklausų skaičius apie asmens kreditingumą per 30 dienų	(5.5, 6.5)	0	>0.1	Skolininkai, kuriems paraiškos teikimo datai per 180 paskutinių dienų buvo užpildytos bent 7 užklauso apie kreditingumą, tikimybė tapti „blogais“ skolininkais yra didesnė.
			(6.5, 9.5)	0.2117	<0.1	
			(9.5, 14.5)	0.5149	0	
			(14.5, 20.5)	0.8837	0	
			(20.5, inf)	1.1725	0	
			20.5	0	0	
			true	0	0	Askačiu gauta paskola per IPPT mažina tikimybę tapti „blogu“ skolininku.
			false	0.2098	0	
			(inf,0.5)	0	0	
			(0.5, 13.5)	-0.4395	<0.001	Skolininkai, kurie geba rasti darbą per 1-13 dienų, turi mažesnę tikimybę tapti „blogais“ skolininkais, lyginant su tais, kuriems esama darbovietė yra pirmoji arba geba rasti darbą per daugiau nei 13 dienų.
			(13.5, 59.5)	0	>0.1	
			(59.5, inf)	0	>0.1	
			(inf,0.06)	0	0	
			(0.06, 5.54)	0	>0.1	
			(5.54, 9.24)	-0.3630	<0.1	
Finansinė			(9.24, 12.95)	0	>0.1	Tikimybė tapti „blogu“ skolininku išauga nuo 37.63% DSTI rodiklio.
			(12.95, 33.04)	0	>0.1	
			(33.04, 37.63)	0	>0.1	
			(37.63, inf)	0.3466	<0.01	
			(inf,39.95)	0	0	
			(49.95, 502.53)	0.3921	0.01	Didesnė turimų skolų suma per paskutinius 3 metus didina tikimybę tapti „blogu“ skolininku.
			(502.53, inf)	0.9367	0	
			(inf,0.5)	0	0	
			(0.5, 1.5)	0.4719	<0.01	
			(1.5, 4.5)	0.5456	<0.001	Didesnis turimų ir buvusių skolų skaičius didina tikimybę tapti „blogu“ skolininku.
			(4.5, 9.5)	0.7494	0	
			(9.5, 19.5)	0.6444	0	
			(19.5, 34.5)	0.7126	0	
			(34.5, inf)	0.7564	0	
			(inf,94.5)	0	0	
Finansinė			(94.5, 313.5)	0.2765	<0.1	Didesnė maksimali skolos padengimo trukmė per paskutinius 5 metus didina tikimybę tapti „blogu“ skolininku.
			(313.5, inf)	0.5755	0	
			false	0	0	Nepadengtų skolų turėjimas didina tikimybę tapti „blogu“ skolininku.
			true	2.0170	0	
			(-inf, -653.74)	0.4273	<0.01	
			(-653.74, -424.08)	0.3080	<0.1	
			(-424.08, -86.53)	0.4412	0	Didesnis skirtumas tarp nurodytų ir Sodrove matomų pajamų didina tikimybę tapti „blogu“ skolininku.
			(-86.53, -22.97)	0.2075	<0.1	
			(-22.97, -2.01)	0	>0.1	
			(-2.01, inf)	0	0	
			(inf,5.5)	0	0	
			(5.5, 7.5)	0	>0.1	
			(7.5, 11.5)	0	>0.1	
			(11.5, 16.5)	-0.2057	<0.1	Didesnė darbo trukmė dabartinėje darbovietėje mažina tikimybę tapti „blogu“ skolininku.
			(16.5, 30.5)	-0.6128	0	
			(30.5, 37.5)	-0.6745	<0.001	
			(37.5, 67.5)	-0.5720	0	
			(67.5, 113.5)	-0.8517	0	
			(113.5, 178.5)	-0.7437	<0.001	
			(178.5, inf)	-0.8910	<0.001	

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Taigi, atlikus keturių etapų nepriklausomų kintamųjų atranką, buvo suformuotas galutinis nepriklausomų kintamųjų sąvadas (žr. 10 lentelę). Nurodyti nepriklausomi kintamieji buvo pritaikyti fizinių asmenų kredito rizikos vertinimo modelio sudarymui.

10 lentelė

Galutinis nepriklausomų kintamųjų sąvadas

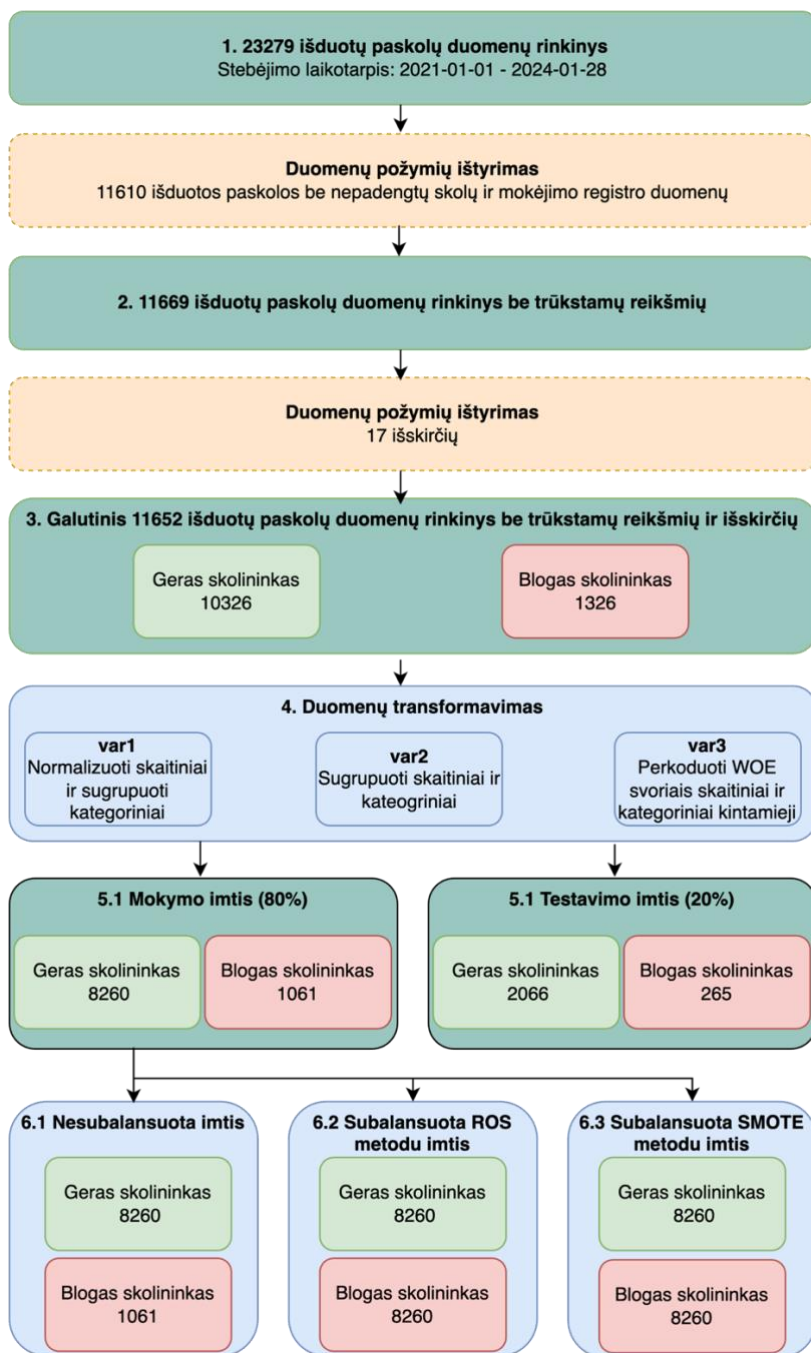
Kintamasis
Debts_Sum_3y
Max_closed_debt_5y_dur
CI_Quer_by_180d
Debts_Count
App_type
Loan_Granted_Amount
Main_income_dur
Age_proj
CI_Quer_by_30d
Income_diff
Product_name
Job_search_dur
Auth_meth
Has_open_debt
Reissuance
App_Applicant_City_fixed
Doc_type
Gender
DSTI

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Modelio sudarymo ir testavimo imčių sudarymas. Suformavus priklausomąjį kintamąjį ir atlikus nepriklausomų kintamųjų atranką, atliekamas duomenų paruošimas, kuris yra vienas iš kertinių etapų sudarant kokybišką modelį. Duomenų paruošimą būtų galima apibendrinti modelio sudarymo proceso etapais, pagrįsto duomenų požymių ištyrimu: 1) trūkstamų reikšmių ir išskirčių analizė, 2) duomenų transformavimas ir 3) duomenų imties balansavimas (žr. 3 pav.).

3 paveikslas

Kredito rizikos vertinimo modelio sudarymo procesas



Šaltinis: sudaryta autorės

Pirma, pradinis duomenų rinkinys buvo sudarytas iš 23279 išduotų paskolų. Atliekant trūkstamų reikšmių analizę, buvo pašalintos 11610 duomenų eilutės. Toks sprendimas buvo pagrįstas tuo, kad trūkstamos reikšmės buvo susijusios su informacija apie fizinio asmens skolas, o šie duomenys yra vieni reikšmingiausių kredito rizikos vertinimo kontekste. Toliau, buvo atliekama išskirčių analizė, kurios galutinis rezultatas pateiktas 1 priede. Šiame kredito rizikos vertinimo modelio sudarymo procese buvo pašalinta 17 duomenų eilučių. Po pirmojo duomenų

požymių ištirimo etapo, imties dydis siekė 11652 išduotas paskolas, iš kurių 10326 (89%) apibrėžė „gerą“ skolininką, o 1326 (11%) – „blogą“ skolininką.

Antra, remiantis atlikta mokslinės literatūros analize, buvo nuspręsta sudaryti kredito rizikos vertinimo modelius, atliekant duomenų transformavimą trimis variantais: 1) normalizuojant skaitinius kintamuosius, o kategorinius – kategorizuojant; 2) kategorizuojant tiek skaitinius, tiek kategorinius kintamuosius; 3) perkoduojant duomenis į WOE įtakos svorius. Transformavimo svarba yra pagrįsta: skirtingų kintamųjų mastelių ir sklaidos suvienodinimu, bei modelio našumo gerinimu. O kelių duomenų transformavimo technikų naudojimas buvo pagrįstas, siekiant sudaryti našiausią fizinių asmenų kredito rizikos vertinimo modelį naudojamų duomenų kontekste.

Trečia, kredito rizikos vertinimą nagrinėjanti mokslinė literatūra išskiria imties disbalanso problemą. Ši problema neišvengtinai atsispindi ir šio tyrimo kontekste, kai imtyje „blogų“ skolininkų apibrėžimą atitinkančių fizinių asmenų yra vos 11%. Todėl, siekiant įvertinti sudarytų modelių našumo skirtumus, modelių sudarymui buvo naudojama nesubalansuota mokymo imtis bei balansuotos dviem technikomis (ROS ir SMOTE). Atitinkamai, pirmuoju būdu mokymo imtį sudarė 8260 „geri“ skolininkai ir 1060 „blogi“ skolininkai, o kitais dviem būdais, atitinkančiais didinimo technikas – 8260 tiek „geri“, tiek „blogi“ skolininkai. ROS balansavimo technika nebuvo pritaikyta šio tyrimo kontekste, dėl ir taip „blogus“ skolininkus atitinkančių fizinių asmenų mažo skaičiaus.

Kai mokymo ir testavimo imtys yra pilnai paruoštos fizinių asmenų kredito rizikos vertinimo modelių sudarymui, pereinama prie klasifikavimo metodų pritaikymo mokymo imčiai.

3.2 Modelių sudarymas ir atranka

Remiantis atlikta literatūros analize, buvo nuspręsta sudaryti kredito rizikos vertinimo modelius, pagrįstus LR, SVM, RF, XGB klasifikavimo metodais. Taip pat buvo nuspręsta sudaryti skirtingomis duomenų transformavimo ir imties balansavimo technikomis pagrįstus modelius. Todėl, iš viso buvo sudaryti 36 modeliai. Geriausio kredito rizikos vertinimo modelio pasirinkimas buvo pagrįstas modelių vertinimo metodais, aprašytais 2.5 dalyje. Visų sudarytų modelių rezultatai pateikiami 4, 5, 6, 7 prieduose.

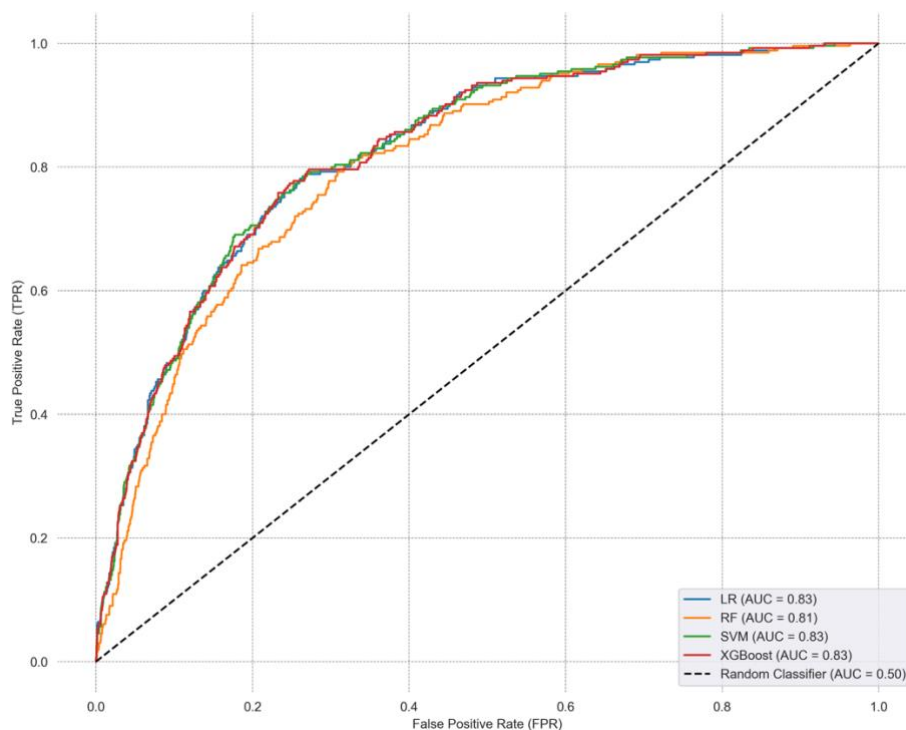
Sudarytų modelių rezultatai atskleidžia, kad pagal AUC vertinimo metodą, geriausiais rezultatais pasižymėjo modeliai, kurių sudarymui buvo naudojami duomenys, pagrįsti var3 duomenų transformavimo metodu (duomenų transformavimu į WOE įtakos svorius). Taip pat remiantis imties balansavimo technikomis, sudarytų modelių rezultatai atskleidžia SMOTE

metodo pranašumą modelio našumo atžvilgiu, lyginant su nesubalansuotos imties ir ROS metodu pagrįstais modeliais.

Apibrėžus geriausiais rezultatais pasižymėjusius duomenų transformavimo ir imties balansavimo technikas, tolimesnė veiksmų seka yra susijusi su geriausio modelio pasirinkimu. Vienas iš pagrindinių vertinimo metodų, siekiant išsirinkti geriausią prognozinę galią turintį modelį, yra ROC kreivė ir ją atitinkantis skaitinis metodas – AUC (žr. 4 pav.). Pastebima, kad RF klasifikavimo metodu pagrįstas modelis yra labiausiai besiskiriantis nuo kitų modelių sudarymui naudotų klasifikavimo metodų. RF modelio rezultatas yra prasčiausias, remiantis tiek ROC kreivės vizualine forma bei jos skaitine išraiška – AUC. Atsižvelgiant tik į ROC ir AUC modelių vertinimo metodus, kitais klasifikavimo metodais (LR, SVM, XGB) pagrįsti modeliai pasižymi reikšmingai nesiskiriančiomis prognozinėmis galimybėmis, jų AUC vertė yra vienoda ir atitinkanti 0.83.

4 paveikslas

Geriausio imties varianto (var3) ir subalansavimo technikos (SMOTE) modelių ROC ir AUC reikšmės



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Visgi, atsižvelgiant į kitus modelių vertinimo metodus, pastebimi minėtais klasifikavimo metodais pagrįstų modelių skirtumai (žr. 11 lentelę). LR AUC skaitinė vertė yra žemesnė, lyginant su SVM, XGB metodais pagrįstų modelių rezultatais. Taip pat vertinant modelių Se, atskleidžianti modelio galimybes tiksliai atskirti „blogus“ skolininkus, LR pasižymi prastesniu rezultatu nei SVM ir XGB lyginamieji modeliai. Kadangi kredito rizikos vertinime didesnis nuostolis

patiriamas neteisingai klasifikuojant „blogus“ skolininkus ir todėl remiantis Se, PPV ir β vertinimo metodų rezultatais, SVM ir XGB modeliai išrenkami kaip aukščiausią prognozinę galią turintys modeliai. O pagal minėtus vertinimo metodus, SVM ir XGB modelių rezultatai yra sąlyginai identiški.

11 lentelė

Geriausio imties varianto (var3) ir subalansavimo technikos (SMOTE) modelių rezultatų vertinimo metodų reikšmės¹

SMOTE				
	LR	SVM	RF	XGB
AUC	0.8269	0.8284	0.8076	0.8284
GINI	0.6538	0.6569	0.6151	0.6569
Ar	0.7683	0.7679	0.7945	0.7675
BAr	0.7542	0.7589	0.7229	0.7586
F1	0.4194	0.4226	0.4108	0.4222
KS	0.5207	0.5211	0.4872	0.5211
Sp	0.7725	0.7706	0.8156	0.7701
Se	0.7358	0.7472	0.6302	0.7472
PPV	0.2932	0.2946	0.3047	0.2942
NPV	0.958	0.9596	0.945	0.9596
α	0.2275	0.2294	0.1844	0.2299
β	0.2642	0.2528	0.3698	0.2528

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Tačiau, dėl praktinio pritaikomumo, potencialios galimybės ateityje permokėti sudarytą fizinių asmenų kredito rizikos vertinimo modelį bei nagrinėjama tema moksliniuose tyrimuose išskiriamo XGB klasifikavimo metodo aukšto rezultato, pastaruoju klasifikavimo metodu sudarytas fizinių asmenų kredito rizikos vertinimo modelis yra pasirenkamas kaip aukščiausią prognozinę galią turintis modelis. Gautus tyrimo rezultatus taip pat patvirtina kitų autorių atlikti fizinių asmenų kredito rizikos vertinimo srities tyrimai, kuriuose geriausi modeliai buvo sudaryti taikant XGB klasifikavimo metodą (Kang ir kt., 2021; Suhadolnik ir kt., 2023; Špicas ir kt., 2023). Visgi pastebima, kad XGB klasifikavimo metodas fizinių asmenų kredito rizikos vertinimo tyrimuose vis dar taikomas rečiau nei LR, kuri išlieka dažniausiai naudojamu baziniu metodu. Toliau darbe pasirinktu klasifikavimo metodu sudarytas modelis buvo naudojamas rangų skalės sistemos sudarymui.

¹ Mėlynai pažymėtos reikšmės atitinka geriausią modelį pagal atitinkamą vertinimo metodą.

3.3 Optimalaus lūžio taško nustatymas ir rangų skalės sistemos sudarymas

Iš sudarytų kredito rizikos vertinimo modelių, XGB modelis paremtas WOE ir SMOTE technikomis buvo išrinktas kaip našiausias modelis. Atrinktas klasifikavimo modelis yra naudojamas optimalaus lūžio taško nustatymui bei rangų skalės sistemos sudarymui.

Atsižvelgus į kredito rizikos vertinimo problemos ypatumus ir FP FN klaidų ekonominio poveikio skirtumą, optimalaus lūžio taško nustatymui yra naudotas Youden indeksas. Šis metodas yra pagrįstas klasifikavimo tikslumo analize.

Rangų skalė sudaryta atsižvelgus į Bazelio bankų priežiūros komiteto rekomendacijomis ir nustatytą lūžio tašką T. Remiantis Bazelio rekomendacijomis, patariama sudaryti rangų skalę iš ne mažiau kaip 8 rangų, iš kurių ne mažiau kaip 6 rangai atitiktų „gerus“ skolininkus ir bent 2 – „blogus“ skolininkus. Todėl siekta, kad lūžio taškas T atitiktų 7-o arba 8-o rango apatinio rėžio reikšmę.

Sudaryta rangų skalės sistema (žr. 12 lentelę) atskleidžia, kad didžiausia dalis „gerų“ skolininkų atitiko 2-5 rangus (60% visų „gerų“ skolininkų). O „blogi“ skolininkai koncentruojasi 7-10 ranguose (76% visų „blogų“ skolininkų). Būtent nuo 7-ojo rango, kurio apatinė rėžio riba atitinka nustatytą lūžio tašką, priskirti skolininkai atitinka „blogų“ skolininko apibrėžimą. Sudaryta rangų skalės sistema atskleidžia, kad 10 range yra didžiausia dalis „blogų“ skolininkų (24%), o šio rangų vėlavimo tikimybė (toliau – PD) ir yra didžiausias.

12 lentelė

Rangų skalės sistema, pasirinkto modelio atžvilgiu ²

Rangas	pd_min	pd_max	n_skaičius	n_geri	n_blogi	pd_avg	GV	GG	BV_ODF	BB	Reitingas	Rizikos lygis
1	0.00	7.99	728	712	16	0.0614	0.978	0.069	0.022	0.0019	A1	žemos
2	8.00	15.99	2308	2097	211	0.1201	0.9086	0.2031	0.0914	0.0248	A2	
3	16.00	23.99	2054	1740	314	0.2	0.8471	0.1685	0.1529	0.0368	B1	vidutinės
4	24.00	31.99	1655	1326	329	0.2795	0.8012	0.1284	0.1988	0.0386	B2	
5	32.00	39.99	1663	1020	643	0.3585	0.6133	0.0988	0.3867	0.0754	C1	aukštesnės
6	40.00	47.99	1370	811	559	0.4407	0.592	0.0785	0.408	0.0656	C2	
7	48.00	60.94	2184	1040	1144	0.5405	0.4762	0.1007	0.5238	0.1342	D1	aukštos
8	60.95	73.88	1986	736	1250	0.6747	0.3706	0.0713	0.6294	0.1466	D2	
9	73.89	86.83	2565	565	2000	0.8046	0.2203	0.0547	0.7797	0.2346	E1	labai aukštos
10	86.84	99.98	2338	279	2059	0.9247	0.1193	0.027	0.8807	0.2415	E2	
11	99.99	100.00									E3	nemokumas

² *pd_min* – nemokumo tikimybės apatinė riba grupėje; *pd_max* – nemokumo tikimybės viršutinė riba grupėje; *n_skaičius* – visų skolininkų skaičius grupėje; *n_geri* – „gerų“ skolininkų skaičius grupėje; *n_blogi* – „blogų“ skolininkų skaičius grupėje; *pd_avg* – nemokumo tikimybės vidurkis grupėje; *GV* – „gerų“ skolininkų dalis nuo visų skolininkų grupėje; *GG* – „gerų“ skolininkų dalis grupėje nuo visų „gerų“ skolininkų grupėje; *BV_ODF* – „blogų“ skolininkų dalis nuo visų skolininkų grupėje; *BB* – „blogų“ skolininkų dalis grupėje nuo visų „blogų“ skolininkų.

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

Apibendrinant, pritaikius Youden indekso metodą, nustatytas lūžio taškas $T=0.48$ nuo kurio priskiriami fiziniai asmenys yra laikomi kaip „blogi“ skolininkai. Rangų skalės sistema sudaryta atsižvelgus į Bazelio bankų priežiūros komiteto rekomendacijas ir nustatytą lūžio tašką. Rangų skalė suformuota iš 11 rangų, iš kurių 6, tikėtina, kad fiziniai asmenys atitiks „gerą“ skolininką ir yra priskirti prie A1-C2 kredito rizikos lygio reitingo, o 3, atitinkantys D1-E2 kredito rizikos lygio reitingą, tikėtina, kad atitiks „blogą“ skolininką, ir 1, priskiriamas prie E3 kredito rizikos lygio reitingo – yra faktiškai nemokūs fiziniai asmenys. Toks rangų skalės sistemos sudarymas atitinka Bazelio bankų priežiūros komiteto rekomendacijas.

IŠVADOS

1. Kredito rizika yra siejama su trimis pagrindiniais aspektais: nemokumo tikimybe, galimų nuostolių dydžiu ir neapibrėžtumu laike. Tikimybė ir laikas yra tarpusavyje susiję veiksniai – kredito davėjui kyla rizika, kad skolininkas laiku neįvykdys prisiimtų finansinių įsipareigojimų. Kredito rizika taip pat yra susijusi su galimų nuostolių dydžiu, todėl didesnės apimties finansiniai sandoriai daro reikšmingesnę įtaką kredito davėjo pelningumui. Ši rizika išsiskiria jautrumu makroekonominiams veiksniams: besikeičiančios ekonominio ciklo fazės tiesiogiai veikia kredito gavėjus, o netiesiogiai – kredito davėjus. Mokslinėje literatūroje pabrėžiama, kad kredito rizika ypač smarkiai veikia kredito gavėjus dėl galimo darbo praradimo ar disponuojamų pajamų sumažėjimo. Visgi, nevaldoma kredito rizika gali paveikti visas finansų rinkos grandis ir realųjį ekonomikos sektorių, o toks reiškinys literatūroje apibūdinamas kaip „drugelio efektas“.
2. Siekiant išvengti nevaldomos kredito rizikos plitimo makroekonominiu mastu, būtina taikyti tinkamas kredito rizikos valdymo priemones. Viena tokių priemonių – kiekybinis kredito gavėjo mokumo vertinimas taikant mašininio mokymosi metodus. Kredito rizikos vertinimo modelio sudarymo procesas apima priklausomojo kintamojo apibrėžimą, nepriklausomų kintamųjų atranką, klasifikavimo metodo parinkimą bei sudaryto modelio įvertinimą. Šiuolaikinėje literatūroje vis dažniau taikomi pažangūs dirbtinio intelekto ir ansamblinio klasifikavimo metodai, kurie pasižymi didesniu tikslumu, tačiau jų trūkumas – žemas paaiškinamumas. Tuo tarpu klasikiniai metodai, pavyzdžiui, LR, išsiskiria aukštu paaiškinamumu, tačiau yra jautrus multikolinearumui, imties disbalansui ir ryšio tarp kintamųjų prielaidoms. Todėl galutinis metodų pasirinkimas priklauso nuo įvertinimo rezultatų, o mokslininkai dažniausiai taiko kelis vertinimo metodus norėdami užtikrinti išsamų modelio efektyvumo įvertinimą.
3. Empirinėje darbo dalyje taikyta metodologija paremta mokslinės literatūros analize, o identifikuotos esminės atliktų tyrimų problemos – nepakankamai atlikta nepriklausomų kintamųjų atranka, nepakankamas duomenų požymių ištyrimas, ypač imties disbalansas bei duomenų transformavimo stoka – buvo integruotos į modelio sudarymą. Tokiu būdu siekta padidinti kredito rizikos vertinimo modelio efektyvumą. Visų minėtų problemų sprendimų taikymas viename empiriniame tyrime pateikiamas pirmą kartą moksliniame lygmenyje.

4. Fizinų asmenų kredito rizikos vertinimo modelis buvo sudarytas naudojant 2021-01-01–2024-01-28 laikotarpio IPPT platformos duomenis. Geriausiomis prognozinėmis savybėmis pasižymintis modelis buvo paremtas XGB klasifikavimo metodu, taikant WOE duomenų transformavimą ir SMOTE imties balansavimo technikas. Modelio prognozinis tikslumas siekė 0.83 AUC. Papildomai, siekiant identifikuoti reikšmingiausius kredito riziką lemiančius veiksnius ir suformuoti galutinį kintamųjų sąvadą, buvo taikyta šalinamoji logistinė regresija. Jos rezultatai parodė, kad kliento skolų rodikliai (skolų skaičius, aktyvių skolų buvimas, skolų suma ir jų padengimo trukmė), pajamų neatitikimai bei darbo stabilumo veiksniai yra reikšmingai susiję su fizinio asmens priskyrimu prie „blogo“ skolininko. Tuo tarpu demografiniai, elgsenos ir finansiniai veiksniai, tokie kaip amžiaus prognozė, atsižvelgiant į paskolos terminą, darbo trukmė esamoje darbovietėje, paskolos ėmimas su bendraskoliu, galimybė nuolat gyventi Lietuvoje, ankstesnė paskolų patirtis IPPT platformoje bei moteriška lytis, mažina kredito riziką. Didžiausią įtaką kredito rizikos lygio prognozavimui turėjo kliento skolų rodikliai bei dažnos paskolų užklauskos, kas patvirtina tiek logistinės regresijos, tiek XGB modelio rezultatus. Remiantis galutinio modelio prognozėmis, buvo sudaryta reitingavimo sistema, segmentuojanti klientus pagal rizikos lygį nuo A1 iki E3. Lūžio taškas nustatytas ties 48%, nuo kurio klientas priskiriamas „blogo“ skolininko kategorijai (D1 reitingas). Šios išvados mokslinis naujumas pasireiškia tuo, kad pirmą kartą moksliniame lygmenyje buvo sukurtas fizinių asmenų kredito rizikos vertinimo modelis, pritaikytas Lietuvos finansų rinkos specifikai, apjungiant pažangius mašininio mokymosi metodus su statistine kintamųjų atranka bei realius IPPT platformos ir kredito įstaigų duomenis. Modelyje sistemingai sprendžiamos mokslinėje literatūroje identifikuotos kredito rizikos vertinimo problemos, tokios kaip nepriklausomų kintamųjų atranka, imties disbalansas ir požymių transformavimo stoka, taip sukuriant metodologiškai pagrįstą ir praktiškai pritaikomą sprendimą Lietuvos finansų sektoriui.

PASIŪLYMAI

1. Šio mokslinio darbo tyrimas buvo paremtas tik vienos Lietuvoje veikiančios IPPT platformos duomenimis, todėl siekiant padidinti modelio reprezentatyvumą ir pritaikomumą ne tik IPPT platformos viduje, bet ir už jos ribų, siūloma įtraukti ir didžiuosius komercinius šalies bankus. Tai leistų įvertinti modelio veikimą platesniame klientų portfelyje ir sumažintų galimą šališkumą.
2. Įtraukti į imtį klientus, kurių paraiškos buvo atmestos. Šio mokslinio tyrimo rezultatas – fizinių asmenų kredito rizikos vertinimo modelis – buvo sudarytas remiantis tik patenkinomomis paskolų paraiškomis. Įtraukus ir atmestas paraiškas būtų galima išvengti atrankos šališkumo, taip pat pagerinti modelio gebėjimą prognozuoti nemokumo tikimybę.
3. Atsižvelgiant į tai, kad kredito rizika yra jautri makroekonominių ciklų pokyčiams, siūloma modelio sudarymo procese įtraukti tokius rodiklius kaip nedarbo lygis, palūkanų normos, infliacija, BVP augimas ar VIX indeksas. Tai padėtų tiksliau įvertinti ekonominių veiksnių įtaką fizinių asmenų mokumui ir padidintų modelio prognozinę vertę.
4. Atliekant mokslinį tyrimą buvo analizuojama tik vartojimo paskolų rinka. Tolimesniuose tyrimuose siūloma įtraukti ir kitas finansavimo formas, pavyzdžiui, Lietuvos rinkoje mokslinių tyrimų, susijusių su būsto paskolų kredito rizikos vertinimo modelių sudarymu ir taikymu, iki šiol praktiškai nėra atlikta. Tai suteiktų galimybę įvertinti modelių universalumą bei skirtingų segmentų rizikos profilius.

FIZINIŲ ASMENŲ KREDITO RIZIKOS VERTINIMO MODELIS

BEATRIČĖ RAUKŠTAITĖ

Magistro baigiamasis darbas

Ekonominė analitika

Vilniaus universiteto Ekonomikos ir verslo administravimo fakultetas

Darbo vadovė – Prof., Dr. (HP) Rasa Kanapickienė

Vilnius, 2026

SANTRAUKA

58 puslapiai, 12 lentelių, 4 paveikslai, 34 literatūros šaltiniai.

Pagrindinis šio magistro baigiamojo darbo tikslas – sukurti fizinių asmenų kredito rizikos vertinimo modelį, pasitelkiant mašininio mokymosi metodus, ir kiekybiškai įvertinti šių metodų tinkamumą vartojimo paskolų kredito rizikos prognozavimui Lietuvos rinkoje. Darbo uždaviniai apėmė kredito rizikos koncepcijos, iššūkių ir ekonominės reikšmės analizę, kredito rizikos vertinimo modelio sudarymo koncepcijos sisteminimą, modelio sudarymo metodologijos pagrindimą, modelio rezultatus bei jų vertinimą.

Darbas susideda iš trijų pagrindinių dalių: mokslinės literatūros analizės, empirinio tyrimo ir jo rezultatų ir suformuluotų išvadų bei pasiūlymų. Literatūros analizei buvo remtasi tarptautiniais ir nacionaliniais šaltiniais, siekiant atskleisti kredito rizikos koncepciją, keliamus iššūkius ir ekonominę reikšmę, kiekybinių valdymo priemonių sudarymo koncepciją ir modernių klasifikavimo metodų taikymo tendencijas. Akcentuojami skirtingi klasifikavimo metodai, jų privalumai ir trūkumai, duomenų požymių ištyrimo, duomenų transformavimo ir imties balansavimo svarba.

Empirinė darbo dalis grindžiama IPPT platformoje suteiktų vartojimo paskolų duomenimis 2021-01-01 – 2024-01-28 laikotarpiu. Naudoti duomenų šaltiniai apėmė nacionalinių institucijų ir registrų, paraiškos formos bei kliento elgseną apibūdinančius duomenis. Bendrai buvo sudaryti 36 modeliai. Išrinktas geriausias prognozinės savybės turintis fizinių asmenų kredito rizikos vertinimo modelis, paremtas XGB klasifikavimo metodu, WOE duomenų transformavimo ir SMOTE imties balansavimo technikomis. Atrinkto modelio prognozavimo tikslumas siekia 0,83 AUC įverčio. Empirinio rezultato vertinimas atskleidė, kad reikšmingiausi rizikos lygi prognozuojantys veiksniai yra kliento skolų rodikliai bei paskolų užklausų dažnis. Sukurta klientų rangų skalės sistema, segmentuojanti klientus pagal rizikos lygius nuo A1 iki E3, o statistiškai pagrįstas lūžio taškas nustatytas ties 48%, nuo kurio klientas priskiriamas prie „blogo“ skolininko kategorijos.

Gauti empirinio tyrimo rezultatai patvirtino, kad pažangūs mašininio mokymosi metodai gali užtikrinti aukštą kredito rizikos modelių tikslumą. Tyrimo rezultatai taip pat atskleidė, kad nepriklausomų kintamųjų atranka, imties disbalanso ir duomenų transformavimo valdymas turi reikšmingą įtaką galutinio modelio efektyvumui.

Darbo rezultatai gali būti publikuojami moksliniuose leidiniuose, nagrinėjančiuose fizinių asmenų kredito rizikos valdymą, mašininio mokymosi taikymą ir vartojimo paskolų rinkas. Tyrimo praktinė reikšmė pasireiškia galimybe finansų įstaigoms bei IPPT platformoms naudoti sukurta modelį sprendimų priėmimo procesuose, o teorinė – kaip pagrindas tolimesniems tyrimams, integruojant būsto paskolų rinką, makroekonominis rodiklius ir didžiųjų komercinių bankų duomenis.

CREDIT RISK ASSESSMENT MODEL FOR INDIVIDUALS

BEATRIČĖ RAUKŠTAITĖ

Master thesis

Economic Analysis

Vilnius university, Faculty of Economics and Business Administration

Supervisor – Prof., Dr. (HP) Rasa Kanapickienė

Vilnius, 2026

SUMMARY

58 pages, 12 tables, 4 pictures, 34 references.

The main objective of this master's thesis is to develop a credit risk assessment model for individuals by applying machine learning methods and to quantitatively evaluate the suitability of these methods for predicting consumer loan credit risk in the Lithuanian market. The tasks of the thesis included the analysis of the concept, challenges, and economic significance of credit risk, the systematization of credit risk assessment model development principles, the justification of the model development methodology, as well as the presentation and evaluation of model results.

The thesis consists of three main parts: a scientific literature review, an empirical study and its results, and formulated conclusions and recommendations. The literature review draws on both international and national sources to explore the concept of credit risk, its challenges and economic importance, the conceptual framework of quantitative risk management tools, and trends in the application of modern classification methods. The emphasis is placed on different classification methods, their advantages and disadvantages, and the importance of data feature exploration, data transformation and sample balancing.

The empirical part of the thesis is based on consumer loan data issued on the IPPT platform during the period from 2021-01-01 to 2024-01-28. The data sources included national institutions and registries, loan application information, and client behavior-related data. In total, 36 models were developed. The best-performing credit risk assessment model for individuals was selected, which was based on the XGB classification method, combined with WOE data transformation and SMOTE sample balancing techniques. The selected model achieved a predictive performance of 0.83 in terms of AUC. The empirical results revealed that the most significant predictors of credit risk level were client debt indicators and the frequency of loan inquiries. A client rating scale system was created to segment clients according to risk levels from A1 to E3, and a statistically validated threshold of 48% was identified, from which a client is assigned to the “bad” borrower category.

The empirical findings confirmed that advanced machine learning methods can ensure high accuracy in credit risk modelling. The results also highlighted that managing selection of independent variables, sample imbalance, and data transformation has a substantial impact on the model's effectiveness.

The results of this thesis may be published in scientific journals focused on individual credit risk management, machine learning applications and consumer lending markets. The practical significance of the study lies in the potential for financial institutions and IPPT platforms to apply the developed model in decision-making processes, while its theoretical contribution provides a basis for further research integrating the mortgage loan market, macroeconomic indicators, and major commercial bank data.

LITERATŪROS SĄRAŠAS

- 2024 m. gegužės 31 d. Europos Parlamento ir Tarybos reglamentas (ES) 2024/1623, kuriuo iš dalies keičiamas Reglamentas (ES) Nr. 575/2013, kiek tai susiję su kredito rizikos, kredito vertinimo koregavimo rizikos, operacinės rizikos, rinkos rizikos ir rezultatų apatinės ribos reikalavimais. (2024). <http://data.europa.eu/eli/reg/2024/1623/oj>
- Basel Committee on Banking Supervision. (2020). *CRE36 IRB approach: minimum requirements to use IRB approach*.
https://www.bis.org/basel_framework/chapter/CRE/36.htm
- Bello, O. A. (2023). Citation: Bello O.A. (2023) Machine Learning Algorithms for Credit Risk Assessment: An Economic and Financial Analysis. *International Journal of Management Technology*, 10(1), 109–133.
<https://doi.org/10.37745/ijmt.2013/vol10n1109133>
- Byanjankar, A., Heikkila, M., & Mezei, J. (2015). Predicting credit risk in peer-to-peer lending: A neural network approach. *Proceedings - 2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, 719–725.
<https://doi.org/10.1109/SSCI.2015.109>
- Boguslauskas, V., & Mileris, R. (2009). Estimation of Credit Risk by Artificial Neural Networks Models. *Izinerine Ekonomika-Engineering Economics* (Numeris 4).
- Boguslauskas, V., Mileris, R., & Adlyte, R. (2011). New internal rating approach for credit risk assessment. *Technological and Economic Development of Economy*, 17(2), 369–381. <https://doi.org/10.3846/20294913.2011.583721>
- Coolen-Maturi, T., & Coolen, F. P. A. (2019). Non-parametric predictive inference for the validation of credit rating systems. *J. R. Statist. Soc. A* (T. 182).
<https://academic.oup.com/jrsssa/article/182/4/1189/7068323>
- Croux, C., Jagtiani, J., Korivi, T., & Vulcanovic, M. (2020). Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior and Organization*, 173, 270–296.
<https://doi.org/10.1016/j.jebo.2020.03.016>
- Dastile, X., & Celik, T. (2021). Making Deep Learning-Based Predictions for Credit Scoring Explainable. *IEEE Access*, 9, 50426–50440.
<https://doi.org/10.1109/ACCESS.2021.3068854>

- Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing Journal*, 91. <https://doi.org/10.1016/j.asoc.2020.106263>
- Dėl pradžinių reikalavimų kredito įstaigoms ir investicinėms įmonėms ir kuriuo iš dalies keičiamas Reglamentas (ES) Nr. 648/2012, Europos Sąjungos oficialusis leidinys (2013). <https://eur-lex.europa.eu/legal-content/LT/TXT/HTML/?uri=CELEX%3A32013R0575>
- Dzidzevičiūtė, L. (2013). Statistinių vertinimo balais modelių taikymo Lietuvos bankuose galimybės, disertacija, Vilniaus universitetas, 237 p.
- Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.752558>
- Kang, Y., Jia, N., Cui, R., & Deng, J. (2021). A graph-based semi-supervised reject inference framework considering imbalanced data distribution for consumer credit scoring. *Applied Soft Computing*, 105. <https://doi.org/10.1016/j.asoc.2021.107259>
- Li, B. (2022). Online Loan Default Prediction Model Based on Deep Learning Neural Network. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/4276253>
- Li, Z., Tian, Y., Li, K., Zhou, F., & Yang, W. (2017). Reject inference in credit scoring using Semi-supervised Support Vector Machines. *Expert Systems with Applications*, 74, 105–114. <https://doi.org/10.1016/j.eswa.2017.01.011>
- Lietuvos Bankas. (2025). *Vartojimo kredito teikimo gairės*. https://www.lb.lt/uploads/documents/docs/40110_ac489893ecfa9bb81b3a2919e2a92f6a.pdf
- Liu, Y., Baals, L. J., Osterrieder, J., & Hadji-Misheva, B. (2024). Network centrality and credit risk: A comprehensive analysis of peer-to-peer lending dynamics. *Finance Research Letters*, 63. <https://doi.org/10.1016/j.frl.2024.105308>
- Malakauskas, A. (2023). *The evaluation of access to credit for small and medium enterprises*, disertacija, Kauno technologijos universitetas, 210 p.
- Markov, A., Seleznyova, Z., & Lapshin, V. (2022). Credit scoring methods: Latest trends and points to consider. *Journal of Finance and Data Science* (T. 8, p. 180–201). KeAi Communications Co. <https://doi.org/10.1016/j.jfds.2022.07.002>
- Mileris, R. (2009). Statistinių kredito rizikos vertinimo modelių efektyvumo analizė, disertacija, Kauno technologijos universitetas, p. 210

- Sanz-Guerrero, M., & Arroyo, J. (2024). *Credit Risk Meets Large Language Models: Building a Risk Indicator from Loan Descriptions in Peer-to-Peer Lending*. <https://ssrn.com/abstract=4979155>
- Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E. (2021). A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Applied Soft Computing*, 98. <https://doi.org/10.1016/j.asoc.2020.106852>
- Shen, F., Zhao, X., Li, Z., Li, K., & Meng, Z. (2019). A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation. *Physica A: Statistical Mechanics and its Applications*, 526. <https://doi.org/10.1016/j.physa.2019.121073>
- Song, Y., Wang, Y., Ye, X., Zaretski, R., & Liu, C. (2023). Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. *Information Sciences*, 629, 599–617. <https://doi.org/10.1016/j.ins.2023.02.014>
- Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine Learning for Enhanced Credit Risk Assessment: An Empirical Approach. *Journal of Risk and Financial Management*, 16(12). <https://doi.org/10.3390/jrfm16120496>
- Špicas, R. (2017). Statistinis mažų ir labai mažų įmonių kredito rizikos vertinimo modelis Lietuvos kredito unijoms, disertacija, Vilniaus universitetas, p. 236
- Špicas, R., Neifaltas, A., Kanapickienė, R., Keliuotytė-Staniulėnienė, G., & Vasiliauskaitė, D. (2023). Estimating the Acceptance Probabilities of Consumer Loan Offers in an Online Loan Comparison and Brokerage Platform. *Risks*, 11(7). <https://doi.org/10.3390/risks11070138>
- Tian, Y., Yong, Z., & Luo, J. (2018). A new approach for reject inference in credit scoring using kernel-free fuzzy quadratic surface support vector machines. *Applied Soft Computing Journal*, 73, 96–105. <https://doi.org/10.1016/j.asoc.2018.08.021>
- Tian, Z., Xiao, J., Feng, H., & Wei, Y. (2020). Credit Risk Assessment based on Gradient Boosting Decision Tree. *Procedia Computer Science*, 174, 150–160. <https://doi.org/10.1016/j.procs.2020.06.070>
- Tran, C. S., Nicolau, D., Nayak, R., & Verhoeven, P. (2021). Modeling Credit Risk: A Category Theory Perspective. *Journal of Risk and Financial Management*, 14(7). <https://doi.org/10.3390/jrfm14070298>
- Verbraken, T., Verbeke, W., & Baesens, B. (2013). A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE*

Transactions on Knowledge and Data Engineering, 25(5), 961–973.

<https://doi.org/10.1109/TKDE.2012.50>

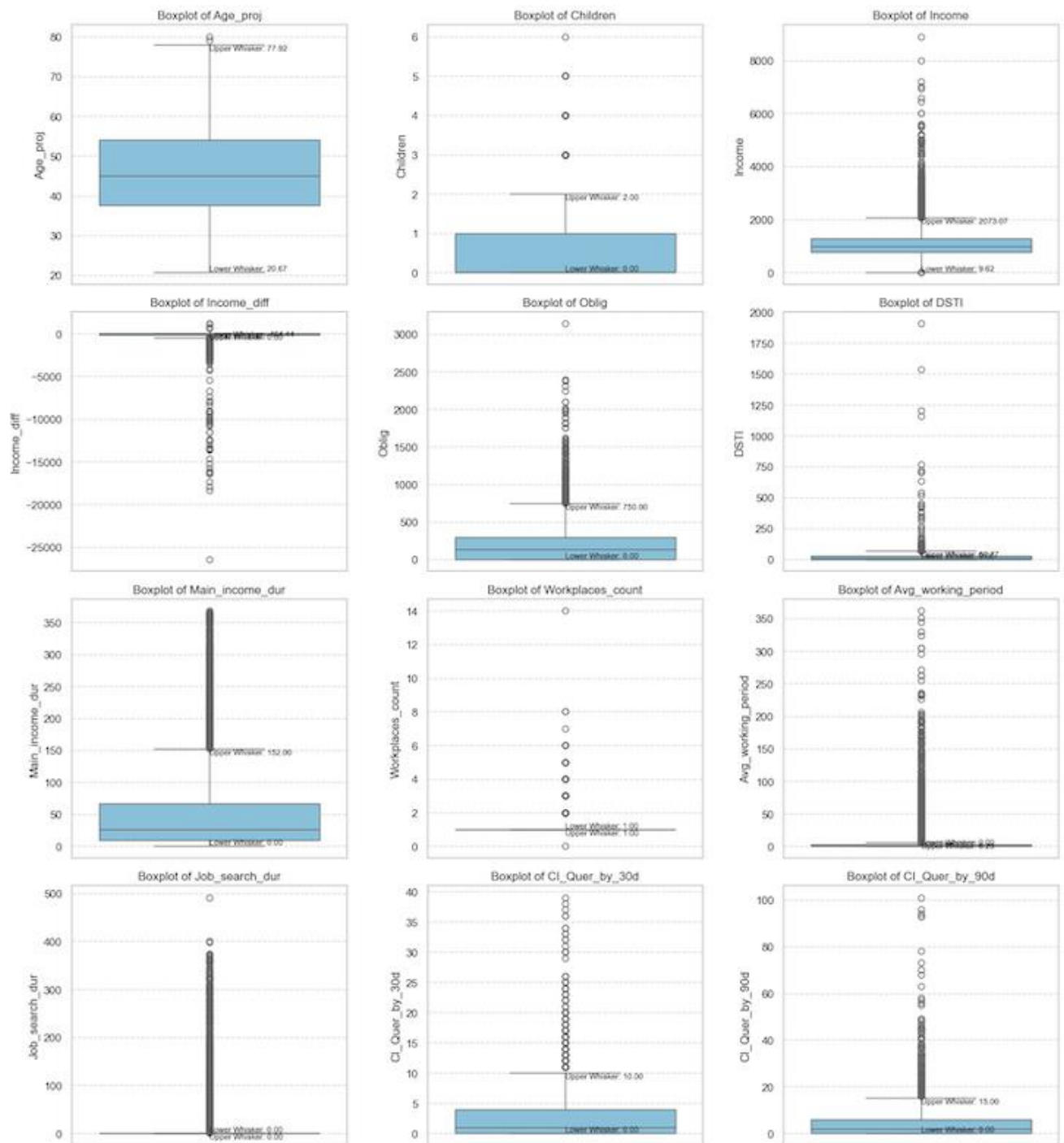
Xiao, J., Wang, Y., Chen, J., Xie, L., & Huang, J. (2021). Impact of resampling methods and classification models on the imbalanced credit scoring problems. *Information Sciences*, 569, 508–526. <https://doi.org/10.1016/j.ins.2021.05.029>

Zhang, X., & Yu, L. (2024). Consumer credit risk assessment: A review from the state-of-the-art classification algorithms, data traits, and learning methods. *Expert Systems with Applications* (T. 237). Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2023.121484>

PRIEDAI

1 priedas

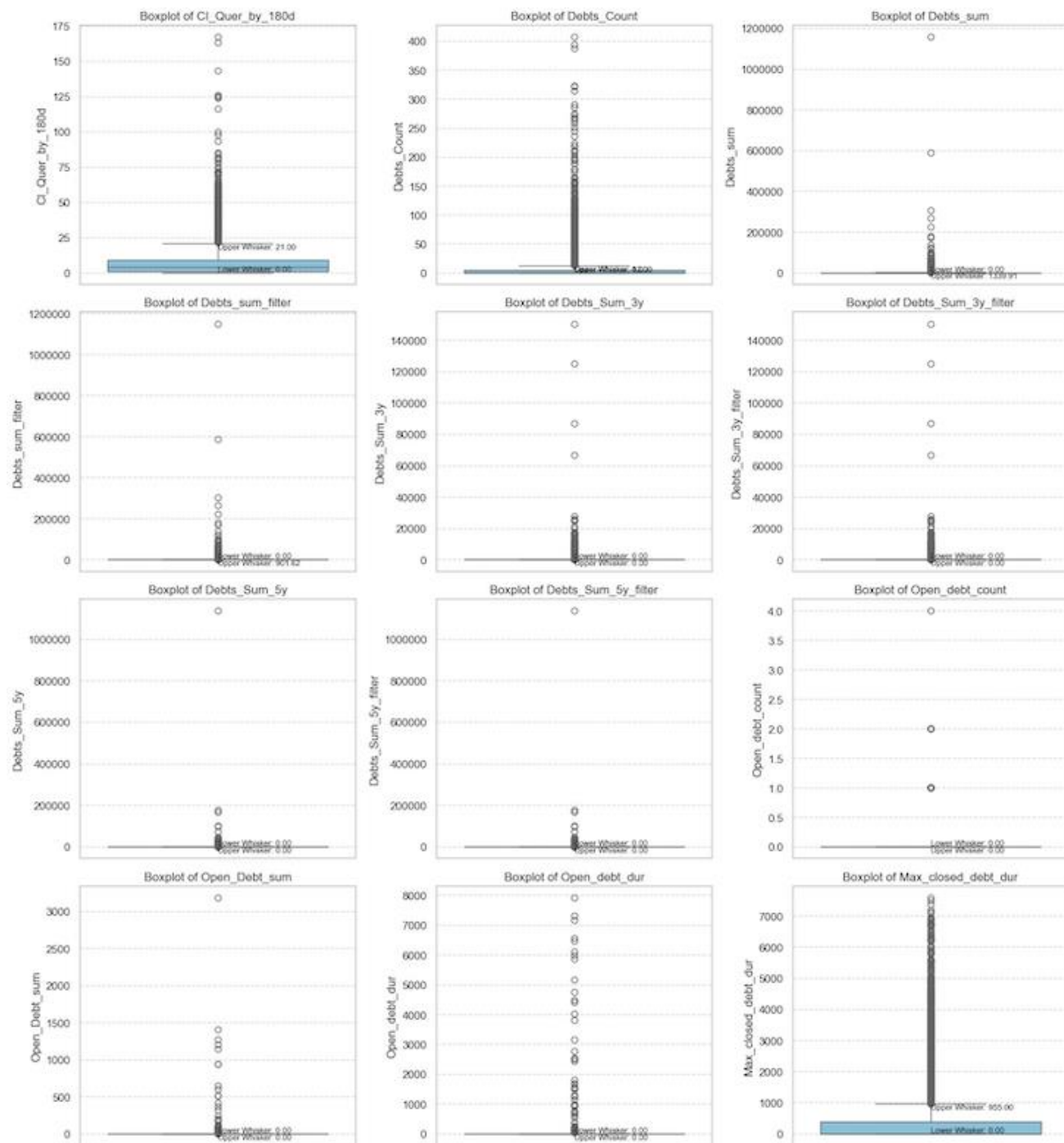
Nepriklausomų kintamųjų stačiakampės diagramos



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

1 priedo tęsinys

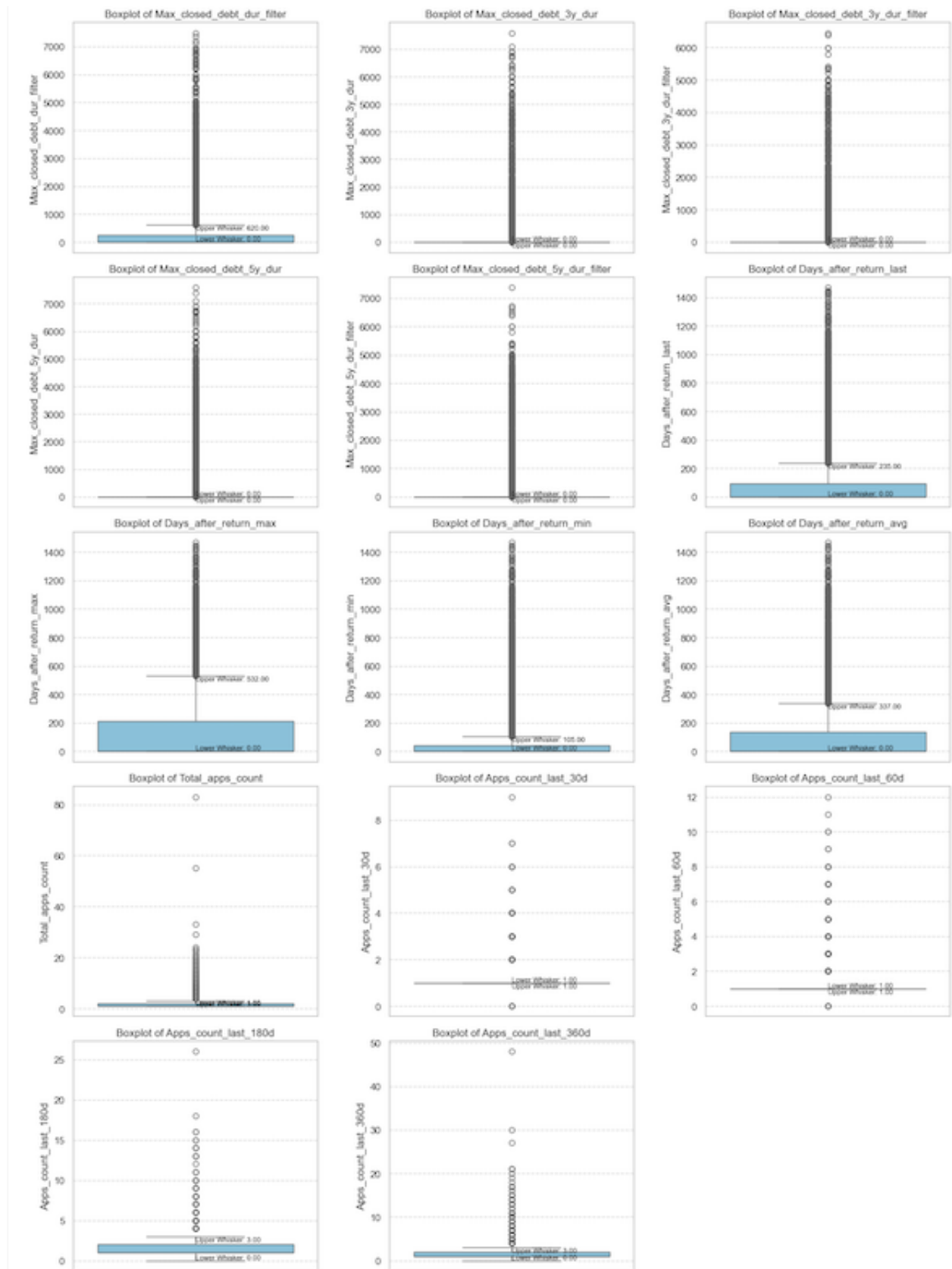
Nepriklausomų kintamųjų stačiakampės diagramos



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

1 priedo tęsinys

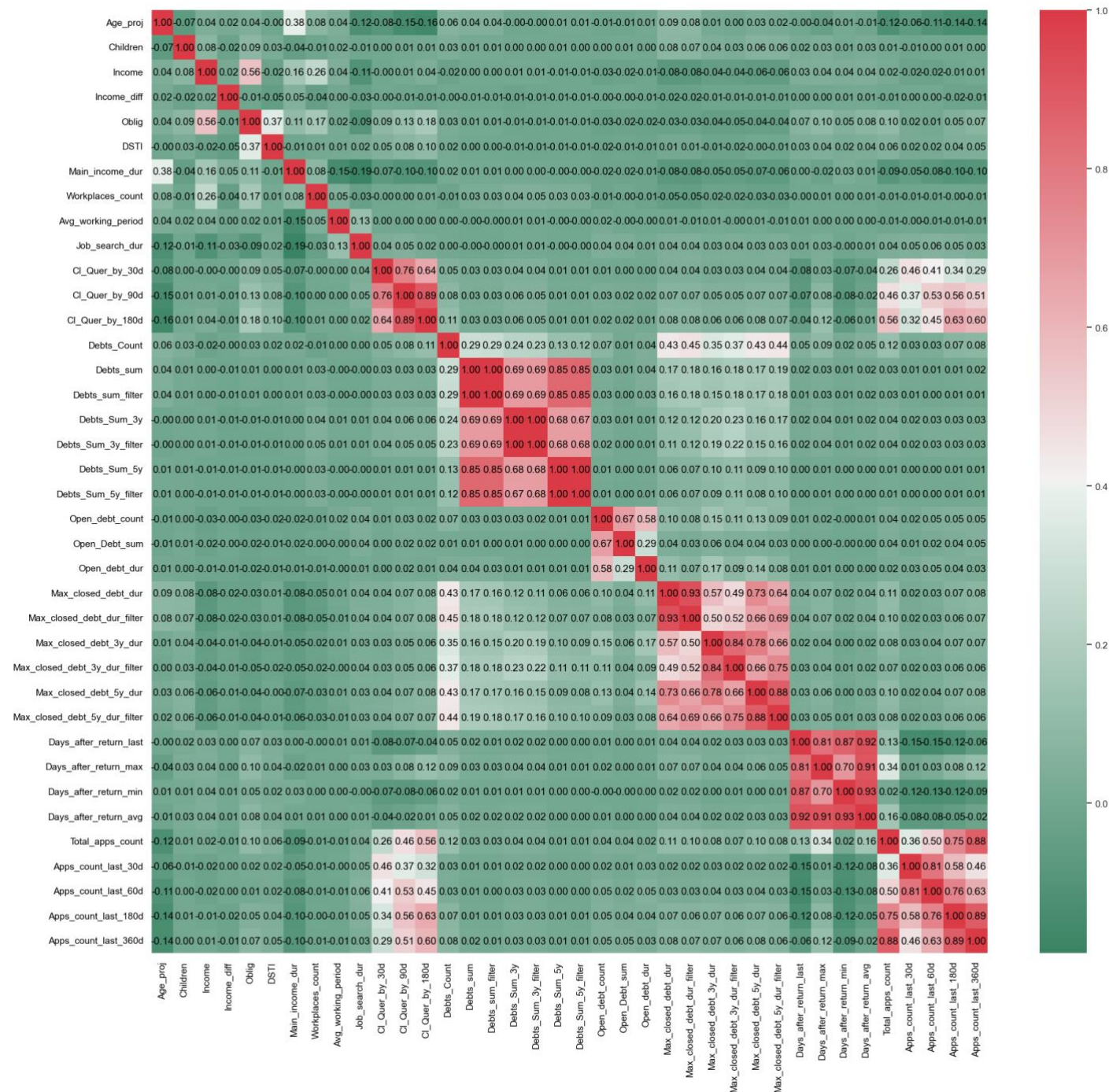
Nepriklausomų kintamųjų stačiakampės diagramos



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

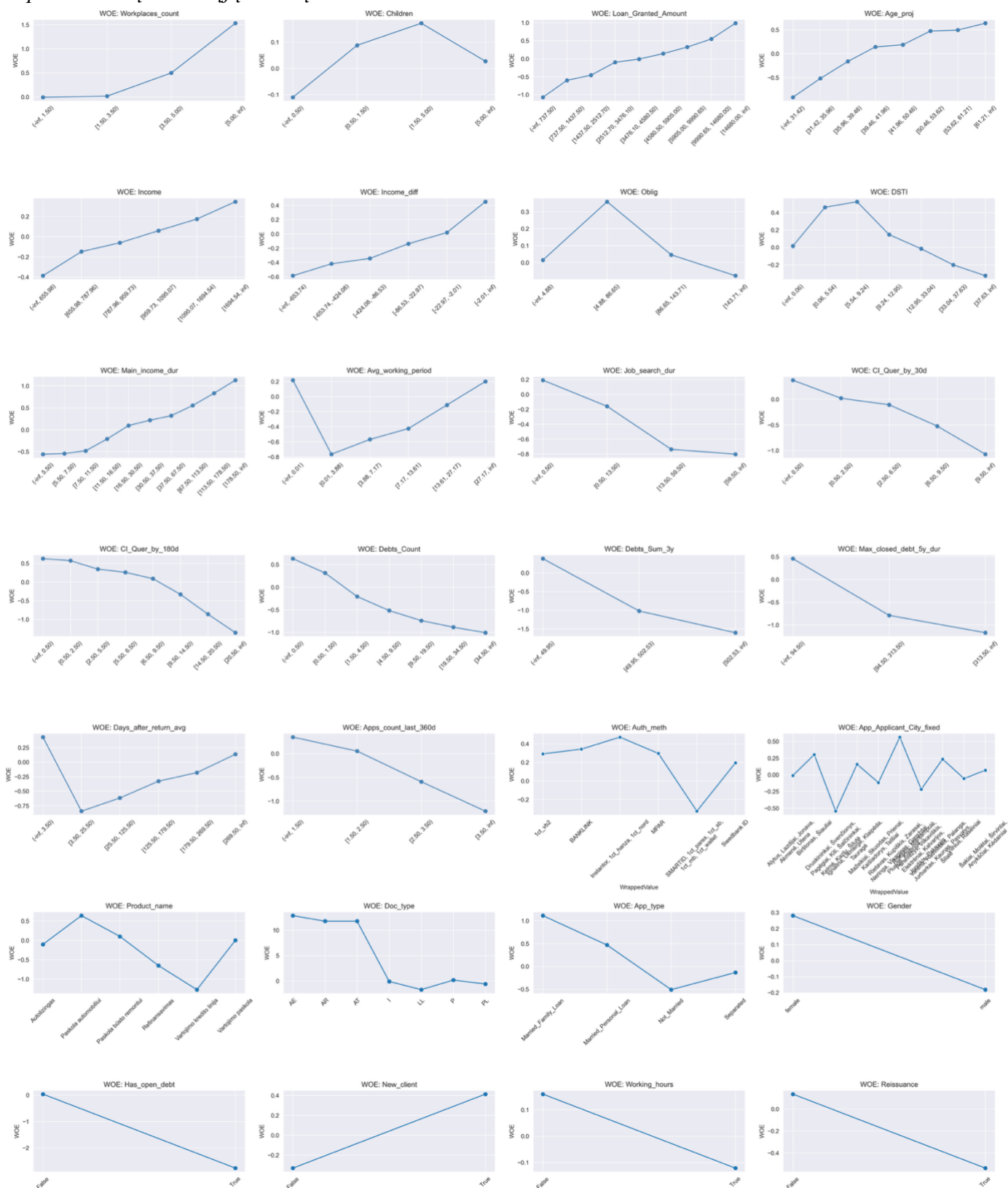
2 priedas

Pirminio nepriklausomų kintamųjų sąvado koreliacijos matrica (Kendall metodas)



3 priedas

Nepriklausomų kintamųjų WOE įverčiai



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

4 priedas

Pirmo varianto (var1) duomenų imties modelių rezultatų vertinimo reikšmės

	var1											
	Not balanced				ROS				SMOTE			
	LR	SVM	RF	XGB	LR	SVM	RF	XGB	LR	SVM	RF	XGB
AUC	0.7907	0.5586	0.8257	0.8227	0.7943	0.7742	0.8128	0.8008	0.6686	0.7034	0.7806	0.7927
GINI	0.5813	0.1172	0.6513	0.6454	0.5887	0.5483	0.6257	0.6015	0.3371	0.4069	0.5613	0.5854
Ar	0.8885	0.8855	0.8927	0.8923	0.7653	0.6444	0.7662	0.6336	0.873	0.8426	0.7937	0.6568
BAr	0.5637	0.511	0.5645	0.5988	0.7163	0.704	0.7415	0.7177	0.5402	0.5872	0.7076	0.7242
F1	0.2262	0.0498	0.2284	0.3161	0.3875	0.3331	0.4083	0.339	0.1638	0.2704	0.3965	0.3496
KS	0.4545	0.0939	0.5146	0.5112	0.4691	0.4094	0.501	0.4742	0.2529	0.3026	0.4446	0.4575
Sp	0.984	0.9956	0.9894	0.9787	0.7798	0.6268	0.7735	0.6089	0.971	0.9177	0.819	0.637
Se	0.1434	0.0264	0.1396	0.2189	0.6528	0.7811	0.7094	0.8264	0.1094	0.2566	0.5962	0.8113
PPV	0.5352	0.4375	0.6271	0.5686	0.2755	0.2117	0.2866	0.2132	0.3258	0.2857	0.297	0.2228
NPV	0.8996	0.8886	0.8996	0.9071	0.946	0.9571	0.954	0.9647	0.8947	0.9059	0.9405	0.9634
α	0.016	0.0044	0.0106	0.0213	0.2202	0.3732	0.2265	0.3911	0.029	0.0823	0.181	0.363
β	0.8566	0.9736	0.8604	0.7811	0.3472	0.2189	0.2906	0.1736	0.8906	0.7434	0.4038	0.1887

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

5 priedas

Antro varianto (var2) duomenų imties modelių rezultatų vertinimo reikšmės

	var2											
	Not balanced				ROS				SMOTE			
	LR	SVM	RF	XGB	LR	SVM	RF	XGB	LR	SVM	RF	XGB
AUC	0.8098	0.5506	0.8094	0.7016	0.8159	0.8153	0.8149	0.6992	0.6149	0.7191	0.737	0.6158
GINI	0.6196	0.1012	0.6188	0.4031	0.6318	0.6306	0.6298	0.3985	0.2298	0.4383	0.474	0.2317
Ar	0.8799	0.8855	0.8906	0.8863	0.7323	0.7036	0.7349	0.3286	0.1896	0.8314	0.7516	0.1137
BAr	0.6247	0.511	0.5386	0.5	0.7404	0.739	0.7452	0.5933	0.5346	0.6319	0.6855	0.5
F1	0.3578	0.0498	0.1472	0.0	0.3894	0.3758	0.3941	0.2407	0.2159	0.335	0.3545	0.2042
KS	0.4909	0.1143	0.4942	0.3155	0.4865	0.4874	0.5029	0.2974	0.193	0.3495	0.3756	0.1781
Sp	0.955	0.9956	0.9942	1.0	0.7299	0.6931	0.7318	0.2507	0.0881	0.8901	0.7711	0.0
Se	0.2943	0.0264	0.083	0.0	0.7509	0.7849	0.7585	0.9358	0.9811	0.3736	0.6	1.0
PPV	0.4561	0.4375	0.6471	0.0	0.2629	0.247	0.2662	0.1381	0.1213	0.3037	0.2516	0.1137
NPV	0.9134	0.8886	0.8942	0.8863	0.9581	0.9617	0.9594	0.9682	0.9733	0.9172	0.9376	0.0
α	0.045	0.0044	0.0058	0.0	0.2701	0.3069	0.2682	0.7493	0.9119	0.1099	0.2289	1.0
β	0.7057	0.9736	0.917	1.0	0.2491	0.2151	0.2415	0.0642	0.0189	0.6264	0.4	0.0

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

6 priedas

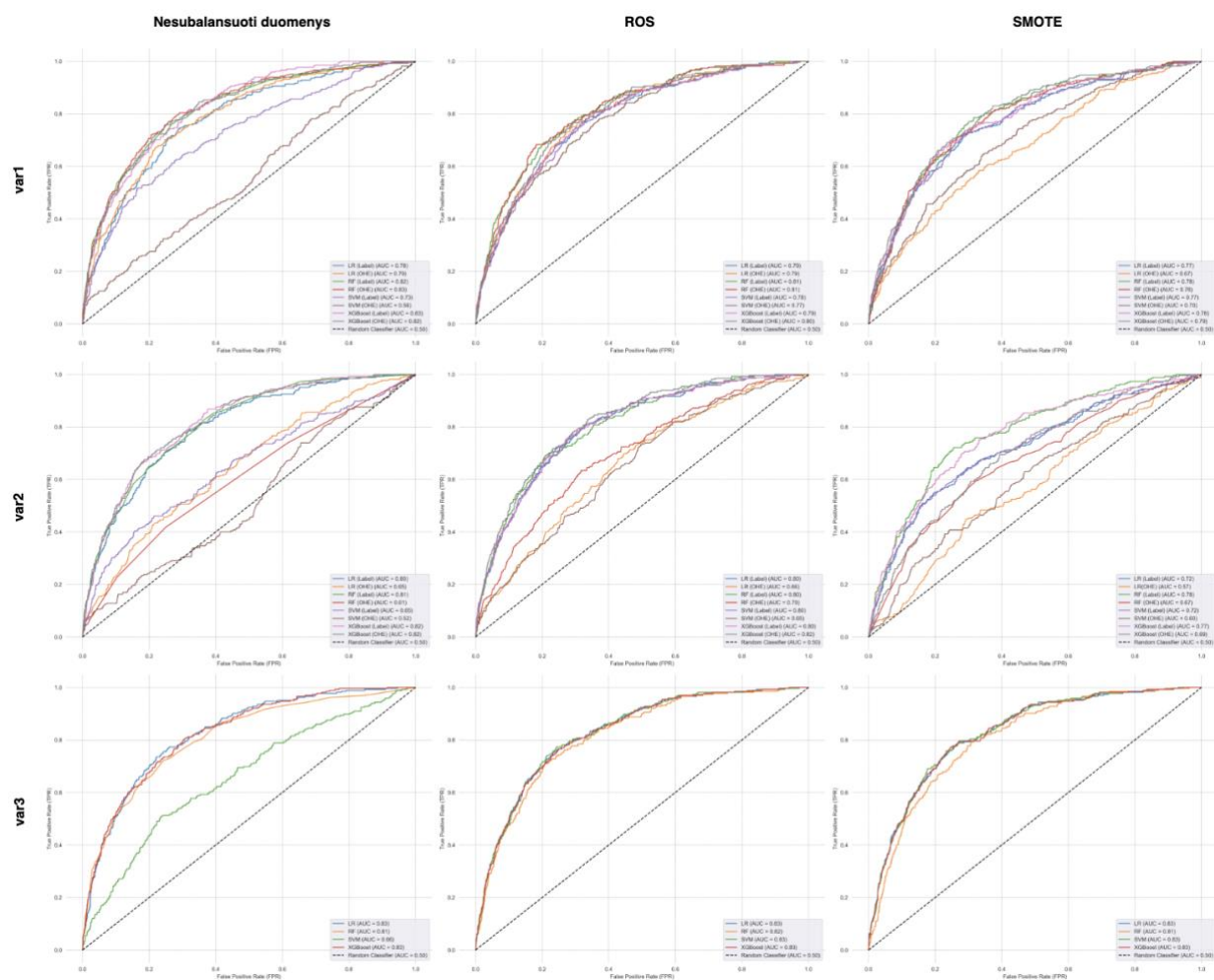
Trečio varianto (var3) duomenų imties modelių rezultatų vertinimo reikšmės

var3												
	Not balanced				ROS				SMOTE			
	LR	SVM	RF	XGB	LR	SVM	RF	XGB	LR	SVM	RF	XGB
AUC	0.7016	0.7016	0.7016	0.8261	0.8284	0.8284	0.8284	0.8275	0.8269	0.8284	0.8076	0.8284
GINI	0.4031	0.4031	0.4031	0.6523	0.6569	0.6569	0.6569	0.6549	0.6538	0.6569	0.6151	0.6569
Ar	0.8863	0.8863	0.8863	0.8936	0.7675	0.7675	0.7675	0.7671	0.7683	0.7679	0.7945	0.7675
BAr	0.5	0.5	0.5	0.5946	0.7586	0.7586	0.7586	0.7551	0.7542	0.7589	0.7229	0.7586
F1	0.0	0.0	0.0	0.3073	0.4222	0.4222	0.4222	0.4193	0.4194	0.4226	0.4108	0.4222
KS	0.3155	0.3155	0.3155	0.4992	0.5211	0.5211	0.5211	0.5187	0.5207	0.5211	0.4872	0.5211
Sp	1.0	1.0	1.0	0.9816	0.7701	0.7701	0.7701	0.7706	0.7725	0.7706	0.8156	0.7701
Se	0.0	0.0	0.0	0.2075	0.7472	0.7472	0.7472	0.7396	0.7358	0.7472	0.6302	0.7472
PPV	0.0	0.0	0.0	0.5914	0.2942	0.2942	0.2942	0.2925	0.2932	0.2946	0.3047	0.2942
NPV	0.8863	0.8863	0.8863	0.9062	0.9596	0.9596	0.9596	0.9585	0.958	0.9596	0.945	0.9596
α	0.0	0.0	0.0	0.0184	0.2299	0.2299	0.2299	0.2294	0.2275	0.2294	0.1844	0.2299
β	1.0	1.0	1.0	0.7925	0.2528	0.2528	0.2528	0.2604	0.2642	0.2528	0.3698	0.2528

Šaltinis: sudaryta autorės, remiantis atliktu tyrimu

7 priedas

Skirtingais imties variantais ir balansavimo technikomis paremtų modelių ROC ir AUC reikšmių rezultatai



Šaltinis: sudaryta autorės, remiantis atliktu tyrimu