



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

Ieva Dirdaitė

**JURIDINIŲ ASMENŲ MOKESČIŲ VENGIMO IDENTIFIKAVIMO
MODELIS**

Baigiamasis magistro projektas

Vadovai

Doc. dr. Alfreda Šapkauskienė

Doc. dr. Vytautas Janilionis

KAUNAS, 2017

KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ. FAKULTETAS

**JURIDINIŲ ASMENŲ MOKESČIŲ VENGIMO IDENTIFIKAVIMO
MODELIS**

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

Vadovai

(parašas) Doc. dr. Alfreda Šapkauskienė
(data)

(parašas) Doc. dr. Vytautas Janilionis
(data)

Recenzantai

(parašas) Doc. dr. Tomas Ruzgas
(data)

(parašas) dr. Ramūnas Mackevičius
(data)

Projektą atliko

(parašas) Ieva Dirdaitė
2017-06-02

KAUNAS, 2017

TURINYS

SANTRAUKA	5
ĮVADAS	7
1. Literatūros apžvalga	9
1.1. Mokesčių vengimo identifikavimo aktualumas.....	9
1.2. Mokesčių vengimui įtaką darantys veiksniai ir mokesčių mokėtojų charakteristikos	13
1.3. Rizikingų mokesčių mokėtojų identifikavimui taikomų algoritmų ir matematinių metodų analizė.....	18
1.4. Klasifikavimo metodų ir programinės įrangos privalumai, trūkumai ir taikymo apribojimai	26
1.5. Apibendrinimas	28
2. TYRIMŲ METODAI.....	Error! Bookmark not defined.
2.1. Tyrimo metodika	Error! Bookmark not defined.
2.2. Juridinių asmenų požymių parinkimas	Error! Bookmark not defined.
2.3. Juridinių asmenų klasifikavimo modeliai.....	Error! Bookmark not defined.
2.3.1. Atsitiktinių miškų metodas	Error! Bookmark not defined.
2.3.2. Neuroniniai tinklai	Error! Bookmark not defined.
2.3.3. Ekstremalaus gradiento sustiprinimo metodas.....	Error! Bookmark not defined.
2.3.4. Adaptyvus sustiprinimo metodas	Error! Bookmark not defined.
2.3.5. Artimiausio kaimyno metodas	Error! Bookmark not defined.
2.4. Klasifikavimo kokybės vertinimas	Error! Bookmark not defined.
2.5. Modelio programinė realizacija.....	Error! Bookmark not defined.
3. Tyrimo rezultatai ir jų aptarimas	Error! Bookmark not defined.
3.1. Tyrimo duomenys.....	Error! Bookmark not defined.
3.2. Juridinių asmenų mokesčių vengimo identifikavimo modelio taikymo rezultatai.....	Error! Bookmark not defined.
3.3. Reikšmingų kintamųjų atrankos metodų taikymo rezultatų palyginimas	Error! Bookmark not defined.
4. Išvados	Error! Bookmark not defined.
5. Literatūros sąrašas.....	Error! Bookmark not defined.
1 Priedas. Tyrimui naudojami santykiniai rodikliai	Error! Bookmark not defined.
2 Priedas. R programavimo kalbos paketai naudojami tyrimui, Programos tekstas	Error! Bookmark not defined.
3 Priedas. Prognozavimo analitikos ir mašininio mokymo programų palyginimas	Error! Bookmark not defined.
4 Priedas. KIntamųjų sklaidos diagramos ir histogramos	Error! Bookmark not defined.

5 Priedas. Mokomosios imties sumažinimo, padidinimo ir papildomų reikšmių generavimo modelių taikymo imties rezultatai**Error! Bookmark not defined.**
6 Priedas. Optimalaus slenksčio nustatymas.....**Error! Bookmark not defined.**



KAUNO TECHNOLOGIJOS UNIVERSITETAS
Matematikos ir gamtos mokslų fakultetas

(Fakultetas)

Ieva Dirdaitė

(Studento vardas, pavardė)

Didžiųjų verslo duomenų analitika, 621G12002

(Studijų programos pavadinimas, kodas)

„Juridinių asmenų mokesčių vengimo identifikavimo modelis“
AKADEMINIO SAŽININGUMO DEKLARACIJA

20 17 m. birželio 2 d.

_____ Kaunas _____

Patvirtinu, kad mano, **Ievos Dirdaitės**, baigiamasis projektas tema „Juridinių asmenų mokesčių vengimo identifikavimo modelis“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

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

(vardą ir pavardę įrašyti ranka)

(parašas)

Dirdaitė, Ieva. Juridinių asmenų mokesčių vengimo identifikavimo modelis. Magistro baigiamasis projektas / vadovai: doc. dr. Alfreda Šapkauskienė, doc. dr. Vytautas Janilionis; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Mokslų kryptis ir sritis: Fiziniai mokslai, matematika (01 P)

Reikšminiai žodžiai: mokesčių vengimas, klasifikavimo metodai.

Kaunas, 2017. 108 p.

SANTRAUKA

Mokesčių vengimas yra aktuali problema, kuri yra analizuojama mokslinėse publikacijose ir tyrimuose. Dėl nevykdomų mokesčių mokėtojų įsipareigojimų surenkama mažiau valstybės biudžeto pajamų, tai paveikia kiekvieną iš mokesčių mokėtojų. Tyrimų duomenimis, Lietuvoje yra didelis šešėlinės ekonomikos ir mokesčių vengimo mastas. Siekiant sumažinti atotrūkį ir identifikuoti mokesčių mokėjimo vengimo priežastis, taikomi prevenciniai ir kontrolės veiksmai. Šiame darbe sprendžiamas aktualus uždavinys. Darbo tiklas – sukurti modelį, kuris pagal mokesčių administratoriams prieinamus didelius duomenų kiekius padėtų identifikuoti rizikingus mokesčių mokėtojus ir automatizuotų šį procesą, tai leis tikslingiau planuoti kontrolės veiksmus ir sąlygos efektyvesnį mokesčių surinkimą.

Remiantis užsienio valstybių patirtimi ir mokslinėmis publikacijomis, nustatyti veiksniai, darantys įtaką juridinių asmenų mokesčių vengimui, sudarytas rizikingų mokesčių mokėtojų požymių (kintamųjų) sąrašas bei išanalizuoti praktiškai taikomi mokesčių vengimo identifikavimo metodai. Panaudojus sudarytą mokesčių mokėtojų požymių sąrašą, reikšmingų kintamųjų atrankos ir klasifikavimo metodus (atsitiktinių miškų, neuroninių tinklų, artimiausio kaimyno, ekstremalaus gradientinio ir adaptyvaus sustiprinimo) sudarytas juridinių asmenų mokesčių vengimo identifikavimo modelis, kuris pritaikytas praktiškai analizuojant nuasmenintus juridinių asmenų 2011 – 2015 metų laikotarpio duomenis. Sprendžiant nesubalansuotų duomenų problemą mokymo imčiai atliktos šios modifikacijos: padidinimas, sumažinimas ir papildomų reikšmių generavimas. Nustatyta, kad RFE metodas yra tinkamiausias reikšmingų požymių atrankai. Analizuotus mokesčių mokėtojus geriausiai klasifikavo ekstremalaus gradientinio sustiprinimo metodas.

Dirdaitė, Ieva. *Tax Evasion by Legal Entities Identification Model.*: Master's thesis in Business Big Data Analytics / supervisor assoc. prof. Alfreda Šapkauskienė, assoc. prof. Vytautas Janilinis. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Research area and field: Mathematics

Key words: Tax evasion, Legal entities, tax evasion identification

Kaunas, 2017. 108 p.

SUMMARY

Tax evasion is a pressing problem, which is analysed in scientific publications and researches. Because of unfulfilled obligations by some taxpayer less income is collected to the state budget, which affects all taxpayers. Research shows that Lithuania has a large shadow economy and high tax evasion rates. In order to reduce the gap and identify the causes of tax evasion, preventative and control measures are taken. The work described solves a pressing issue. The objective of the work is to create a model that would help identify risky legal entities using big data available to tax authorities and to automate this process in order to plan control measures and more effectively collect taxes.

Based on the experience of other nations and scientific publications, actions were identified that influence legal entity tax evasion, a list of features (variables) that describe risky taxpayers was created and practical tax evasion identification methods were analysed. Using the list of features, significant feature selection and classification methods (random forest, neural network, nearest neighbour, extreme gradient and adaptive strengthening) were used to create a legal entity tax evasion identification model, which was applied in practise analysing anonymised legal entity data from the 2011 - 2015 year period. To solve unbalanced training set problem these modifications were made: increase, decrease and additional entries were generated. It was found that RFE method is the best in finding significant features. Analysing tax payers the best classification method was the extreme gradient support method.

IVADAS

Mokesčių vengimas plačiai analizuojamas moksliniuose tyrimuose ir praktiniuose taikymuose. Vieni pirmųjų mokesčių vengimo problemą analizavo Allinghamas ir Sandmo (1972), kurie suformulavo proporcingo apmokestinimo ir mokesčių slėpimo modelį, pagal kurį mokesčių mokėtojas įvertina audito identifikavimo tikimybę ir pajamų sumą, kurią gali nuslėpti. Lietuvoje mokesčių vengimo mastą iliustruoja Murphey'aus (2012), Schneiderio (2015), atlikti tyrimai, kurie rodo, kad Lietuvoje yra aukštas mokesčių nesumokėjimo lygis, šešėlinės ekonomikos dydis yra vienas didžiausių Europoje.

Mokesčių surinkimas užtikrina viešųjų paslaugų prieinamumą ir palaikymą, tačiau praktikoje susiduriama su vengiančio mokėti mokesčius mokėtojo problema. Mokesčių mokėtojai suvokia mokesčių mokėjimo naudą, tačiau vengia vykdyti įsipareigojimus. Tai paveikia valstybės biudžetą, kadangi skiriamos lėšos vengiančių mokėti mokesčius mokėtojų identifikavimui, bei sukuria nelygybę tarp tų, kurie sąžiningai moka mokesčius ir tų, kurie stengiasi rasti būdų kaip išvengti apmokestinimo.

Moksliniuose straipsniuose analizuojami socialiniai, demografiniai bei finansiniai rodikliai, darantys įtaką mokesčių vengimui. Skirtingų rodiklių derinių parinkimas leidžia analizuoti mokesčių mokėtojus ir prognozuoti elgseną remiantis istoriniais duomenimis. Didėjantis kaupiamų duomenų kiekis, apibūdinantis mokesčių mokėtojus, ir pažangūs duomenų tyrybos metodai leidžia detalai analizuoti ir identifikuoti rizikingus mokesčių mokėtojus bei tiksliau atlikti patikrinimus. Todėl kyla klausimas kokius veiksnius reikia parinkti ir taikyti metodus, kurie leistų identifikuoti juridinius asmenis, vengiančius mokėti mokesčius.

Tyrimo naujumas – praktinis duomenų tyrybos metodų taikymas Lietuvos juridinių asmenų mokesčių vengimo identifikavimui, parengta metodika uždaviniui spręsti.

Pirmiausia, baigiamajame darbe analizuojama mokesčių vengimo problematika ir tyrimų būtinybė, analizuojami moksliniai tyrimai ir praktiškai taikyti matematiniai metodai, tiriant rizikingų juridinių asmenų mokesčių vengimą. Antroje dalyje apibrėžiami pagrindiniai tyrime taikomų metodų aspektai, suformuojama tyrimo metodika uždavinio sprendimui. Tyrimų dalyje analizuojamam duomenų rinkiniui pateikiami praktiškai pritaikytos tyrimų metodikos rezultatai, lyginami ir interpretuojami modelių rezultatai, formuojamos išvados.

Darbo tikslas – sudaryti Lietuvos juridinių asmenų mokesčių vengimo identifikavimo modelį.

Darbo uždaviniai:

1. Atlikti literatūros analizę ir atrinkti veiksnius (kintamuosius) darančius įtaką juridinių asmenų mokesčių vengimui;

2. Susisteminti ir parengti analizei Lietuvos juridinių asmenų atrinktų kintamųjų duomenis už paskutinių 5 metų laikotarpį;
3. Sudaryti Lietuvos juridinių asmenų mokesčių vengimo identifikavimo modelį ir įvertinti jo kokybę;
4. Pateikti pasiūlymus dėl sudaryto modelio taikymo ir jo tobulinimo.

Tyrimo tematika parengtas ir publikuotas pranešimas konferencijoje „Matematika ir gamtos mokslai: teorija ir taikymas“, 2017.

1. LITERATŪROS APŽVALGA

Šioje darbo dalyje pateikiamas problemos aktualumas, analizuojamos literatūroje nagrinėjamos mokesčių vengimo priežastys, įvertinami veiksniai ir mokesčių mokėtojų charakteristikos, kurios lemia mokesčių mokėtojų nenorą mokėti mokesčius. Analizuojami praktikoje taikomi mokesčių vengimo identifikavimo metodai, leidžiantys identifikuoti rizikingus juridinius asmenis.

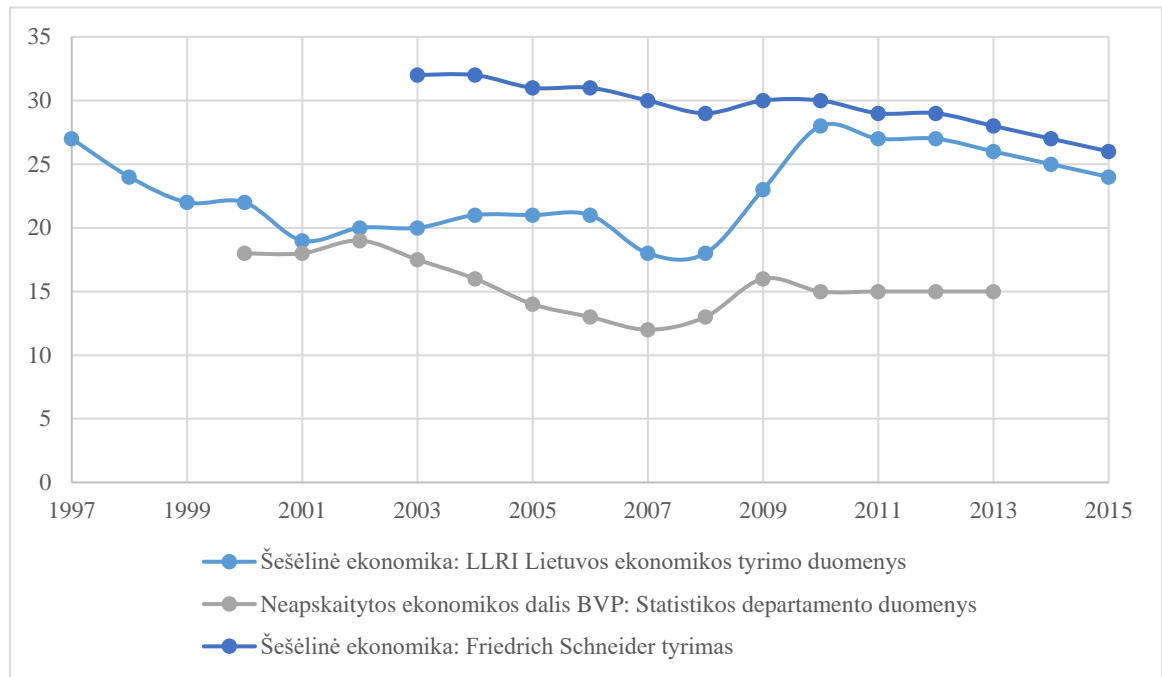
1.1. Mokesčių vengimo identifikavimo aktualumas

Pajamos iš mokesčių sudaro didžiąją Lietuvos Respublikos biudžeto dalį (apie 80 proc.), o tinkamas mokesčių surinkimas užtikrina viešųjų paslaugų prieinamumą, funkcionavimą ir socialinių paslaugų palaikymą. Visgi Lietuva kaip ir kitos valstybės susiduria su mokesčių vengimo problema. Ji yra itin aktuali, nes Schneiderio tyrimo rezultatai rodo, kad, Lietuvoje šešėlinė ekonomika 2015 metais sudarė 25,8 %, tuo tarpu, Europos vidurkis siekė 18 % [1]. Lietuvos laisvosios rinkos institutas (2016) pateikė šešėlinės ekonomikos apibrėžimą. Šešėlinė ekonomika yra ekonominė veikla, vykdoma nesilankant galiojančių teisės aktų ir jų reikalavimų, siekiant išvengti mokesčių ir/ar reguliavimų. Svarbu paminėti, jog mokesčių slėpimas yra žala ne tik valstybės biudžetui, kai valstybė negali užtikrinti teikiamų paslaugų, bet taip pat padidina išlaidas ir nelegalių veiklų identifikavimui [2,3].

Hunterio ir Nelsono nuomone [4], mokesčių slėpimo identifikavimas apima dvi funkcijas: mokesčių vengimo prevenciją (užkirsti kelią vykdyti naujus nusikaltimus) ir mokesčių vengimo identifikavimą (identifikuoti klaidingai pateikiamas deklaracijas). Paulauskas (2006) mokesčių vengimą apibrėžė kaip meną išsisukti nuo mokesčių, formaliai nepažeidžiant mokesčių įstatymų ir nesumažinant mokesčių naštos. Mokesčių vengimas ne tik paveikia valstybės biudžetą, tačiau sukuria nelygybę tarp tų, kurie sąžiningai moka mokesčius ir tų, kurie stengiasi rasti būdų kaip išvengti apmokestinimo [6].

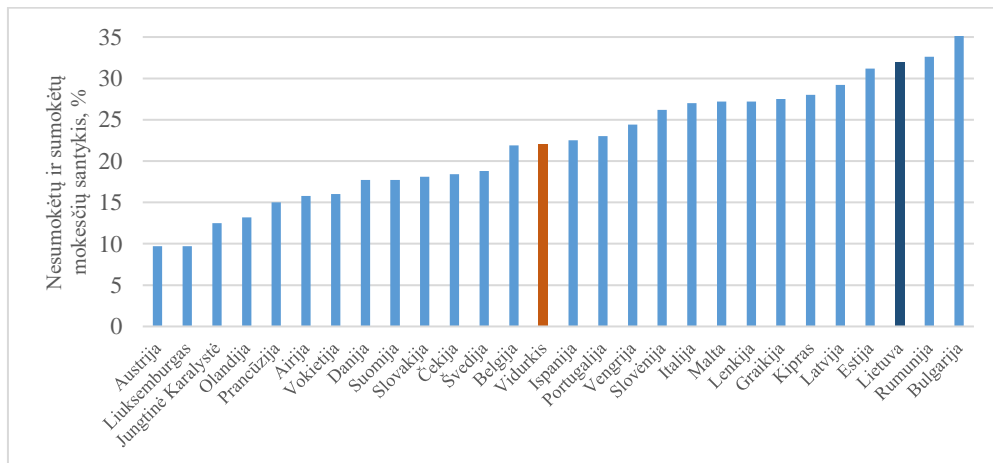
Remdamiesi šiuolaikinės ekonomikos teorija, Rutkauskas ir Ivaškaitė-Tamošiūnė (2015) teigia, kad mokesčių mokėjimas siejamas su ūkio subjekto ir valstybės įsipareigojimu. Ūkio subjektas sutinka mokėti mokesčius, o valstybė įsipareigoja teikti socialines gėrybes ir užtikrinti saugumą. Tačiau praktikoje aptinkama nemokančio mokėtojo problema, kuris suvokia mokesčių ir išlaidų surinkimo būtinybę, kuri užtikrina viešųjų paslaugų prieinamumą ir funkcionavimą, tačiau elgiasi priešingai ir vengia vykdyti įsipareigojimus.

Lietuvos laisvosios rinkos instituto ataskaitoje [1] pateikta diagrama, apibrėžianti šešėlinės ir oficialiai nepaskaitytos ekonomikos dalį procentais nuo BVP. Remiantis Lietuvos laisvosios rinkos pateiktais skaičiavimais, pastebima, kad šešėlinės ekonomikos lygis po truputį mažėja, Statistikos departamento skaičiavimai rodo, kad neapaskaitytos ekonomikos dalis išlieka pastovi (žr. 1.1.1 pav.).



1.1.1 pav. Šešėlinės ir oficialiai neapaskaitytos ekonomikos dalis procentais nuo BVP

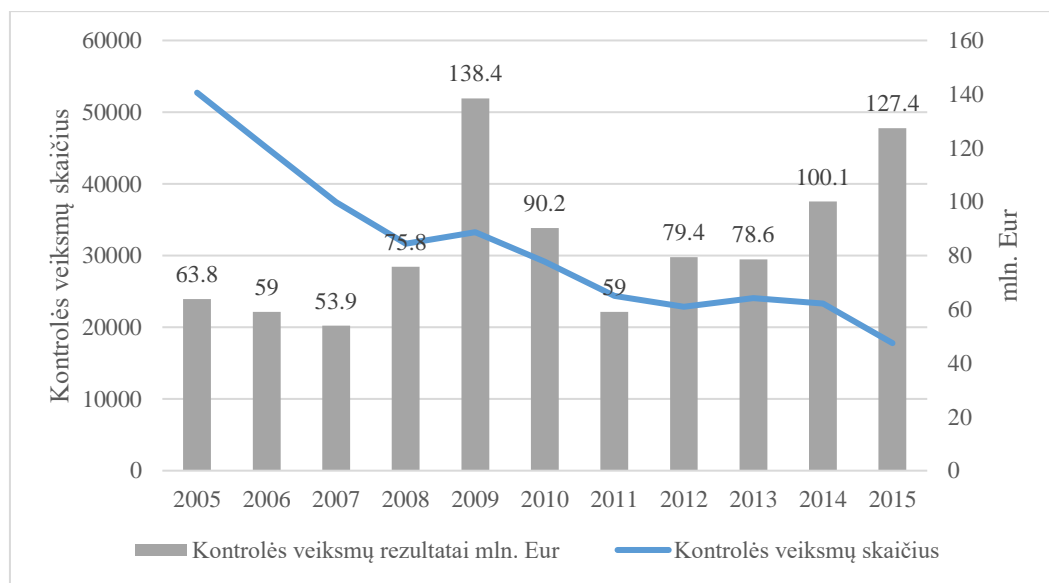
Mokesčių mokėtojų vengimą mokėti mokesčius geriausiai iliustruoja Murphy'aus (2012) atlikti skaičiavimai, kurie įvertinama nesumokėtų ir sumokėtų mokesčių santykį (žr. 1.1.2 pav.). Murphy'aus palygino Europos valstybių mokesčių nesumokėjimo mastą. 1.1.2. pav. grafikas rodo, kad Lietuvoje mokesčių vengimo lygis 2012 metais buvo 32 %, o Europos vidurkis 22,1%.



1.1.2 pav. Nesumokėtų mokesčių ir sumokėtų mokesčių santykis (Murphy, 2012)

Apskaičiuota, kad mokestinių pajamų jautrumas (angl. *tax revenue sensitivity*) Lietuvoje yra 0,26, Europos valstybių vidurkis 0,42. Mokestinių pajamų jautrumo indeksas rodo, kad Lietuvoje jautriai reaguojama į mokesstinės sistemos pokyčius ir nuolatos ieškoma naujų būdų, kaip išvengti įsipareigojimų. Europos komisijos pateikti skaičiavimai [8] rodo, kad Lietuvoje pridėtinės vertės mokesčio atotrūkis 2014 metais sudarė 36,84 %. Pridėtinės vertės mokesčio atotrūkis įvertina skirtumą tarp teoriškai apskaičiuotos mokesčių sumos ir faktiškai sumokėtų mokesčių, parodo kiek pridėtinės vertės mokesčio pajamų prarandama dėl sukčiavimo, mokesčių vengimo, bankrotų ar įmonių nemokumo.

Lietuvos Respublikos Vyriausybės kanceliarijos šešėlinės ekonomikos apžvalgoje (2016) pateikė mokesčių slėpimo apibrėžimą. Mokesčių slėpimas (angl. *tax evasion*) suprantamas kaip mokesčių mokėjimo vengimas, pasireiškiantis veikla, skirta mokesčių bazei nuslėpti, pavyzdžiui, sukčiavimas. Šešėlinės ekonomikos ir nuslėptų mokesčių mastas gali būti įvertintas analizuojant Valstybinės mokesčių inspekcijos atliktų kontrolės veiksmų skaičių ir rezultatus.



1.1.3 pav. VMI kontrolės veiksmų rezultatai (sprendimais patvirtintos mokesčių, baudų ir delspinigių bei patikslintų deklaracijų po mokestinių tyrimų ir operatyvių patikrinimų sumos) ir kontrolės veiksmų skaičius.

Analizuojant mokestinių pažeidimų duomenis, pastebėta, kad bendrai apskaičiuota papildomai į biudžetą mokėtinų mokesčių, baudų ir delspinigių suma nuo 2013 metų didėja, tuo tarpu kontrolės veiksmų skaičius mažėja. Taigi vienam kontrolės veiksmui tenkančių patvirtintų mokesčių, baudų ir sumų skaičius didėja, tai rodo didėjantį kontrolės veiksmų efektyvumą.

Mokslininkai (Mittone; Russellas, 2010; Ameer ir Thiouat, 2012) analizuoja priežastis, kurios lemia mokesčių mokėtojų nenorą laikytis įsipareigojimų. Mokesčių mokėtojų vengimą mokėti

mokesčius pirmieji pastebėjo Allinghamas ir Sandmo 1972 metais. Pasak autorių, kiekvienas mokesčių mokėtojas remdamasis neoklasikiniu ekonomikos požiūriu, atlieka kaštų ir naudos analizę. Autoriai (Allinghamas ir Sandmo) pateikė priklausomybės tarp proporcingo apmokestinimo ir mokesčių slėpimo modelį, kuris išreiškiamas formule:

$$\max_U E(U) = (1 - p)U(W - \theta X) + pU(W - \theta X - \pi(W - X)) \quad (1.1.1)$$

čia U – mokėtojo naudingumo funkcija, W – deklaruojamos pajamos, X – pajamų suma nurodyta mokesčių gražinime, $W=X$, kai mokėtojas yra sąžiningas ir $X < W$, kai nusprendžia nenurodyti visų pajamų, θ – pajamų mokesčių dydis, p – tikimybė būti patikrintam mokesčių administratoriaus. Patikrinimo atveju visa nedeklaruota suma ($W-X$) – bus nustatyta, jei mokesčių mokėtojas nuslėpė dalį pajamų, paskirtos baudos dydis bus didesnis nei mokesčio dydis θ . Prieš deklaruojant pajamas mokesčių mokėtojas pasirenka pajamų dydį X , kuris turi būti deklaruojamas, įvertinus mokesčių audito tikimybę ir galimas pasekmes. Tyrimai parodė, kad baudos padidinimas teigiamai veikia mokesčių mokėtoją ir jis labiau linksta sumokėti ir deklaruoti teisingai. Vėliau modelis (1.1.1) buvo kritikuojamas ir patobulintas, įtraukiant išorinius faktorius. 2003 metais Feldas ir Frėjus pastebėjo, kad Allinghamo ir Sandmo modelis nuvertina mokesčių vengimo ir tikrasis mokesčių mokėjimo slėpimo mastas yra daug didesnis nei rodo modelis (1.1.1). 2003 metais buvo pasiūlyta alternatyva, kurioje siūloma atkreipti dėmesį į elgesio ekonomiką ir psichologinius tyrimus, kuriose nurodoma, jog žmogus yra veikiamas vidinių motyvų [10,11]. Ekonominio bendradarbiavimo ir plėtros organizacijos (angl. *Organisation for Economic Co-operation and Development, OECD*) [12] ataskaitoje išskirtos įmonių pasirinkimą dalyvauti šešėlinėje ekonomikoje lemiančios priežastys: mokesčių mokėtojas įvertina galimybę būti mokesčių administratoriaus identifikuotam, bei galimybę sutaupyti ir nuslėpti mokesčius. Mokesčių mokėtojų įstatymų nesilaikymas dažniausiai nagrinėjamas iš fizinio asmens perspektyvos, remiantis klasikine laukiamo pelningumo prielaida (angl. *expected utility maximization*). Andreonis, Erardas ir Feinsteinas 1998 metais suformulavo prielaidą, kad mokesčių mokėtojas nevykdys įsipareigojimų tol, kol nevykdymo kaštai neviršys prievolių nevykdymo privalumų [13]. Vadinasi, mokesčių mokėtojas yra linkęs įvertinti galimą riziką būti identifikuotam ir naudą, kuri egzistuoja, jei mokesčių administratorius nepastebi piktavališkos veiklos.

Mokesčių vengimas yra aktuali problema. Mokslininkų atlikti tyrimai rodo, kad Lietuvoje egzistuoja aukštas šešėlinės ekonomikos lygis, susiduriama su mokėtojų nenoru vykdyti įsipareigojimus. Taigi svarbu identifikuoti veiksniai, kurie rodytų galimą mokesčių vengimą.

1.2. Mokesčių vengimui įtaką darantys veiksniai ir mokesčių mokėtojų charakteristikos

Didėjantis susidomėjimas elgsenos ekonomika ir taikymu mokesčių administravimo srityje, geresnis ir sklandesnis suvokimas apie mokesčių mokėtojų motyvus ir jų požiūrį gali padėti išsiaiškinant mokėtojų taisyklių laikymosi priežastis. Taip pat geresnis priežasčių suvokimas ir požiūris į apmokestinimą gali padėti suvokti kas galėtų padidinti įstatymų laikymąsi ir mokesčių surinkimo efektyvumą. Svarbu nustatyti veiksnius, kurie lemia mokesčių mokėtojų pasirinkimą dalyvauti nelegalioje veikloje ir nevykdyti įsipareigojimų.

Atlikus mokslinės literatūros analizę, galima išskirti tokius socialinius, elgsenos veiksnius, kurie daro įtaką mokesčių vengimui.

1.2.1 lentelė. Mokesčių mokėtojų socialiniai, elgsenos veiksniai, darantys įtaką mokesčių vengimui

Autorius	Veiksniai
Beron (1988)	Demokratijos palaikymas, pasitikėjimas valdžia.
Torgler (2006)	Stipri koreliacija tarp mokesčių moralės ir mokesčių įsipareigojimų laikymosi tiek besivystančiose, tiek išsivysčiusiose valstybėse. Mokesčių moralė yra svarbus determinantas šešėlinėje ekonomikoje ir taip pat turi įtakos mokesčių vengimui.
Šinkūnienė (2009)	Mokesčių moralė.
Russell (2010)	Teisingumas gali būti vienas faktorių, kurie daro įtaką mokesčių mokėtojui. Išskirtos 3 teisingumo rūšys: pasiskirstymo (mokesčių administratorius veikia kaip geras saugotojas, elgiasi išmintingai), procedūrinis (laikomasi procedūrų, susijusių su mokesčių mokėtojais), baudžiamoji – (suvokiama, kad taikomos bausmės, kai taisyklės yra pažeistos).
Ameur ir Tkiouat (2012)	Vengimo priežastys: techninės – sudėtinga mokesstinė sistema ir įstatymų taikymas, klaidos dėl blogai interpretuojamų įstatymų; politinės priežastys – skirtingi įstatymų taikymai asmenų ar įmonių grupėms; moralinės ir psichologinės priežastys – mokesčių mokėtojo elgesys priklauso nuo mokesčių administratoriaus vertinimo ir pačio asmens moralinių įsitikinimų.
Walsh (2012)	Penki faktoriai lemiantys mokesčių mokėtojų elgesį: atgrasymas, socialinės ir asmeninės normos, teisingumas ir pasitikėjimas mokesčių administravimu, mokesstinės sistemos sudėtingumas, valdžios vaidmuo ir ekonominė aplinka. Asmeninės nuostatos svarbesnės nei socialinės.
Pantoja, Rodrigas (2014)	Egzistuoja stilizuoti faktai mokesčių mokėtojas elgiasi taip kaip leidžia moralės principai. Mokesčių vengimas taip pat pasireiškia dėl kitų mokesčių mokėtojų. Tarkime, sukčiauja, nes turtingas taip pat sukčiauja.
Putniūša, Saukabas (2014)	Pasitikėjimas valdžia ir mokesstinė sistema, neigiamai susiję su įmonės polinkiu vykdyti nelegalią veiklą, tokiu atveju, įmonės mažiau linkusios slėpti.
Stankevičius ir Leonas (2015)	Mokesčių moralė, aptikimo tikimybė, bauda, baimė, sistemos sudėtingumas, kontroliuojančios institucijos palaikymas.
Rutkauskas ir Ivaškaitė-Tamošiūnė (2015)	Religingumas, nacionalinis pasididžiavimas, požiūris į korupciją, pasitikėjimas valdžia.

Apibendrinti mokslinių tyrimų rezultatai (žr. 1.2.1 lent.) rodo, kad mokesčių mokėtojų pasirinkimą elgtis teisingai ir laikytis įstatymų, labiausiai lemia mokesčių moralė, pasitikėjimas valdžia ir asmeninės nuostatos bei normos. Tačiau analizuojami ne tik socialiniai – ekonominiai veiksniai, bet ir individualaus asmens charakteristikos.

Mokslinėje literatūroje akcentuojama, kad svarbu analizuoti juridinio asmens vadovo charakteristikas, kadangi tai sprendimus priimantis asmuo, galintis daryti didelę įtaką įmonės veiklai ir pasirinkimams [14]. Be to, Stankevičius ir Vasiliauskaitė (2014) siūlo atkreipti dėmesį į kapitalo struktūrą, kuri rodo efektyvų įmonės valdymą, tai yra asmenis, kurie gali prisiimti įsipareigojimus dėl sukčiavimo. 1.2.2 lentelėje pateikiama išanalizuota apibendrinta mokslinių publikacijų santrauka, kurioje įvardijamos žmogaus asmeninės charakteristikos, kurios gali rodyti potencialų mokesčių vengimą.

1.2.2 lentelė. Mokesčių mokėtojo asmeninės savybės, lemiančios mokesčių vengimą

	Jackson ir Millironas (1986)	Beron (1988)	Torgler (2006)	Richardsin (2006)	Russell (2010)	Demir (2011)	Gupta ir Gili (2012)	OECD (2012)	OECD (2013)	Pickhardt (2014)	Rutkauskas ir Ivaškaitė-Tamošiūnė (2015)	Stankevičius ir Leonas (2015)
Amžius	+	+	+	+	+			+		+	+	+
Lytis	+	+	+	+	+	+		+	+	+		+
Išsilavinimas	+	+		+	+		+	+	+	+		+
Gyvenamoji vieta: miestas, kaimas	+											
Pajamų lygis	+			+								
Finansinė būklė	+		+	+			+		+	+	+	+
Šeimyninė padėtis		+	+			+						+
Nuosavo turto turėjimas		+										
Gimimo vieta		+										
Rasė		+										
Religija			+					+	+		+	
Baudos, nusižengimai						+						
Asmenybė										+		

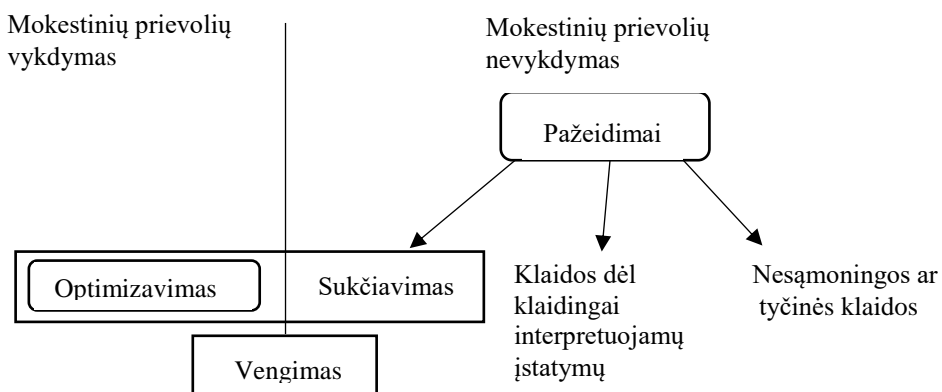
Atsižvelgiant į 1.2.2 lentelėje pateiktus duomenis, svarbu analizuoti mokesčių mokėtojo amžiaus, lyties, išsilavinimo charakteristikas. Straipsnių autoriai (Beron, 1988; Russell, 2010; Demir, 2011) pastebėjo, jog vyresni mokesčių mokėtojai yra labiau linkę laikytis taisyklių nei jaunesni, kadangi jauni yra labiau linkę rizikuoti ir nebijo būti identifikuoti. Išsilavinimas yra vienas svarbiausių veiksnių, OECD pateiktoje ataskaitoje [24] išsilavinęs žmogus yra vertinamas kaip mažiau linkęs rizikuoti, tuo tarpu, Fjeldstadas ir Heggstadas [26] teigia, jog labiau išsilavinę gali gerai išmanyti įstatymus ir lengvai pastebėti spragas. Pastebima, kad žemesnis išsilavinimas apsunkina sugebėjimą suprasti su mokestine sistema susijusius įstatymus ir procedūras. Demiro (2011) atliktame tyrime, apibendrinęs apklausų rezultatus padarė išvadas, jog moterys yra labiau linkusios laikytis įstatymų ir mokėti mokesčius, susituokę asmenys turi mažesnę mokesčių vengimo riziką nei vieniši asmenys.

Moksliniuose darbuose [13,14] svarbus veiksnys yra mokėtojo vyresnis amžius, kuris yra mažiau rizikingas.

Tarptautinės apklausos (OECD, 2013) rezultatai parodė, kad mokesčių mokėtojai viešai deklaruojantys religiją, turi teigiamą požiūrį į mokesčių mokėjimą; moterų mokestinė moralė yra aukštesnio lygio nei vyrų; vyresni žmonės yra mažiau linkę apgaudinėti nei jauni asmenys. Darbuotojai, dirbantys ne visą darbo dieną ir užsiimantys individualia veikla, turi žemesnę mokesčių moralę nei pilną darbo dieną dirbantys darbuotojai. Apibendrinant apklausos rezultatus, pastebėta, kad didžiausią įtaką mokesčių moralei turi demokratijos palaikymas (12,4 %), amžius (7,5 %), pasitikėjimas valdžia (5,5 %), lytis – moteris (3,5 %), religija (2,7 %), įgytas išsilavinimas (1,5 %) nuo respondentų skaičiaus [24]. Pastebėta, kad didesnis mokesčių laikymasis pastebimas vietovėse su mažesniu išsilavinimo lygiu ir vyresne populiacija ir populiacija, kur šeimoje vadovavo moteris. [14] Sociodemografiniai rodikliai yra reikalingi, nes padeda paaiškinti atrodytų netinkamus rezultatus literatūroje ir pateikia svarbios informacijos dėl ateities tyrimų.

Mokesčių mokėtojų socialiniai – elgsenos veiksniai, vadovo asmeninės savybės gali padėti identifikuoti ar sukurti įmonės rizikingumo profilį. Mokslinėje literatūroje, susijusioje su mokesčių vengimo problematika, dažniausiai akcentuojamos įmonės rizikingumo charakteristikos. Ameuras ir Tkouatas atliktame atvejo analizės tyrime [6] pastebėjo, kad mokesčių mokėtojų sukčiavimo veiklos identifikavimas yra vienas sudėtingiausių etapų visoje mokesčių administravimo srityje. Todėl svarbus tinkamų rodiklių ir parametrų rinkinys, kuris leistų identifikuoti sudėtingiausius sukčiavimo atvejus.

Svarbu akcentuoti ne tik veiksnius, kurie lemia mokesčių mokėtojų įsipareigojimų vykdymą, bet ir analizuoti mokėtojų elgesį ir galimas mokesčių vengimo schemas. Straipsnyje [6] analizuojama Prancūzijoje (angl. *The Council of compulsory contributions of France*) sudaryta mokestinių prievolių vengimo schema (1.2.1. pav.).



1.2.1 pav. Mokesčių vengimo ir prievolių vykdymo schema

Išskirtos dvi mokestinių prievolių vengimo strategijos (1.2.1 pav.): vengiantys vykdyti mokestines prievoles ir vykdantys mokestines prievoles. Mokestinių prievolių vykdymo atveju, legaliais būdais stengiamasi sumažinti mokestines prievoles, pasinaudojant įstatymų spragomis. Antruoju mokestinių prievolių nevykdymo atveju, vykdomi sąmoningi veiksmai, mokesčių planavimas, sąmoningai daromos klaidos, kurios lemia prievolių nevykdymą.

Išanalizavus mokslinę literatūrą, galima apibendrinti mokslininkų analizuojamus veiksnius, susijusius su juridiniais asmenimis, vengiančiais mokėti mokesčius.

1.2.3 lentelė. Juridinius asmenis, vengiančius mokėti mokesčius, apibūdinantys veiksniai

	Beron (1988)	OECD (2004)	Lin, Lin, Wu, Yang ir Roan (2010)	Kotsiantis, Koumanakos, Tzelepis	Russell (2010)	Ravisankar, Ravi, Rao ir Rose	Gupta ir Gili (2012)	Putniņa ir Saukab (2014)	Fiscalis (2016)	Dias, Pinto, Batista, Neves (2016)
Veiklos rūšis, sektorius	+	+			+			+	+	
Įmonės amžius		+			+			+	+	
Įmonės dydis		+						+		
Ilgalaikio ir turto santykis				+		+	+			
Turto apyvartumas				+			+			+
Turto pelningumas						+	+			
Trumpalaikio turto ir turto santykis						+	+			
Grynujų pinigų ir turto santykis							+			
Nepaskirstyto pelno ir turto santykis							+			
Bendrojo pelno ir turto santykis							+			
EBIT ir turto santykis				+			+			+
Pelningumo koeficientas							+			
Bendrasis pelningumas							+			
Veiklos pelningumas						+				
Finansinis svertas				+						
Altmano z įvertis				+		+	+			
Sąnaudos		+	+					+		
Pardavimo pajamų ir ilgalaikio turto							+			
Grynojo pelno ir turto santykis						+	+			
Einamojo likvidumo rodiklis			+	+		+				+
Bendrasis likvidumas										+
Skolos santykis			+				+			
Apyvartinis kapitalas ir turtas iš viso				+			+			
Skolos ir turto santykis				+						
Mokestinė grąža										+
Pelno mokestis			+							
Bendrasis pelnas (nuostoliai)			+			+				
Grynasis pelnas (nuostoliai)			+			+				
Tipinės veiklos pelnas (nuostoliai)						+				
Tipinės veiklos pelno ir pardavimo pajamų						+				
Atsargų ir pardavimo pajamų santykis			+	+			+			
Atsargų ir turto santykis						+	+			
Atsargų ir trumpalaikių įsipareigojimų santykis						+				

Remiantis analizuotomis mokslinėmis publikacijomis, 1.2.3 lentelėje pateikta agreguota informacija apie veiksnius, kurie rodo juridinių asmenų vengiamą mokėti mokesčius. Straipsnio [2] autorių nuomone, sudarant rizikingų ir įtartinų mokėtojų sąrašus, svarbu atkreipti dėmesį ne tik į tiesiogiai deklaruojamus dydžius, bet ir į santykinius rodiklius. Tiesioginiai dydžiai gali būti nepakankamai informatyvūs, kadangi neatskleis įmonės galimybių ir rizikų. Taip pat ir daugelio analizuotų straipsnių autoriai (žr. 1.2.3 lent.) siūlė taikyti išvestinius ir santykinius rodiklius, o ne absoliučius dydžius.

Allinghamas ir Sandmo (1972) pastebėjo, kad egzistuoja ryšys tarp praeities veiksmų, dabarties ir esamos situacijos. Šios dienos sprendimas yra paveiktas praeities deklaracijų, taip pat sprendimus sukčiauti ateityje lemia esama situacija. Tokiu atveju sudarant stochastinius modelius, turėtų būti įtrauktas baudos kintamasis, bei nagrinėjami istoriniai duomenys [11]. Pantoja 2014 metais parengtame straipsnyje [17] išsiaiškino, jog mokesčių mokėtojas elgiasi taip, kaip leidžia moralės principai, mokesčių vengimas taip pat pasireiškia ir dėl kitų mokesčių mokėtojų veiksmų. Vadinas, reikėtų analizuoti ne tik konkretaus mokesčio mokėtojo charakteristiką, tačiau atsižvelgti ir į supančią aplinką, bei galinčius daryti įtaką asmenis. Kitame darbe [13] siūloma nagrinėti panašaus pobūdžio įmones, nes jų elgsena galimai bus sąlygota kito mokėtojo. Straipsnyje [19] siūloma išsiskiriančias rodiklių reikšmes laikyti rizikingomis, jei priežastis negali būti paaiškinta.

Nuoseklus ir atidus finansinių duomenų analizavimas gali parodyti: įmonėms sekasi ar įmonės patiria krizę. Įmonei esančiai krizėje, nagrinėjant dokumentus galima bandyti suprasti krizės priežastį. Vertinant finansinių ataskaitų duomenis galima įvertinti skirtingus finansinius rodiklius: likvidumo, saugumo, pelningumo, efektyvumo. Tačiau finansinės ataskaitos gali būti netikslios, kai siekiama paskolos, norint pagerinti finansinius rodiklius [29]. Tuo tarpu Mostak Ahamedo atlikti tyrimai [9] parodė, kad įmonės finansinė plėtra sumažina įmonės mokesčių vengimą, išorinių finansinių šaltinių (kreditų) paieška yra per brangi įmonėms su aukštesne mokesčių vengimo veikla. Mostak Ahamedas suformavo prielaidą, kad įmonės norinčios gauti papildomą finansavimą, yra mažiau linkusios sukčiauti.

Mokesčių mokėtojas, norėdamas slėpti mokesčius, gali nepateikti arba pateikti nepilną informaciją. Todėl rizikos, kurias administratorius galėtų stebėti registruojantis mokėtoju yra kitų mokėtojų egzistavimas tuo pačiu adresu, žinomas veiklos vykdymo adresas [30]. Straipsnių [30, 33] autoriai, pasiūlė į kuriamus modelius įtraukti trečiųjų šalių informaciją, pavyzdžiui, bankų, laisvai prieinamos informacijos internete duomenis. Autoriai pastebi, kad trečiųjų šalių šaltinių įterpimas pagerina kuriamų modelių tikslumą.

Veiksnių ir rodiklių, kurie leistų identifikuoti rizikingus mokėtojus, paieška yra pastoviai vykstantis ir sudėtingas procesas. Daugelis valstybių remiasi kitų šalių gerąja praktika, tačiau tik šio metodo naudojimas yra ydingas, kadangi neįvertina turimų duomenų resursų ir nesuteikia naujų galimybių ieškant naujų rizikų [33]. Vis labiau akcentuojama ne tik finansinių, bet ir kokybinių rodiklių svarba ir nauda dėl naujų išvalgų ir neįprastų sprendimų. Mokslinėje literatūroje analizuojamos įmonės vadovo charakteristikos, įmonių rodikliai, ekonomiškai pagrįstiems veiksniams taikomi matematiniai metodai, leidžiantys identifikuoti mokesčių mokėtojus, kurie vengia mokėti mokesčius.

1.3. Rizikingų mokesčių mokėtojų identifikavimui taikomų algoritmų ir matematinių metodų analizė

Šioje dalyje apžvelgiami rizikingų mokesčių mokėtojų identifikavimo (rizikos balų priskyrimo) ir duomenų tyrybos metodai. Duomenų tyrybos metodai dar nėra plačiai taikomi mokesčių administratorių veikloje, viešai publikuojami pavieniai straipsniai apie taikytus metodus ar gautus rezultatus. Daugeliu atveju, rezultatai pristatomi praktiniuose seminaruose, kurių informacija nėra viešai publikuojama dėl modelių taikymo tikslo ir duomenų konfidencialumo.

Mokestiniai auditai yra pagrindinė mokesčių užtikrinimo priemonė visose valstybėse, dažnesni auditai leidžia sumažinti mokesčių vengimą [15]. Vienas iš rizikos nustatymo būdų – atsitiktiniai auditai, kurie praplečia nagrinėjamų rizikų sąrašą. Praktiniai pavyzdžiai rodo, kad atsitiktiniai auditai yra vertingi dėl naujų rizikos požymių suradimo ir įvertinimo. Dažniausia mokesčių administratoriai sudaro atskiro mokesčių mokėtojo rizikos profilio įvertinimą. Nors vertinamas atsitiktinių auditų privalumas, tačiau metodas kritikuojamas dėl rezultatų neapibrėžtumo, atsirenkant mokesčių mokėtojus atsitiktinai, iš anksto negalima spręsti apie audito efektyvumą [30, 33]. Apskritai, rizikos valdymas yra nuolatinis procesas – modeliavimo ciklas apimantis rizikos identifikavimą, rizikos analizę, rizikos įvertinimą ir prioritetų suteikimą, sprendimo būdų paiešką ir modelio įvertinimą [30].

Sėkmingas audito parinkimas yra vienas svarbiausių rizikos nustatymo metodų. Planuojant auditą siekiama maksimizuoti audito rezultatus ir minimizuoti audito sąnaudas [34]. Remiantis klasikiniu požiūriu, auditas atliekamas suteikus mokesčių mokėtoju rizikos įvertį. Rizikos balo įvertis rodo mokesčių mokėtojo galimą mokesčių vengimo riziką. Šis įvertis yra paremtas mokesčių mokėtojo savybėmis: dydžiu, ekonominės veiklos sritimi, mokestinių įsipareigojimų vykdymu ir kitais. Taip pat įtraukiamos ir anksčiau įgautos žinios iš mokestinių patikrinimų. Rizikos įverčių priskyrimo metodas yra priimtinas daugeliui praktikų, kadangi aiškiai apibrėžiamos taisyklės, pagal kurias vertinamas ir priskiriamas rizikos balas [33]. Tačiau šis metodas, nepalankiai vertinamas

mokslininkų dėl galimo tyrėjų subjektyvumo [29]. Šiuolaikinis audito atrankos procesas remiasi duomenimis, išskiriami du pagrindiniai šaltiniai: anksčiau atliktų patikrinimų rezultatai, bei individualių mokėtojų duomenys, dažniausiai – ekonominiai rodikliai: sumokėtų mokesčių suma, grąžinimai [33].

Mokesčių galimo slėpimo prevencijai taikomi duomenų tyrybos metodai. Pagal turimos ir kaupiamos informacijos pobūdį, išskiriamos kelios tyrybos metodų rūšys. Metodai su mokytoju, taikomi esant tikslo kintamajam, kuris gali būti prognozuojamas statistiniais metodais. Metodai be mokytojo, pavyzdžiui, klasterizavimas, taikomi siekiant išsiaiškinti galimas išskirtis ar neleistiną ar netipišką elgesį. Mokesčių mokėtojo elgesys dažnai lyginamas su vidutiniu standartiniu mokesčių mokėtojo elgesiu. Vertinant taikomų metodų ir modelių efektyvumą, svarbūs rodikliai yra ne tik atliktas didelis kontrolės veiksmų (auditų) skaičius, bet tie veiksmai, kurie buvo tikslingi ir galėjo turėti poveikį. Svarbu kiek buvo efektyvių auditų, kiek rizikingų mokėtojų pavyko identifikuoti [30, 34].

Analitinės įžvalgos apie mokesčių mokėtojų elgesį ir efektyvesnį mokesčių surinkimą yra aktualios temos mokesčių administratoriams. Ilgą laiką buvo naudojami metodai, remiantis ekspertine patirtimi. Neseniai pradėta taikyti metodus, kurie leidžia automatizuoti procesus, lengviau identifikuoti rizikingus mokėtojus. Šiuo metu bandoma taikyti sudėtingesnius metodus, kurie gautų tikslesnius rezultatus.

1.3.1 lentelė. Klasifikavimo metodų taikymas rizikingų juridinių asmenų identifikavimui

	Bonchi, Gianotti, Mainetto ir Pedreshi (1999)	Kotsiantis, Koumanakos, Tzelepis ir Tampakos (2006)	Fiscalis ataskaita (2010)	Lin, Lin, Wu, Yang ir Roan (2010)	Zhou, Kapoor (2011)	Kallo ir Back (2011)	Ravisankar, Ravi, Raghava, Rao ir Bose (2011)	Ameur ir Tkiouat (2012)	Gupta ir Gili (2012)	Walsh (2012)	Hsu, Nismith, Srivastava, Tschida, Bjorklind (2015)
Sprendimų medis	+	+		+	+			+	+		
Neuroniniai tinklai			+	+	+	+	+				+
Bajeso		+			+				+		+
Genetinis programavimas									+		
Artimiausių kaimynų metodas		+									
SOM						+					
Logistinė regresija					+		+			+	
MultiBoosting											+
SVM		+		+							+

Turint istorinius duomenims, bei taikant analitinius metodus gali būti sudaryti algoritmai ir modeliai, leidžiantys prognozuoti elgesį ar mokėtojo veiksmus. Praktikoje rizikingų juridinių asmenų identifikavimo uždaviniai sprendžiami taikant klasifikavimo metodus, kur priklausomas kintamasis rodo juridinio asmens rizikingumą. 1.3.1 lentelėje pateikti apibendrinti praktiškai taikytų klasifikavimo metodai. Lentelėje 1.3.1 pateikti rezultatai rodo, kad dažniausiai mokslinėje literatūroje analizuojami sprendimų medžio, neuroninių tinklų ir Bajeso metodai.

Bonchis, Giannottis, Mainettas ir Pedreschis [34] praktiškai taikė sprendimų medžio metodą atrenkant rizikingus mokesčių mokėtojus. Sprendimų medžio metodas pasirinktas dėl metodo paprastumo ir lengvo rezultatų interpretavimo. Tyrimo tikslas identifikuoti mokesčių mokėtojus, kuriems turi būti atliktas auditas bei maksimizuoti galimą gauti pinigų sumą atlikus patikrinimą. Tyrimo metu duomenų rinkinys yra suskaidytas į mokymo ir testavimo imtis, prieš tai duomenų parengimo žingsnyje pašalinti įrašai, turintys praleistų reikšmių. Modelio tyrimo imtį sudaro 4103 ekspertiskai atrinkti mokesčių mokėtojai, kuriems buvo atliktas auditas, duomenys. Sudarant modelį įvertintos audito išlaidos ir gauta papildoma pinigų suma atlikus auditą. Tyrimo metu autoriai pastebėjo, kad kainos kintamojo įtraukimas, leido sudaryti tinkamesnį modelį ekonomine prasme. Tyrimo autoriai pastebėjo, kad rezultatai gerėja didinant mokomosios imties dydį, kurioje yra daugiau istorinių audito duomenų.

Duomenų tyrybos metodų svarbą ir taikymą mokesčių administratoriaus veikloje rodo atliktas Minesotos atvejo analizės tyrimas [37]. Mokslininkai taikė klasifikavimo metodus sudarant rizikingų mokėtojų sąrašus. Tyrimo tikslas mokėtojų identifikavimas pardavimo mokesčių srityje, kai reikia įvertinti tikėtiną mokestinių kontrolės veiksmų efektyvumą. Analizuojama tyrimo imtis sudaryta apibrėžiant taisykles ir ekspertiniu būdu parenkant didžiausias pastoviai tikrinamas įmones, didžiausias įmones valstijoje ir pagal audito atrankos procesą. Ekspertiškai nustatyta suteikti auditui požymį sėkmingai identifikuotas auditas, kai papildomai nustatyta suma didesnė nei 1500 dolerių, priešingu atveju auditas laikoma nesėkmingu. Duomenų rinkinys suskirstytas į tris dalis: mokymo, tikrinimo ir testavimo. Tyrimui naudota per 200 kintamųjų, kurių didžioji dalis atrinkta ekspertiniu būdu. Dalis kintamųjų susiję su įmonės charakteristikomis: įmonės registravimo data, geografinė vieta, verslo tipas. Modeliavimui pasirinkti MultiBoosting ir Naive Bayes metodai. Buvo analizuotos ir atraminių vektorių ir neuroninių tinklų metodų galimybės. Modelių sudarymui taikytas kryžminio patikrinimo metodas. Tyrimo rezultatai parodė, kad daugelis klasifikavimo modelių buvo tikslesni didelėms įmonėms ir mažiau tikslūs mažesnėms įmonėms, todėl sudaryti atskiri modeliai pagal įmonių dydį. Rezultatų efektyvumas įvertintas ir palygintas skaičiuojant ROI įvertį, kuris apibrėžia

santykį tarp papildomai gautų pajamų audito metu ir išlaidų auditui atlikti. Gauti rezultatai rodo, kad taikant klasifikavimo metodus pavyko identifikuoti 72,3 % sėkmingų auditų didelių įmonių atveju ir 56,3 % mažų įmonių atveju.

Dias, Pinto ir kt. (2016) atlikto mokesčių mokėtojų rizikingumo įvertinimą, naudojant finansinių ataskaitų ir mokesčių administratoriaus disponuojamus duomenis. Portugalijos mokslininkai analizavo kasybos ir gamybos sektoriaus įmones, apskaičiavo pelningumo, likvidumo, mokestinius ir aktyvią ekonominę veiklą rodančius rodiklius. Straipsnyje pastebima, kad finansinė situacija gali būti atvaizduota naudojant finansinius rodiklius, kurie dažnai naudojami bankroto tikimybės modeliuose. Tiriama duomenų imtis sudaryta naudojant 2010 – 2011 metų finansinius duomenis, įtrauktos aktyvios įmonės (pardavimai neneigiami), apyvarta didesnė nei 100000€, bet mažesnė nei 1000000€. Apskaičiavus tiesinės koreliacijos koeficientus, pašalinti 3 kintamieji. Gautajam duomenų rinkiniui pritaikytas k-vidurkių metodas, išskirti 3 klasteriai: mažos, vidutinės ir aukštos rizikos įmonės [31].

Ameuras ir Tkiouatas (2012) nagrinėjo metodus, galinčius padėti identifikuoti rizikingus mokėtojus. Straipsnyje autoriai pateikia metodiką ir tyrimo planą, kuris turėtų būti įgyvendintas. Dėl konfidencialumo, straipsnyje pristatomi rezultatai, gauti atsitiktinai sugeneravus duomenų rinkinį, turintį 3500 įrašus. Analizuojama situacija, kai kiekvienas įrašas atspindi juridinį asmenį iš kurių 500 buvo atrinkti auditui. Modeliavimui naudotos apyvartos, pelno mokesčio ir sumokėtų mokesčių reikšmės Rizikingų mokesčių mokėtojų identifikavimui taikytas sprendimo medžių metodas. Gautas modelis, leidžiantis identifikuoti rizikingas įmones ir jų polinkį sukčiauti. Analizuojamas duomenų rinkinys buvo suskaidyti į mokymo ir testavimo imtis. Rezultatų tikslumas įvertintas suskaičiuojus klaidingai suklasifikuotų įrašų skaičių. Testinės imties atveju 1,7 % įrašų buvo klaidingai suklasifikuota.

Gupta ir Gilis suklasifikavo įmones į rizikingas ir nerizikingas pagal finansinių santykinių rodiklių reikšmes, susijusias su pelningumu, likvidumu ir einamuoju efektyvumu. Nagrinėtos įmonės, kurios dalyvauja akcijų biržose, privalo pateikti finansines ataskaitas, pinigų srautų duomenis kas ketvirtį ir metus. Analizuoti finansinių ataskaitų santykiniai rodikliai siekiant apskaičiuoti finansinės apgaulės tikimybę. Tyrimui pasirinkta 114 įmonių, kurios yra akcijų biržose, iš kurių 85 nėra pateikusios finansinių ataskaitų. Atrinktos 29 įmonės, kurios nusižengė ir buvo patikrintos. Apskaičiuoti 52 finansiniai rodikliai kiekvienai įmonei: likvidumo, saugumo, pelningumo ir efektyvumo. Rodikliai normalizuoti, modeliavimui taikytas kryžminio patikrinimo metodas su 10 blokų, atliktas modelio tikrinimas. Požymių parinkimui, Gupta ir Gilis taikė vienfaktorinę dispersinę analizę, kurios tikslas buvo sumažinti dimensiją ir patikrinti ar egzistuoja statistiškai reikšmingi

skirtumai tarp dviejų klasių mokesčių mokėtojų, įvardintų kaip vengiančių ir nevengiančių mokėti mokesčius. Vienfaktorinė dispersinė analizė leido išskirti 22 informatyvius ir statistiškai reikšmingus kintamuosius, kurie toliau buvo panaudoti modelių sudarymui. Autoriai pasirinko taikyti sprendimų medžio, Bajeso ir genetiniais algoritmais grįstą klasifikavimą. Gupta ir Gilis išskyrė klasifikavimo medžių metodo privalumą, nes nereikia iš anksto atlikti kintamųjų atrankos, metodas pats geba susidoroti su šia užduotimi. Tačiau, sprendimų medis yra labai jautrus išskirtims mokymo imtyje. Autoriai pastebėjo, kad finansinių duomenų analizėje tai yra gana dažnas reiškinys dėl finansinių krizių. Autoriai analizuoja Bajeso metodų taikymo prielaidų apribojimus dėl kintamųjų nepriklausomumo, todėl daugeliu atveju metodas negali būti taikomas. Sudarytų modelių: sprendimų medžių, Bajeso ir genetinių algoritmų tikslumas vertinamas skaičiuojant specifiškumo ir jautrumo metrikas. Palyginus rezultatus gauta, jog tiksliausiai klasifikuoja sprendimų medžio metodas (jautrumas 86,2 %, specifiškumas 100 %), blogiausiai – genetiniai algoritmai (jautrumas 53%, specifiškumas 99,2 %). Gupta ir Gilio nuomone, modelių tikslumas galėtų būti didesnis, jei būtų įtraukti ne tik finansinių ataskaitų, bet ir kokybiniai kintamieji.

Kitame praktiniame pavyzdyje [28] analizuojami 2001-2002 laikotarpio Graikijos įmonių duomenys ir naudojamas modelis skirtas neteisingai pateiktų deklaracijų identifikavimui. Duomenų rinkinį sudaro 164 įmonių, prekiaujančių akcijų biržoje, duomenys, neįtraukiant finansinių institucijų. Ekspertai patikrino 41 įmonę, šios įmonės laikomos teisingai pateikusias deklaracijas, likusios 123 įmonės – klaidingai pateikusios deklaracijas. Pirmiausia atrinkti 24 kintamieji, norint įsitikinti, kad rodikliai yra reikšmingi, ranguotos kintamųjų reikšmės pagal jų įtaką. Rangavimas atliktas skaičiuojant relief reikšmę, kuri įvertina kintamųjų įtaką ir svarbą. Straipsnio autoriai (Kotsiantis, Koumanakos, Tzelepis, Tampakas) pastebi, jog nepakanka taikyti vieną klasifikavimo metodą, rekomenduojama įtraukti daugiau metodų ar naudoti jų kombinacijas, siekiant padidinti modelio tikslumą, sumažinti klaidos tikimybę. Uždavinio sprendimui taikyti sprendimų medžio, artimiausio kaimyno, Bajeso, atraminių vektorių (SVM) metodai. Tiksliausias iš taikytų metodų buvo sprendimų medžio metodas, kurio tikslumas 91,2 %. Taikant skirtingų metodų kombinacijas, pasiektas tikslumas 95,1 % [28].

Mokslininkai (Lin, Lin, Wu, Yang ir Roan, 2010) sukūrė pelno siekiančių juridinių asmenų pajamų mokesčio vengimo identifikavimo modelį. Straipsnio autoriai [38] naudojo 17 kintamųjų sąrašą, pasiūlytų auditorių ekspertų (žr. 1.2.3 lent.). Modelio sukūrimui naudoti 5 metų duomenys, duomenų rinkinys suskaidytas į mokymo, tikrinimo ir testavimo imtis, kintamųjų reikšmės

standartizuotos, taikyti sprendimų medžio ir neuroninių tinklų metodai. Tyrimo rezultatai parodė, kad sukurtas neuroninių tinklų modelis yra tikslesnis nei sprendimų medis.

Suomijoje buvo sukurtas modelis, kurio sudarymui panaudoti 2014 juridinių asmenų audito duomenys, į eksperimentą neįtrauktos vieno asmens ar užsienio įmonės. Nagrinėtą imtį sudaro 4355 įmonės, kur 107 (2,4 %) įmonės buvo patikrintos ir mokesčių suma buvo koreguota. Modelio kintamieji parinkti bendradarbiaujant su ekspertais. Iš viso parinkti 8 kintamieji (atlyginimai, skolos, ir t.t.). Visais atvejais nagrinėtas santykis su apyvarta, kad būtų sumažinta įmonės dydžio įtaka. Apdorotas duomenų rinkinys suskaidytas į mažesnes imtis ir sudaryti SOM žemėlapiai. Gauti rezultatai leido išskirti klasterius, į kuriuos pateko įmonės, kurioms buvo atliktas auditas. Autoriai nurodė, kad pasirinktas metodas davė gana tikslius rezultatus, tačiau pasiekti geresnius rezultatus sutrukdė netiksliai sudaryta mokomoji imtis, kai į imtį pateko labai nedidelis tikrintų įmonių skaičius ir daug didesnis patikrintų audito. Autoriai kelia prielaidą, kad toks rezultatas gautas dėl neteisingai parinktų kintamųjų [36].

Praktikoje sprendžiant klasifikavimo uždavinius susiduriama su klasių disbalanso problema, kai prognozuojamo kintamojo reikšmių, rodančių požymio egzistavimą, skaičius yra nedidelis [39, 40, 41, 42]. Nesubalansuotumo problemos sprendimas reikalauja kūrybiškumo ir įvairių klasifikavimo metodų išmanymo [41, 42]. Siekiant sumažinti klasių disbalansą, naudojami metodai: priklausomo kintamojo reikšmių papildomas generavimas (angl. *oversampling*) ar sumažinimas (angl. *undersampling*).

Q.Jangas ir X.Wu ėmėsi iniciatyvos identifikuojant 10 sudėtingiausių problemų duomenų tyrybos srityje. Apibendrinę apklausų rezultatus jie išsiaiškino, kad viena iš dešimties problemų yra darbas su nesubalansuotais ir nestatiniiais bei ekonomiškai jautriais duomenimis [27]. Mokesčių vengimo identifikavimo atveju, nesubalansuotų duomenų problema yra itin aktuali.

Šeifertas, Košgoftaras, Hulse ir Napolitanas pateikė du būdus duomenų nesubalansuotumo problemai spręsti, tai duomenų modifikavimas arba sustiprinimo metodų taikymas. Pirmuoju atveju taikomas priklausomo kintamojo įvykių skaičiaus didinimas arba mažinimas, tai leidžia sumažinti modelio sudarymo laiką, tačiau prarandama dalis informacijos. Antruoju atveju, duomenų nesubalansuotumo problemai spręsti taikomi AdaBoost, RUSBoost, SMOTEBoost, RUS, SMOTE metodai. Vienas populiariausių yra AdaBoost metodas. Tyrimais įrodyta, jog šis metodas leidžia padidinti modelio tikslumą. Dėl metodo specifikos, kiekvienoje iteracijoje stengiamasi apmokyti silpniausią klasifikatorių, AdaBoost metodas yra tinkamas dirbant su nesubalansuotais duomenų rinkiniais, kadangi suteikiamas didesnis svoris priklausomo kintamojo įvykių klasei, tai leidžia

pagerinti modelio prognozės tikslumą [43]. Esant geram dalykinės srities supratimui gali būti taikomi klasifikavimo metodai, kurie naudoja svorių arba klaidos vertės matricas. Šiuo atveju įvedamos baudos už blogą klasifikavimą. Siekiama, kad nuostoliai būtų minimalūs. Straipsnio [41] autoriai pastebėjo, kad nėra reikšmingo rezultatų tikslumo pagerėjimo, tarp 10 ir 50 iteracijų. Atlikti empiriniai tyrimai su 4 metodais ir 14 duomenų modifikavimo būdų bei 1500 duomenų rinkiniais parodė, kad metodai, naudojantys duomenų modifikavimą, pavyzdžiui, papildomų reikšmių kūrimą ir sustiprinimo metodų taikymą yra tikslesni nei tik modelių taikymas modifikuotiems duomenims.

Nesubalansuotų duomenų atvejo analizė taikyta Ispanijos mokesčių administratoriaus veikloje, siekiant identifikuoti pridėtinės vertės mokesčio karuselės atvejus. Atlikti tyrimai pagrindė prielaidą, jog klasifikatoriai yra tikslesni, kai taikomi duomenų rinkinio modifikavimo metodai, taip pat taikant sustiprinimo metodus, gaunamas didesnis prognozės tikslumas. Tyrimo metu gauta išvada rodo, kad dirbant su nesubalansuotais duomenimis ir norint aptikti sukčiavimo atvejus, didžiausias modelio tikslumas gaunamas taikant sustiprinimo metodus, kai pradinio duomenų rinkinio priklausomo kintamojo įvykių skaičius yra sumažintas [60]. Straipsnio [61] autoriai taikė atsitiktinių miškų algoritmą bioinformatikos nesubalansuotų duomenų prognozavimui. Atliktų skaičiavimų metu gautos išvados, kad tinkamiausiems duomenų rinkinio balansavimo metodas, kai santykis yra 50:50. Nors logistinė regresija yra viena populiariausių klasifikavimo metodų, tačiau metodas yra nerekomenduotinas, kai klasės yra nesubalansuotos, kadangi retų įvykių klasės yra blogai klasifikuojamos [40].

Įprasti klasifikavimo metodų klasifikavimo kokybės įverčiai ne visada tinka dirbant su nesubalansuotais duomenimis. Straipsnyje [39] rekomenduojama nesubalansuotų duomenų rinkinio klasifikavimo tikslumą įvertinti skaičiuojant jautrumo ir specifiškumo reikšmių geometrinį vidurkį. Prognozavimo tikslumas (angl. *Accuracy*) yra vienas populiariausių klasifikavimo metodų įvertinimo rodiklių, tačiau nesubalansuotų duomenų atveju, klasifikavimo kokybės įvertis gali būti netinkamas, kadangi nesubalansuotos imties atveju neteisingai identifikuoto požymio kaina skiriasi. ROC kreivė yra laikoma tinkama metrika modelių klasifikavimo kokybės įvertinimui. Pagrindinis tikslas naudojant nesubalansuotus duomenis yra pagerinti atsako (angl. *recall*) įvertį ir nedarant įtakos preciziškumas (angl. *precision*) [44]. Tuo tarpu straipsnyje [42] akcentuojama, kad reikia vertinti ne tik ROC kreivę, bet ir kitas metrikas, suradus optimalaus slenksčio (angl. *threshold*) reikšmę. Slenkstis – reikšmė, pagal kurią nusprendžiama kokia bus priklausomo kintamojo reikšmė. Sprendžiant klasifikavimo uždavinius su subalansuotu duomenų rinkiniu dažniausiai taikoma slenksčio vertė yra 0,5, kuri daugeliu atveju yra nepriimtina analizuojant nesubalansuotus duomenis. Įprasti

klasifikavimo algoritmai veikia netinkamai nesubalansuotiems duomenų rinkiniams. Nesubalansuotų duomenų atveju vertinant klasifikavimo kokybės rodiklius, modelio tikslumas bus artimas 1, kadangi modelis gerai identifikuos priklausomo kintamojo didesnės klasės reikšmes, mažesnės klasės įvykiai bus blogai identifikuojami. Atliktas tyrimas parodė, kad apskaičiuota didelė ploto po ROC kreive reikšmė – AUC, nebūtinai lemia didelę harmoninio vidurkio tarp atsako ir tikslumo prognozės tikslumo F1 reikšmę. Todėl buvo pasiūlyta parinkti slenksčio vertę testavimo imčiai. Straipsnio [42] autoriai kryžminio patikrinimo metodą, skirtingas slenksčio vertes, stengėsi įvertinti prognozės reikšmių pasiskirstymą.

Kintamųjų parinkimo bei duomenų nesubalansuotumo problemų sprendimas dažnai ženkliai pagerina klasifikavimo metodų tikslumą. Kintamųjų parinkimas yra sudėtingas uždavinys, kai yra daug kintamųjų ir kintamuosius sieja stiprus tarpusavio ryšys. Šio uždavinio sprendimui [45] straipsnyje siūloma taikyti rekursinį požymių eliminavimo algoritmą (angl. *recursive function elimination*), kuris įvertina kintamųjų kombinacijas. Dideliuose duomenų rinkiniuose, ypač esant didesniai stebėjimų skaičiui nei stebinių, ne visi kintamieji yra reikšmingi. Didelis kintamųjų skaičius gali būti nenaudingas, pavyzdžiui, turėti neigiamos įtakos modelio tikslumui. Kintamųjų parinkimo ar požymių atrankos metodai leidžia sumažinti kintamųjų skaičių, palengvina modelio interpretavimą, gali padidėti modelio prognozės tikslumą. Guyonas ir Eliseffas išskiria tris požymių parinkimo būdus:

- 1) filtravimas – atrenkamas kintamųjų poaibis duomenų paruošimo etape, nepriklausomai nuo naudojamo prognozavimo kintamojo ar metodo;
- 2) įterpimas (angl. *embedded*) - kintamųjų parinkimas modelio apmokymo procese, rezultatas priklauso nuo pasirinkto modelio.
- 3) detalaus analizavimo (angl. *wrapper*) – įvertinami visi kintamieji ir jų svarba visiems poimčiams [47].

Tyrimė [46] atskleidžiama idėja, kad ne visada reikia įvertinti klasifikavimo kintamųjų tarpusavio ryšio stiprumą, kadangi daugelis klasifikavimo metodų geba susidoroti su kintamųjų priklausomumo problema. Be to, Pirsono koreliacijos koeficiento naudojimas požymių atrankai yra kritikuojamas, nes atliekamas prieš modelio sudarymą ir neatsižvelgia į kintamųjų svarbą sudaromo modelio tikslumui.

Atlikus literatūros analizę, nustatyta, kad daugelis mokslininkų mokesčių vengimo identifikavimui arba panašių uždavinių sprendimui taikė sprendimų medžio, neuroninių tinklų ir

sustiprinimo metodus. Daugeliu atveju, atlikti tyrimai ir gauti modelių prognozių įverčiai su nedideliu duomenų rinkiniu ir iš anksto suformuotomis prielaidomis.

1.4. Klasifikavimo metodų ir programinės įrangos privalumai, trūkumai ir taikymo apribojimai

Šiame skyrelyje glaustai apžvelgsime 1.3 skyrelyje analizuotų klasifikavimo metodų privalumus ir trūkumus, siekiant pasirinkti tinkamiausius metodus baigiamojo projekto uždavinių sprendimui.

Atsitiktinių miškų metodas priskiriamas kolektyvų metodų klasei, gebantis susidoroti su dideliais duomenų kiekiais, nereikalauja specialaus nepriklausomų kintamųjų paruošimo prieš metodo taikymą, pavyzdžiui, standartizavimo. Sudarant modelį atsitiktinių miškų metodas identifikuoja reikšmingiausius kintamuosius prognozės sudarymui, trūkumas – modelis gali persimokyti analizuojant triukšmingus duomenis.

Artimiausių kaimynų metodas – priskiriamas tingių metodų klasei, modelio sudarymo metu metodas nenaudoja mokymo imties informacijos. Metodo prognozės tikslumas priklauso nuo k parametro pasirinkimo [47]. Metodas yra tinkamas, kai nėra pirminės informacijos apie duomenų pasiskirstymą, efektyvus analizuojant didelius duomenų kiekius, atsparus triukšmingiems duomenims, kai atstumas vertinamas skaičiuojant ne Euklido atstumą. Metodo realizavimui ir modelių sudarymui reikia daug kompiuterio atminties.

Logistinė regresija – vienas dažniausiai taikomų metodų. Siekiant sudaryti tikslesnį logistinės regresijos modelį galima papildomai naudoti reguliarizavimo parametrus, tai sumažina modelio permokymo galimybę, taip pat leidžia įvertinti kintamųjų svarbą, modelio rezultatus. Siekiant gauti stabilų sprendinį, reikalinga didelė mokomoji imtis. Metodas tinka kategoriniams atsako kintamiesiems prognozuoti, tačiau netinka tolydiems kintamiesiems. Vienas iš pagrindinių metodo reikalavimų, kad kintamieji būtų nepriklausomi, tačiau sprendžiant praktinius uždavinius, kintamieji dažniausiai yra priklausomi, todėl apsunkina metodo taikymą. Jei egzistuoja priklausomybė tarp kintamųjų, metodas pervertina kintamųjų svarbą [48].

Ekstremalaus gradientinio sustiprinimo metodas XgBoost priklauso sustiprinimo metodams, kai kiekvienos iteracijos metu yra sudaromas medis, kuris stengiasi pataisyti prieš tai iteracijoje buvusias klaidas. Metodas gali lengvai susidoroti su dideliais duomenų kiekiais ir retomis matricomis, tačiau praktiškai gali būti pritaikytas tik skaitinio tipo kintamiesiems. Sudarytas modelis gali būti permokytas, jei netaikoma papildomų apsaugų, pavyzdžiui, reguliarizavimo parametrų įtraukimo į modelį.

Kitas metodas - adaptyvus sustiprinimo AdaBoost yra tinkamas tiek sudėtingoms, tiek paprastoms problemoms spręsti. AdaBoost metodu galima sudaryti tikslų modelį iš silpnų klasifikatorių, tačiau metodas gali būti jautrus išskirtims ir triukšmingiems duomenims.

Neuroniniai tinklai nereikalauja specialaus kintamųjų paruošimo, gaunamas tikslus prognozės rezultatas, galima spręsti skirtingus uždavinius, pavyzdžiui, vaizdų atpažinimo. Metodą gana sudėtinga apmokyti ir parinkti geriausius parametrus, rezultatų negalima interpretuoti, kadangi sudaryto modelio rezultatas yra tik prognozė, reikalauja daug laiko norint analizuoti didelį duomenų rinkinį. Neuroninių tinklų privalumas – gebėjimas dirbti su triukšmingais duomenimis taip pat ir mokėjimas klasifikuoti modelius nors ir nebuvo apmokyti.

Atraminių vektorių metodas –klasifikavimo metodas. Galima išvengti modelio permokymo, parinkus tinkamą branduolį modelio sudarymui. Taikant modelį gaunami gana tikslūs rezultatai. Metodas gali aptikti tiek tiesinį, tiek netiesinį duomenų pasiskirstymą. Taikant metodą praktiškai susiduriama su problema, kad modelio kūrimo procesas reikalauja daug kompiuterio atminties, norint surasti geriausius modelio parametrus, modeliavimo procesas užtrunka, taip pat modelio kūrimo sėkmė priklauso nuo branduolio parinkimo, metodas gana lėtas.

Taikant Naive Bajeso metodą, formuluojama prielaida apie kintamųjų nepriklausomumą, sprendžiant realius uždavinius ši sąlyga dažniausiai negalioja.

Sprendimų medžiai – neparimetrinis metodas, kurio rezultatai lengvai interpretuojami, tačiau metodas yra linkęs persimokyti.

Svarbu analizuoti ne tik taikomus metodus, bet ir programines priemones, kurios leistų realizuoti modelius, užtikrintų skaičiavimų tikslumą. Duomenų tyrybos metodų srityje naudojamos įvairios programinės priemonės. 2014 metų apklausų duomenimis [49, 67], R, SAS, SQL ir Python sudarė populiariausių programavimo kalbų ketvertuką analitinėms, duomenų tyrybos ir duomenų mokslo uždaviniams spręsti.

R yra atviro programavimo kalba, skirta statistiniams skaičiavimams atlikti ir grafinei vizualizacijai pateikti. Programavimo kalbos galimybės yra praplečiamos papildomai įdiegiant paketus, kurie sukurti konkrečioms problemoms spręsti. R programavimo kalbos taikymui naudota RStudio programinė įranga, kurioje integruota R kalba. Programoje galima tiesiogiai vykdyti kodą, atvaizduoti grafikus, išsaugoti skaičiavimo istoriją, tikrinti ir suderinti programos kodą. RStudio yra atviro kodo programinė įranga, suderinama su skirtingomis operacinėmis sistemomis.

Forrester sudarytoje 2017 metų I ketvirčio ataskaitoje įvertintos programinės priemonės prognozuojamos analitikos ir mašininio mokymo (angl. *Predictive analytics and machine learning* –

PAML) srityje atsižvelgus į 23 kriterijus. Vizualus programinių priemonių palyginimas pateiktas (3 Priede 1 pav.) rodo, kad SAS išlieka lyderiu lyginant su kitomis programomis. Čia vertinami šiuo metu egzistuojantys modeliavimo, duomenų paruošimo, algoritmų sprendimai (angl. *current offering*) ir strategija (angl. *strategy*) – galimybė varžytis algoritmų srityje ir užimamą rinką. SAS sudaro vieningą sistemą duomenų paruošimo, vizualizavimo, duomenų analizės, modelių kūrimo ir modelio taikymo, tai sukuria nuoseklią vartotojo patirtį.

Python – atviro kodo plačiai naudojama programavimo kalba, galinti apdoroti didelius duomenų kiekius, turinti specialų paketą – scikit-learn, duomenų tyrybos uždaviniams spręsti. Python programavimo kalba nebuvo pritaikyta matematinių uždavinių sprendimui, todėl neturi daug statistinių paketų.

SQL – struktūrinė užklausų (angl. *Structured Query Language*), programavimo kalba skirta dirbti su reliacinėmis duomenų bazėmis ar apdoroti srautinius (angl. *stream*) duomenis. Ši programavimo kalba skirta duomenų saugojimui, manipuliavimui ir duomenų gavimas iš duomenų bazių. SQL programavimo kalba pasižymi greitu duomenų apdorojimu, tačiau nėra matematinių funkcijų, leidžiančių atlikti skaičiavimus.

Visi šiame skyriuje išanalizuoti metodai (k artimiausių kaimynų, Naive Bajeso, atsitiktinių miškų, sprendimų medžio, adaptyvaus ir ekstramalaus gradiento sustiprinimo, neuroninių tinklų) gali būti sprendžiami SAS, R ir Python priemonėmis.

Taigi, uždavinio sprendimui atlikti ir metodikai sudaryti pasirinkta R programavimo kalba R Studio aplinkoje ir SAS programinė įranga. Pasirinkimą taikyti šias priemones lėmė tai, jog pradiniai duomenų šaltiniai yra saugomi SAS, be to tai galingas įrankis duomenų paruošimo, turintis ir SQL procedūras. R programavimo kalba pasirinkta dėl plataus naujausių ir klasikinių klasifikavimo metodų pasirinkimo bei nesudėtingo naudojimosi.

Atsižvelgiant į klasifikavimo metodų taikymo privalumus ir trūkumus, nuspręsta darbe taikyti atsitiktinių miškų, ekstramalaus gradientinio sustiprinimo, adaptyvaus sustiprinimo, artimiausių kaimynų ir neuroninių tinklų metodus dėl metodų prognozės tikslumo, panašių atliktų studijų mokesčių vengimo identifikavimo srityje.

1.5. Apibendrinimas

Išanalizavus temos aktualumą, mokesčių vengimo problemos mastą ir apibendrinus mokslininkų analizuotus juridinių asmenų rizikingumo požymius, išskirti veiksniai, lemiantys mokesčių mokėtojų nenorą mokėti mokesčius. Atsižvelgiant į mokslininkų taikytų metodų įvairovę

bei pasiūlymus dėl duomenų rinkinio subalansavimo, baigiamajame darbe pasirinkta taikyti klasifikavimo metodus. Remiantis analizuotomis mokslinėmis publikacijomis, matematiškai bus vertinamas pasirinktų požymių įtaka klasifikavimo tikslumui, atliekamas priklausomo kintamojo klasių balansavimas, taikomi klasifikavimo metodai, lyginami modelių taikymo rezultatai, skaičiuojami klasifikavimo tikslumo įverčiai bei. sudaroma teoriškai pagrįsta metodika, padedanti įvertinti juridinių asmenų rizikingumą pagal pasirinktą kintamųjų sąrašą.