



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

# **Lietuvos įmonių bankroto prognostika**

Magistro studijų baigiamasis projektas

---

Projektą parengė

**Rasa Škiudaitė**

Projektui vadovavo

**Doc. dr. Šviesa Leitonienė**  
**Prof. dr. Evaldas Vaičiukynas**

---

**Kaunas, 2026**



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

## **Lietuvos įmonių bankroto prognostika**

Magistro studijų baigiamasis projektas  
Didžiųjų verslo duomenų analitika (6213AX001)

---

Projektą parengė

**Rasa Škiudaitė**

Projektui vadovavo

**Doc. dr. Šviesa Leitonienė**

**Prof. dr. Evaldas Vaičiukynas**

Projektą recenzavo

**Doc. dr. Kęstutis Lukšys**

**Doc. dr. Lina Sinevičienė**

---

**Kaunas, 2026**



**Kauno technologijos universitetas**

Matematikos ir gamtos mokslų fakultetas

Rasa Škiudaitė

## **Lietuvos įmonių bankroto prognostika**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdamas (-a) kitų asmenų autoriaus ar kitų teisių, laikydamasis (-i) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo, Kauno technologijos universiteto (toliau – Universitetas) intelektualinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. visi baigiamajame projekte pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena projekto dalis nėra plagijuota nuo spausdintinių ar elektroninių šaltinių, o visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. baigiamajame projekte tinkamai laikiausi asmens duomenų apsaugos reikalavimų, nenaudojau neskelbtinų ar konfidencialių duomenų be teisėto pagrindo, o jei juos naudoju, jie yra tinkamai nuasmeninti;
4. jei rengiant baigiamąjį projektą naudojausi dirbtinio intelekto (toliau – DI) ar kitais automatizuotais įrankiais, juos taikiau pagal Universitete nustatytą tvarką, nepažeisdamas (-a) akademinio sąžiningumo principų;
5. nesumokėjau ir nesu įsipareigojęs (-usi) mokėti jokių įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis jokiam fiziniam ar juridiniam asmeniui;
6. suprantu, kad išaiškėjus akademinio nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikoma atsakomybė pagal Universitete nustatytą tvarką ir galiu būti pašalintas (-a) iš Universiteto; akademinio nesąžiningumo atvejis gali būti nagrinėjamas ir po studijų baigimo, inicijuojant kvalifikacinio laipsnio atšaukimo procedūrą.

Škiudaitė Rasa. Lietuvos įmonių bankroto prognostika. Magistro studijų baigiamasis projektas / projektui vadovavo doc. dr. Šviesa Leitonienė ir prof. dr. Evaldas Vaičiukynas; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir studijų kryptių grupė: Taikomoji matematika (Matematikos mokslai).

Reikšminiai žodžiai: Bankrotas, klasifikacija, prognozė, klasių disbalansas, makroekonominiai rodikliai.

Kaunas, 2026. 62 p.

## Santrauka

Šiame magistro studijų baigiamajame projekte yra nagrinėjama viena iš verslo ir finansinių problemų – įmonių bankroto prognostika kintančioje ekonominėje aplinkoje. Verslų bankrotai sukelia tiek tiesioginių, tiek netiesioginių nuostolių investuotojams, kreditoriams, įmonių vadovams, bei valstybinėms institucijoms. Tradiciniai prognostikos modeliai, kurie remiasi statiniais finansiniais rodikliais, keičiantis ekonominėms sąlygoms, praranda savo tikslumą. Todėl kyla poreikis kurti pažangius mašininio mokymosi modelius, gebančius integruoti tiek vidinius įmonės finansinius, tiek makroekonominis rodiklius, atspindinčius ekonomikos ciklo etapus ir su jais susijusią sisteminę riziką.

Tyrimo objektas – Lietuvos įmonių bankroto rizikos prognozavimo galimybės. Pagrindinis šio darbo tikslas – atlikti šiuolaikinių įmonių bankroto prognozavimo tyrimų analizę ir jos pagrindu sudaryti Lietuvos įmonių duomenims tiksliausią bankroto prognozės modelį, apjungiantį finansinius ir makroekonominis rodiklius. Šiam tikslui pasiekti iškelti uždaviniai: išanalizuoti įmonių bankroto prognostikos teorinius aspektus, identifikuojant bankroto riziką lemiančius veiksnius, taikomų modelių ypatumus, bei duomenų struktūros ir makroekonominės aplinkos įtaką prognozių tikslumui; suformuoti galutinį Lietuvos įmonių duomenų rinkinį; atlikti detalų duomenų paruošimą, užtikrinant jų vientisumą ir suvaldant trūkstamas reikšmes; išanalizuoti logistinės regresijos, atsitiktinių miškų ir gradientinio stiprinimo algoritmų taikymo galimybes bei naudojant standartines metrikas įvertinti sudarytų modelių tikslumą; palyginus gautus rezultatus, išrinkti tiksliausią prognozės modelį bei pateikti rekomendacijas dėl jo taikymo praktikoje.

Darbo metodologijoje remiamasi tiek tradiciniais statistiniais metodais, tiek medžių ansamblių mašininio mokymosi metodais. Tyrimui naudoti realūs Lietuvos įmonių 2013-2022 m. duomenys, apimantys 1454 bankrutavusias ir 39011 nebankrutavusių įmonių. Siekiant objektyviai įvertinti modelių gebėjimą prognozuoti, duomenys buvo padalinti į du laikotarpius: 2013-2020 m. (1024 bankroto atvejai) skirtas apmokyti, o 2021-2022 m. (430 bankroto atvejų) – testuoti. Į modelius įtrauktas 81 nepriklausomas kintamasis, šį sąrašą sudaro pagrindiniai bei santykiniai finansiniai rodikliai (pelningumo, likvidumo, mokumo, efektyvumo) bei makroekonominiai rodikliai su vėlavimais bei pokyčiais (BVP augimas, EURIBOR vidurkis, VKI ir nedarbo lygis). Duomenų paruošimo etape trūkstamos reikšmės užpildytos MICE metodu. Klasifikavimo uždavinys spęstas R aplinkoje, darbe palyginama logistinė bei reguliarizuota logistinė regresija, atsitiktinių miškų ir GBM modeliai su trimis

skirtingomis parametru variacijomis. Modeliu kokybe vertinta naudojant tokias metrikas kaip AUC, jautrumas (atkuriamumas) ir F1 reiksme.

Tyrimo rezultatai parode, kad klasikinė logistinė regresija yra netinkama nagrinėjamam klasifikavimo uždaviniui (AUC = 0,5069), esant dideliu klasiu disbalansui. Regularizacijos taikymas pastebimai pagerino regresijos tiksluma (AUC = 0,7742). Logistinės regresijos modelių rezultatus pranoksta ansambliniai medžių metodai. Geriausiai rezultatus parode gilusis GBM modelis, pasiekęs aukštą tikslumą (AUC = 0,9104) ir aukščiausiai atkuriamumo rodiklio reiksme (0,9). Tai leido teisingai identifikuoti 387 bankroto atvejus iš 430, toks rezultatas leido sumažinti antro tipo klaidų skaičių, lyginant su kitu modelių rezultatais.

Kintamųjų svarbos analizė atskleidė, kad stipriausią įtaką modelio rezultatams turi naujai sukurtas kintamasis, parodantis finansinių ataskaitų teikimo vėlavimo apimtį dienomis („reporting\_delay“). Svarbią reiksme taip pat turi nuosavo kapitalo, turto, pelno ir įsipareigojimo rodikliai, bei jų pokyčiai laike. Makroekonominių rodiklių integravimas leido sėkmingai identifikuoti verslo ciklo kontekstą – tirtu ekonomikos atsigavimo laikotarpiu užfiksuotas nedarbo lygio mažėjimas ir dėl augančios infliacijos didėjančios įmonių sąnaudos.

Škiudaitė Rasa. Bankruptcy Prediction for Lithuanian Companies. Master's Final Degree / Project supervisors doc. dr. Šviesa Leitonienė and prof. dr. Evaldas Vaičiukynas; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and study field group: Applied Mathematics (Mathematical Sciences).

Keywords: Bankruptcy, classification, prediction, class imbalance, macroeconomic indicators.

Kaunas, 2026. 62.

## Summary

The aim of this master's thesis is to analyse corporate bankruptcy prediction as one of the key challenges in business and finance. Business failure can lead to direct and indirect losses for various stakeholders, including investors, creditors, corporate managers, and regulatory authorities. Traditional bankruptcy models, that mostly rely on financial ratios can often lose accuracy in a constantly changing economic environment. Therefore, it is relevant to develop dynamic and localised models capable of integrated both firm-level financial data and national macroeconomic factors, reflecting business cycle phases related to systemic risk.

The object of this research is potential ways of predicting bankruptcy risk of Lithuanian companies. The main goal of this work is to analyse research on corporate bankruptcy prediction and, based on this analysis, develop the most accurate bankruptcy prediction model. This model for Lithuanian company data has to combine both, financial and macroeconomic indicators. In order to achieve this goal, the following tasks were set: to analyse the theoretical aspects of corporate bankruptcy prediction, identifying the factors that determine bankruptcy risk, the characteristics of the models used, and the impact of data structure and the macroeconomic environment on the accuracy of predictions; to compile a final dataset of Lithuanian companies; perform detailed data preparation, ensuring data integrity and handling missing values; to analyse the applicability of logistic regression, random forests and gradient boosting algorithms, and evaluate the accuracy of the developed models using standard metrics; to select the most accurate prediction model, and provide practical recommendations, according to the given results.

The methodology of this work is based on both traditional statistical and advanced machine learning methods. Real data of Lithuanian companies from the period 2013-2022 was used for the study. The final dataset consists of 1454 bankrupt and 39011 non-bankrupt companies. In order to objectively evaluate the reliability of all the models, the data was split chronologically into two periods. 2013-2020 period (containing 1024 bankruptcy cases) for training the models and 2021-2022 period (containing 430 bankruptcy cases) was used for testing. The final data used for the models consists of 81 independent financial variables (covering profitability, liquidity, solvency, operational efficiency) and key macroeconomic indicators (GDP growth, EURIBOR, consumer price index and unemployment rate) with lags and changes. During the data preparation phase, due to a severe class imbalance in the data, MICE method was applied. Classification problem was solved by using and comparing several models. Logistic regression (with and

without elastic net regularisation), random forests and GBM. For the evaluation, the standard metrics were used: AUC, sensitivity (recall), F1-score.

The results of the study demonstrated that logistic regression (a traditional statistical method) is ineffective when dealing with severe class imbalance and it fails to accurately identify bankrupt companies. Meanwhile, machine learning algorithms significantly outperform traditional models. The deep GBM model demonstrated the best performance, achieving a high AUC score of 0.9104 and highest sensitivity (recall) of 0.9. This indicates that this model is capable of successfully identifying 90 % of companies that are about to bankrupt within a one-year horizon.

Analysis of variable importance revealed that financial reporting delay (`reporting_delay`) is a critical internal risk signal for a company. Furthermore, indicators related to equity, assets and liabilities also have a significant impact on bankruptcy prediction. The integration of macroeconomic factors provided models with an analytical advantage in assessing systemic risk. During the analysed period, the methods successfully identified specific risks related to business cycle factors: labor market overheating (unemployment rate changes) and rising raw material prices. The results of the study prove that bankruptcy prediction cannot be separated from the macroeconomic context. The combination of micro and macroeconomic variables ensures a higher prediction accuracy.

## Turinys

Lentelių sąrašas .....	10
Paveikslų sąrašas .....	11
Santrumpų ir terminų sąrašas .....	12
Įvadas.....	13
1. Įmonių bankroto prognostikos teorinis pagrindimas .....	14
1.1. Įmonių bankroto ekonominiai, teisiniai ir vadybiniai aspektai .....	14
1.2. Įmonių bankroto prognostikos modelių taikymo problematika .....	15
1.3. Makroekonominių veiksnių įtaka bankroto prognozavimo modeliams .....	16
1.4. Finansinių, makroekonominių ir ne finansinių rodiklių integravimas bankroto prognostikoje .....	18
1.5. Duomenų imties ir struktūros įtaka bankroto prognozių tikslumui.....	21
1.6. Įmonių bankroto prognostikos teorinių įžvalgų apibendrinimas .....	22
2. Tyrimo metodologija ir duomenų analizės metodai .....	24
2.1. Logistinė regresija .....	24
2.1.1. Logistinė regresija su bauda.....	24
2.1.2. Logistinė regresija su regularizacija .....	25
2.1.3. Praktinis logistinės regresijos taikymas .....	26
2.2. Atsitiktiniai miškai .....	26
2.2.1. Atsitiktinių miškų algoritmas.....	27
2.2.2. Atsitiktinių miškų klasifikavimo procesas .....	27
2.2.3. Atsitiktinių miškų metodo privalumai.....	27
2.2.4. Praktinis atsitiktinių miškų metodo taikymas.....	28
2.3. Gradientinio stiprinimo algoritmai .....	28
2.3.1. GBM matematinė išraiška ir sąsaja su R funkcijomis .....	28
2.4. Statistinės metrikos .....	29
2.4.1. ROC kreivė .....	29
2.4.2. Preciziškumo-atkuriamumo kreivė .....	30
2.4.3. Sumaišymo matrica .....	31
2.5. Kintamųjų svarba.....	33
2.6. Metodai ir duomenys .....	33
2.6.1. Tradiciniai vienos reikšmės užpildymo metodai.....	33
2.6.2. MICE trūkstamų reikšmių užpildymo metodas.....	34
2.6.3. Duomenų aprašymas .....	34
2.6.4. Kintamųjų sąrašas .....	34
2.6.5. Žvalgomoji analizė .....	36
3. Empirinis Lietuvos įmonių bankroto prognozavimo tyrimas.....	40
3.1. Duomenų analizė .....	40
3.2. Rezultatai .....	45
3.2.1. Logistinė regresija .....	45
3.2.2. Atsitiktiniai miškai .....	48
3.2.3. Gradientinio stiprinimo metodai .....	49
3.2.4. Modelių rezultatų palyginimas .....	51
3.2.5. Ekonominė rezultatų prasmė .....	53
3.3. Apibendrinimas ir rekomendacijos.....	55

Išvados .....	57
Literatūros sąrašas .....	59
Priedai .....	63
1 priedas .....	63
2 priedas .....	65
3 priedas .....	69

## Lentelių sąrašas

<b>1 lentelė.</b> Literatūroje randamų finansinių rodiklių sąrašas .....	18
<b>2 lentelė.</b> Literatūroje randamų makroekonominių rodiklių sąrašas .....	19
<b>3 lentelė.</b> Literatūroje randamų ne finansinių rodiklių sąrašas .....	20
<b>4 lentelė.</b> Literatūroje gaunami įmonių bankroto prognozavimo modelių tikslumo rezultatai .....	22
<b>5 lentelė.</b> Išvestinės sumaišymo matricos metrikos .....	33
<b>6 lentelė.</b> Bendrinė informacija apie įmonę pateikiama duomenyse .....	34
<b>7 lentelė.</b> Pagrindiniai įmonės finansiniai rodikliai pateikiami duomenyse .....	35
<b>8 lentelė.</b> Santykiniai įmonės finansiniai rodikliai pateikiami duomenyse .....	36
<b>9 lentelė.</b> Bankroto indikacijos stulpelių reikšmės pateikiamos duomenyse .....	36
<b>10 lentelė.</b> Įmonių pasiskirstymas pagal finansinių metų pradžios mėnesį .....	37
<b>11 lentelė.</b> Įmonių pasiskirstymas pagal finansinių metų pradžios mėnesio keitimosi dažnį .....	37
<b>12 lentelė.</b> Įmonių, turinčių vieną įrašą, pasiskirstymas pagal teisinį statusą .....	38
<b>13 lentelė.</b> Bankrutavusių įmonių pasiskirstymas pagal teisinį statusą .....	38
<b>14 lentelė.</b> Bankrutavusių įmonių dalis pagal įrašų skaičių .....	39
<b>15 lentelė.</b> Vėlavimo pateikti ataskaitas (dienomis) pasiskirstymas duomenyse .....	40
<b>16 lentelė.</b> Įmonių teisinio statuso reikšmės .....	40
<b>17 lentelė.</b> Vėlavimo pateikti ataskaitas (dienomis) pasiskirstymo duomenyse palyginimas .....	42
<b>18 lentelė.</b> Santykinų finansinių rodiklių reikšmių pasiskirstymo duomenyse palyginimas .....	43
<b>19 lentelė.</b> Papildomi finansinių rodiklių stulpeliai .....	44
<b>20 lentelė.</b> Bankrutavusių įmonių skaičius skirtinguose laiko intervaluose .....	44
<b>21 lentelė.</b> Papildomi makroekonominių rodiklių stulpeliai .....	45
<b>22 lentelė.</b> Logistinės regresijos modelių parametrai .....	46
<b>23 lentelė.</b> Atsitiktinių miškų modelių parametrai .....	48
<b>24 lentelė.</b> GBM modelių parametrai .....	50
<b>25 lentelė.</b> Geriausių klasifikavimo modelių rezultatai .....	52
<b>26 lentelė.</b> Makroekonominių rodiklių reikšmės modelio testavimo periode [42] .....	55

## Paveikslų sąrašas

<b>1 pav.</b>	ROC kreivių palyginimas [1] .....	30
<b>2 pav.</b>	PR kreivė [34].....	31
<b>3 pav.</b>	PR kreivių palyginimas [34] .....	31
<b>4 pav.</b>	Sumaišymo matrica [35].....	32
<b>5 pav.</b>	Bankrutavusių įmonių skaičius laike .....	44
<b>6 pav.</b>	Optimalios alfa reikšmės paieškos grafikas.....	46
<b>7 pav.</b>	Logistinės regresijos modelių ROC palyginimas .....	47
<b>8 pav.</b>	Logistinės regresijos modelių PR palyginimas .....	47
<b>9 pav.</b>	Atsitiktinių miškų modelių ROC palyginimas .....	48
<b>10 pav.</b>	Atsitiktinių miškų modelių PR palyginimas .....	49
<b>11 pav.</b>	GBM modelių ROC palyginimas.....	50
<b>12 pav.</b>	GBM modelių PR palyginimas.....	51
<b>13 pav.</b>	Geriausių bankroto prognozavimo modelių ROC palyginimas .....	51
<b>14 pav.</b>	Geriausių bankroto prognozavimo modelių PR palyginimas .....	52
<b>15 pav.</b>	Giliojo GBM modelio 30 svarbiausių indikatorių .....	54

## Santrumpų ir terminų sąrašas

AUC – plotas po kreive (angl. *area under the curve*) [1];

BVP – bendrasis vidaus produktas;

EURIBOR – Europos tarpbankinės rinkos palūkanų norma. (angl. *euro interbank offered rate*);

GBM – gradientinio stiprinimo metodai (angl. *gradient boosting method*) [2];

grynasis pelnas (nuostoliai) atspindi galutinį įmonės veiklos rezultatą - uždirbtą pelną arba nuostolius. Jis yra apskaičiuojamas iš pelno (nuostolių) prieš mokesčius atėmus pelno mokesčio sąnaudas [3];

ilgalaikis turtas – tai turtas, kurį įmonė planuoja laikyti ilgiau nei vienus metus, tokį turtą ji registruoja ilgalaikio turto sąskaitose [4];

įsipareigojimai – tai yra pinigų ar pinigų ekvivalentų suma, kurios dabartinė vertė yra gaunama diskontuojant, taikant arba palūkanų normą galiojusią įsipareigojimų atsiradimo metu, arba galiojančią balanso sudarymo dieną [5];

LASSO – mažiausio absoliutaus susitraukimo ir atrankos operatorius (angl. *least absolute shrinkage and selection operator*);

MICE – daugiamačių duomenų užpildymas susijusių lygčių metodu (angl. *multivariate imputation by chained equations*);

nuosavas kapitalas – tai įmonės turto dalis, gaunama iš viso turto atėmus įsipareigojimus [6]. Kai kuriais atvejais nuosavas kapitalas gali būti apibūdinamas paprastųjų akcijų rinkos verte [5];

pardavimo pajamos – tai per ataskaitinį laikotarpį gauta ekonominė nauda iš prekių pardavimo ir paslaugų teikimo, dėl kurios didėja įmonės turtas arba sumažėja įsipareigojimų, tuo pačiu padidėja nuosavas kapitalas, išskyrus papildomus savininkų įnašus [3];

pelnas (nuostoliai) iki mokesčių yra gaunamas iš bendrojo pelno (nuostolių) atėmus pardavimo, bendrąsias ir administracines sąnaudas, finansinio turto ir trumpalaikių investicijų vertės sumažėjimą, palūkanų, ir pridėjus kitos veiklos rezultatą, pajamas iš investicijų į patronuojančiąsias, patronuojamąsias bei asocijuotąsias įmones, taip pat kitas palūkanas ir panašias pajamas [3];

ROA – tai yra turto pelningumo rodiklis (angl. *return on assets*) [7];

ROE – tai yra nuosavo kapitalo grąža (angl. *return on equity*) [7];

trumpalaikis turtas – tai turtas, kurį tikimasi panaudoti arba parduoti per 12 mėnesių nuo paskutinės einamojo ataskaitinio laikotarpio dienos [4].

## Ivadas

Įmonių bankroto prognostika apima daug suinteresuotų grupių – jos apima investuotojus, kreditorius, įmonių vadovus, priežiūros institucijas. Įvairiose srityse nuolatos patiriamos išlaidos, susijusios su verslų bankrotais. Norėdamos tokių nuostolių išvengti, suinteresuotos grupės stengiasi tokioms situacijoms užbėgti už akių, bandydamos nuspėti ateitį, remdamosi, praeities duomenimis. Nepaisant plataus masto mokslinių tyrimų ir daugybės prognostikos modelių, pradedant ankstyvomis statistinėmis analizėmis ir baigiant sudėtingais mašininio mokymosi (angl. *machine learning*) ML ir dirbtinio intelekto (DI) metodais, iššūkis laiku pastebėti būsimo bankroto ženklus – išlieka. Tai rodo, kad bankroto reiškiniui būdingas sudėtingumas, glaudžiai susijęs su dinamiška ir nuolat kintančia verslo aplinka bei ekonominėmis sąlygomis [8]. Dėl šių kintančių sąlygų, nuolat peržiūrimas esamų modelių patikimumas ir pritaikomumas, todėl būtina nuolat atlikti mokslinius tyrimus ir modelius tobulinti.

**Darbo problema** – mokslinėje literatūroje ir praktikoje bankroto prognostikai yra taikomi įvairūs statistiniai bei mašininio mokymosi metodai, tačiau jų rezultatai dažnai priklauso nuo ekonominio konteksto, duomenų struktūros ir naudojamų rodiklių. Dėl šios priežasties, Lietuvos ekonominei aplinkai tinkamo bankroto prognozavimo modelio sudarymo, gebančio tiksliai įvertinti bankroto riziką, problema išlieka.

**Pagrindinis tikslas** – sudaryti ir empiriškai įvertinti Lietuvos įmonių bankroto prognozavimo modelį, paremtą finansinių ir makroekonominių rodiklių integracija bei pažangių prognostikos metodų taikymu.

Darbo uždaviniai:

1. išanalizuoti įmonių bankroto prognostikos teorinius aspektus, identifikuojant bankroto riziką lemiančius veiksnius, taikomų modelių ypatumus bei duomenų struktūros ir makroekonominės aplinkos įtaką prognozių tikslumui;
2. suformuoti galutinį duomenų rinkinį, apimantį sėkmingai veikiančių ir bankrutavusių įmonių finansinius bei makroekonominius rodiklius;
3. parengti ir apdoroti duomenis, taikant trūkstamų reikšmių užpildymo, kintamųjų atrankos ir modelių vertinimo metodus;
4. išanalizuoti logistinės regresijos, atsitiktinių miškų ir gradientinio stiprinimo algoritmų taikymo galimybes, įvertinti sudarytų modelių tikslumą ir patikimumą, naudojant standartines metrikas;
5. palyginus rezultatus, išrinkti tiksliausią prognostikos modelį, pateikiant rekomendacijas dėl jo taikymo praktikoje.

## 1. Įmonių bankroto prognostikos teorinis pagrindimas

Šiame skyriuje apibrėžta ekonominė ir vadybinė įmonių bankroto problema, kaip ji pateikiama ir suprantama pasirinktoje literatūroje. Čia apžvelgiami atrinktuose moksliniuose straipsniuose taikomi finansiniai, ne finansiniai, bei makroekonominiai rodikliai, skirti bankrotui prognozuoti. Taip pat apibrėžiama duomenų struktūros ir imties įtaka bankroto prognozavimo modelių rezultatams.

### 1.1. Įmonių bankroto ekonominiai, teisiniai ir vadybiniai aspektai

Literatūroje apibrėžiamas įmonių bankrotas yra suprantamas kaip teisinis procesas, kai įmonei yra iškeliami bankroto byla, bei kaip kompleksinė problema susijusi su ribotų išteklių paskirstymu, kai įmonė savo kapitalu negali padengti visų turimų įsipareigojimų. Pati sąvoka yra kilusi iš lotyniškų žodžių *bancus* ir *ruptus*, reiškiančių suolą arba stalą ir žodį „sulūžęs“. Istoriskai tai reiškia turgaus pardavėjo nesugebėjimą įvykdyti nustatytų įsipareigojimų [9].

Ekonominis bankroto aspektas yra išskaidomas pagal du pagrindinius sunkumų tipus su kuriais gali susidurti įmonė: ekonominius ir finansinius. Ekonominiai sunkumai atsiranda tada, kai įmonė savo generuojamomis pajamomis nesugeba padengti veiklos sąnaudų. Straipsnio autorių teigimu [9], tokias įmones likvidavus, socialinė gerovė yra padidinama. Finansiniai sunkumai atsiranda tada, kai įmonė veikia pelningai, tačiau ją slegia per didelė įsipareigojimų našta. Autorių nuomone [9], jei tokioms įmonėms būtų leidžiama veiklą tęsti, socialinė gerovė išaugtų.

Efektyvaus bankroto procedūra nusakoma kaip tokia, kurios metu yra stengiamasi maksimizuoti įmonės turto vertę. Turto vertės maksimizavimas gali būti atliekamas dalimis likviduojant turtą arba parduodant visą įmonę, taip siekiant užtikrinti kuo didesnės skolų dalies padengimą [10]. Problema gali iškilti, jei nėra pradedama oficiali bankroto procedūra, kadangi tada atskiri kreditoriai gali kelti atskiras bylas, siekdami areštuoti įmonės turtą. Tokie nesuderinti veiksmai gali lemti įmonės vertės sumažėjimą, o tai tiesiogiai sumažina sumą, kurią įmonė gyvavimo pabaigoje išmoka kreditoriams. Įprasto bankroto proceso metu, turtas turi būti paskirstomas pagal sutartyse nustatytą pirmenybės tvarką, vadovaujantis ekonominiu principu, vadinamu absoliučios pirmenybės principu [10].

Vadybinis bankroto aspektas yra susijęs su įmonės vadovybės elgesiu, bei atskaitomybe krizės metu. Veiksmingai veikiančioje sistemoje, bankrutuojančių įmonių vadovai turėtų patirti neigiamas pasekmes (pavyzdžiui – prarasti darbą), tokiu būdu įmonių vadovai jaustų didesnę poreikį pasverti priimamų rizikingų sprendimų pasekmes. Įmonei atsidūrus ant bankroto ribos, vadovai, veikiantys akcininkų vardu, gali būti suinteresuoti imtis itin rizikingų projektų. Tokiems projektams įvykus sėkmingai, įmonė lieka gyvuoti, tačiau priešingu atveju, didžiąją nuostolių dalį patiria kreditoriai, kadangi nuosavas kapitalas tuo momentu jau gali būti beveik bevertis. Taip pat praktikoje sutinkamas bankroto bylų iškėlimo atidėjimas: įmonių vadovai, siekdami išlaikyti kontrolę, atidėlioja bankroto bylą, nors ankstesnis bylos iškėlimas galėtų išsaugoti didesnę vertę, o tai leistų kokybiškiau atsiskaityti su kreditoriais [10]. Onokoya ir Olota [9] pabrėžia dvi vertės ir rizikos pasidalijimo teorijas. Vertės teorija teigia, kad įmonės ištekliai yra ne tik turtas, tačiau į juos patenka įmonės augimo ar nuostolių potencialas, kuris priklauso nuo vadovybės sprendimų, o jiems įtaką daro įvairūs socialiniai, politiniai ir moraliniai veiksniai. Rizikos pasidalijimo teorija teigia, kad akcininkai turėtų dalytis

kolektyvinę įmonės riziką, taip pat įskaitant riziką kylančią dėl vadovybės priimamų rizikingų sprendimų.

Apibendrinant galima teigti, kad bankrotas šaltiniuose yra pateikiamas kaip priemonė, padedanti išvengti pažeidimų, susijusių su teise į išmokų paskirstymą. Ekonominis poreikis efektyviai paskirstyti turtą turi būti suderintas su įmonės vadovybės drausme priimant rizikingus sprendimus.

## **1.2. Įmonių bankroto progностikos modelių taikymo problematika**

Didžioji dalis mokslinių tyrimų, nagrinėjančių įmonių bankroto prognozavimą, kaip vieną esminių iššūkių įvardija universalus modelio, tinkančio visoms situacijoms, paiešką. Literatūroje pabrėžiama, kad bankroto prognozių modeliai turi būti pritaikyti konkrečiam ekonominiam kontekstui, sektoriaus struktūrai ar regioniniams ypatumams, kadangi vienas metodologinis sprendimas retai gali užtikrinti vienodą prognozių tikslumą skirtingose aplinkose.

Šią tendenciją iliustruoja Lietuvos statybos sektoriaus tyrimai. Vienas iš jų [11] yra darbas, kuriame lyginamos skirtingos bankroto prognozavimo metodologijos, taikomos statybos sektoriaus įmonių duomenims. Autoriai parodė, kad skirtingi modeliai (tradiciniai statistiniai ir alternatyvūs metodai) pasižymi nevienodu tikslumu, o jų efektyvumas priklauso nuo sektoriaus specifikos. Tyrime pabrėžiama, kad statybos sektorius pasižymi išskirtiniu rizikos profiliu, susijusiu su projektų cikliškumu, ilgalaikėmis sutartimis ir dideliu finansinių srautų nepastovumu, todėl universalūs bankroto prognozavimo modeliai ne visada tinkamai atspindi realią rizikos struktūrą [11].

Dar labiau sektoriaus specifikos svarbą išryškina kitas tyrimas, nagrinėjantis Lietuvos statybų sektorių [12]. Šiame darbe analizuojamos mažos ir mikro statybos sektoriaus įmonės, o tyrimo imtis sudaryta iš 581 įmonės: 321 bankrutavusios įmonės, kurios bankrutavo arba pradėjo bankroto procesą 2009–2013 m., ir 260 nebankrutavusių įmonių, kurios iki 2021 m. darbą tęsė įprastai. Tyrimas išsiskiria tuo, kad bankrotui prognozuoti integruojami ne tik finansiniai, bet ir nefinansiniai, sektoriaus bei makroekonominiai rodikliai. Tokiu būdu yra sukuriamas kompleksinis modelis, leidžiantis plačiau įvertinti bankroto rizikos veiksnius. Šių dviejų tyrimų palyginimas leidžia daryti svarbią išvadą: net analizuojant tą patį sektorių, skirtingos metodologinės priemonės lemia skirtingus rezultatus ir interpretacijas. Jei [11] daugiausia dėmesio skiria palyginti metodus ir jų tinkamumą sektoriaus kontekste, tai mažų ir mikro įmonių tyrimas išplečia analizę, įtraukdamas makroekonominis ir struktūrinius veiksnius. Tokiu būdu Lietuvos statybos sektoriaus tyrimai ne tik patvirtina modelių lokalizacijos svarbą, bet ir parodo, kad bankroto rizika yra kompleksinis reiškinys, kurį lemia ne vien finansiniai rodikliai, bet ir platesnis ekonominis bei sektoriaus kontekstas.

Panaši tendencija pastebima ir kitose šalyse. Apžvalgoje [13], skirtoje Vengrijos bankroto prognozavimo modelių raidai, išsamiai aptariama daugiau nei trisdešimties metų vietinių modelių kūrimo istorija. Autoriai pažymi, kad tarptautiniai modeliai pasirodė sunkiai pritaikomi Vengrijos ekonominėms sąlygoms. Buvo pradėtos dėti intensyvios nacionalinio lygio pastangos kurti empirinius modelius, kurie geriau atitiktų vietos ekonominę struktūrą, teisinę aplinką ir duomenų specifiką [13]. Šią išvadą papildė kitas tyrimas, kuriame

naudojami itin didelės apimties Vengrijos įmonių duomenys. Autoriai parodė, kad net ir taikant pažangius ekonometrinius modelius, jų prognozių tikslumas priklauso nuo vietinių duomenų struktūros ir ekonomikos ciklo ypatumų. Makroekonominiai kintamieji ne visada reikšmingai pagerina modelio tikslumą, tačiau yra svarbūs siekiant paaiškinti bendrą bankroto rizikos dinamiką [14].

Panašias įžvalgas pateikia, straipsnio autoriai, kurie analizuoja įmonių bankroto prognozavimą pasitelkdami mašininio mokymosi metodus [15]. Autoriai nustatė, kad pažangūs algoritmai dažnai pranoksta tradicinius statistinius modelius pagal prognozių tikslumą, tačiau jų efektyvumas taip pat priklauso nuo duomenų struktūros ir laikotarpio. Tai leidžia teigti, kad net pažangiausi metodai nėra universalūs ir jų pritaikomumas turi būti vertinamas atsižvelgiant į kontekstą.

Dar platesnį požiūrį į kontekstinių veiksnių reikšmę pateikia autoriai, analizavę didelių Europos įmonių bankroto rizikos raidą skirtingose šalyse [8]. Straipsnyje parodoma, kad net taikant tuos pačius bankroto prognozavimo modelius, rizikos vertinimo rezultatai skiriasi tarp šalių ir laikotarpių, o tai leidžia daryti išvadą, jog makroekonominė aplinka ir instituciniai skirtumai daro reikšmingą įtaką bankroto rizikos struktūrai.

Toks nuoseklus siekis kurti vietinei rinkai pritaikytus modelius leidžia daryti išvadą, kad ekonominis kontekstas yra vienas svarbiausių, nors dažnai ir netiesiogiai vertinamų, aiškinamųjų veiksnių bankroto prognozavimo tyrimuose. Modelio pritaikymas, pasirenkant konkrečios šalies duomenis ir sektoriui būdingus finansinius rodiklius, gali būti interpretuojamas kaip bandymas įtraukti šio konteksto poveikį į modelio struktūrą. Santykinė tokių lokalizuotų modelių sėkmė patvirtina, kad kontekstiniai veiksniai turi reikšmingą aiškinamąją galią, o tai leidžia teigti, jog universalus bankroto prognozavimo metodas dažnai nėra optimalus sprendimas.

### **1.3. Makroekonominių veiksnių įtaka bankroto prognozavimo modeliams**

Vienas iš esminių bankroto prognozavimo modelių ribotumų yra jų statiškumas. Daugelis tradicinių modelių kuriami remiantis konkretaus laikotarpio istoriniais duomenimis ir daro prielaidą, kad ekonominė aplinka išlieka santykinai stabili. Tačiau realybėje makroekonominės sąlygos nuolat kinta – ekonomikos ciklai, finansų krizės ir struktūriniai pokyčiai daro reikšmingą poveikį įmonių veiklai. Dėl šios priežasties modeliai, kurie nėra pritaikyti prie kintančios ekonominės aplinkos, gali prarasti prognozavimo tikslumą ir tapti mažai patikimi [11].

Šią problemą patvirtina ir empirinių tyrimų rezultatai, kurie rodo, kad bankroto prognozavimo modelių efektyvumas priklauso ne tik nuo naudojamų kintamųjų, bet ir nuo ekonomikos ciklo etapo. Vengrijos kontekste straipsnio [13] autoriai apžvelgia ilgalaikę bankroto prognozavimo modelių raidą ir pabrėžia, kad modelių struktūra kito kartu su ekonomine aplinka. Autoriai pažymi, kad modeliai buvo adaptuojami atsižvelgiant į skirtingus ekonomikos ciklo etapus, o įmonių rangavimas pagal riziką tapo svarbia priemone, leidžiančia įvertinti dinamiškai kintančius kapitalo ir rizikos reikalavimus. Tai leidžia daryti išvadą, kad bankroto prognozavimas negali būti atsietas nuo makroekonominio konteksto – svarbu vertinti ne tik regioninius skirtumus, bet ir ekonomikos ciklo etapą.

Makroekonominių veiksnių svarbą patvirtina ir [14] tyrimas, kuriame analizuojami Vengrijos įmonių duomenys. Autoriai nustatė, kad makroekonominiai kintamieji, tokie kaip BVP augimas, kreditų plėtra ar palūkanų normos, ne visada reikšmingai pagerina modelio tikslumą trumpuoju laikotarpiu, tačiau jie yra būtini siekiant paaiškinti bendrą bankroto rizikos dinamiką ir jos svyravimus ekonomikos ciklo metu. Šis rezultatas rodo, kad makroekonominiai veiksniai dažniau veikia ne kaip tiesioginiai bankroto prognozavimo indikatoriai, bet kaip sisteminiai rizikos veiksniai, lemiantys bendrą bankrotų lygį.

Lietuvos statybos sektoriaus tyrimai taip pat patvirtina makroekonominės aplinkos svarbą. Giriūnienė [11] nustatė, kad įmonių veiklą ir sėkmingą plėtrą reikšmingai lemia tokie išoriniai veiksniai kaip nedarbo lygis, statybos prekių kainų indeksas ir būsto kainų pokyčiai. Tai reiškia, kad modeliai, kurie remiasi vien tik įmonės finansiniais rodikliais, gali neįvertinti esminių sisteminių rizikos veiksnių, galinčių lemti bankrotą nepriklausomai nuo individualios įmonės finansinės būklės. Tokiu būdu išryškėja problema – sudėtinga įmonės lygmens charakteristikų ir makroekonominių veiksnių sąveika, kuri dažnai lemia bankroto rizikos didėjimą.

Šią sąveiką dar labiau išplečia tyrimai, taikantys pažangius modelius ir prognozavimo metodus, taikomus keliems periodams. Straipsnis [15] parodė, kad mašininio mokymosi modelių prognozavimo tikslumas priklauso nuo laikotarpio ir ekonominės aplinkos stabilumo. Autoriai nustatė, kad modelių tikslumas mažėja ilgėjant prognozavimo horizontui ir didėjant ekonominiam neapibrėžtumui, o latentinių rizikos veiksnių įtraukimas leidžia geriau atspindėti sisteminius ekonominius svyravimus. Tai leidžia daryti išvadą, kad net pažangiausi modeliai nėra visiškai atsparūs ekonominių sąlygų pokyčiams, todėl jų rezultatai turi būti interpretuojami remiantis kontekstu.

Panašias įžvalgas pateikia ir autoriai [8], analizavę didelių Europos įmonių bankroto rizikos raidą, skirtingais laikotarpiais, įvairiose šalyse. Autoriai parodė, kad net taikant tuos pačius bankroto prognozavimo modelius, rizikos vertinimo rezultatai skiriasi, priklausomai nuo makroekonominių sąlygų ir institucinės aplinkos. Tai leidžia teigti, kad bankroto rizika nėra vien tik įmonės vidinių veiksnių rezultatas, bet ir platesnės ekonominės struktūros išraiška. Ekonominių sąlygų poveikį modeliams, patvirtina į atskirus sektorius orientuoti tyrimai. Nehrebecka [16] tyrime, lyginančiame *logit* ir atraminių vektorių klasifikatorius (angl. *support vector machines*) SVM, nustatyta, kad skirtingų metodų prognozavimo tikslumas kinta priklausomai nuo ekonominio laikotarpio ir duomenų struktūros. Tai rodo, kad net metodologiškai pažangūs modeliai nėra universalūs ir jų efektyvumas priklauso nuo ekonominės aplinkos stabilumo.

Dar vieną svarbią įžvalgą pateikia [17], analizuodami Altmano Z-score modelio taikymą gamybos sektoriaus įmonėms. Autoriai pateikia pavyzdį, kai 2020–2021 m. staigus įmonės Z įverčio sumažėjimas greičiausiai atspindėjo išorinius ekonominius sukrėtimus – COVID-19 pandemiją – o vėliau, stabilizuojantis ekonominei situacijai, įmonės finansiniai rodikliai pagerėjo. Šis pavyzdys parodo, kad bankroto rizika gali būti laikinai padidėjusi dėl makroekonominių šokų, todėl modeliai, kurie neatsižvelgia į laiko dimensiją ir ekonominį kontekstą, gali pateikti klaidinančius rezultatus.

Apibendrinant galima teigti, kad norint veiksmingai prognozuoti bankrotą, būtinas integracinis požiūris, kuriame derinami endogeniniai (įmonės lygmens) ir egzogeniniai (makroekonominiai) veiksniai. Tai daro įtaką duomenų rinkimo etapui, kadangi kuriant modelius, būtina naudoti ne tik finansinius, bet ir makroekonominis duomenis. Be to, literatūra rodo, kad sisteminių krizių laikotarpiais net ir finansiškai stabilios įmonės gali tapti rizikingomis, todėl bankroto prognozavimo modeliai turi būti dinamiški ir gebėti prisitaikyti prie besikeičiančių ekonominių sąlygų. Tokiu būdu bankroto prognozavimas tampa ne vien statistiniu uždaviniu, bet kompleksiniu procesu, reikalaujančiu nuolatinio modelių prisitaikymo prie ekonominės aplinkos pokyčių.

#### 1.4. Finansinių, makroekonominių ir ne finansinių rodiklių integravimas bankroto prognostikoje

Bankroto prognozavimo modeliuose pagrindinę vietą užima finansiniai rodikliai, apskaičiuojami remiantis įmonių finansinėmis ataskaitomis. Šie rodikliai leidžia kiekybiškai įvertinti įmonės finansinę būklę, veiklos efektyvumą, kapitalo struktūrą ir likvidumą, todėl jie laikomi pagrindiniu informacijos šaltiniu, modeliuojant bankroto riziką [18]. Literatūroje dažniausiai naudojami rodikliai apima pelningumo, likvidumo, mokumo ir veiklos efektyvumo aspektus. Pagrindiniai autorių aprašomi finansiniai rodikliai yra nurodyti žemiau esančioje lentelėje.

1 lentelė. Literatūroje randamų finansinių rodiklių sąrašas

Rodiklis	Kategorija	Autoriai
ROA	Pelningumas	Giriūnienė (2019), Kanapickienė (2023), Onofrei (2014), Bauer (2016), Brédart (2022), Sigrist (2023), Nehrebecka (2018), Devi (2025)
Turto apyvartumo rodiklis	Efektyvumas	Giriūnienė (2019), Kanapickienė (2023), Onofrei (2014), Sigrist (2023), Bärbuță-Mișu (2020), Devi (2025)
Bendrasis skolos rodiklis	Mokumas	Giriūnienė (2019), Kanapickienė (2023), Bauer (2016), Brédart (2022), Sigrist (2023)
ROE	Pelningumas	Onofrei (2014), Brédart (2022), Sigrist (2023)
Skolos ir nuosavo kapitalo santykis	Mokumas	Onofrei (2014), Bauer (2016), Sigrist (2023), Bärbuță-Mișu (2020)
Trumpalaikio turto ir įsipareigojimų santykis	-	Giriūnienė (2019), Kanapickienė (2023), Bärbuță-Mișu (2020)
Apyvartinio kapitalo ir turto santykis	Likvidumas	Kanapickienė (2023), Sigrist (2023), Bärbuță-Mișu (2020), Devi (2025)
Grynoji pelno marža	Pelningumas	Giriūnienė (2019), Kanapickienė (2023), Onofrei (2014), Sigrist (2023)
Visas turtas	Pelningumas	Kanapickienė (2023), Sigrist (2023), Bärbuță-Mișu (2020), Devi (2025)
Investicijų grąža	Pelningumas	Nehrebecka (2018), Onofrei (2014)

Finansiniai rodikliai dažnai naudojami ne atskirai, bet yra integruojami į kompleksinius bankroto prognozavimo indeksus, tokius kaip *Altmano Z*, *Ohlsono O*, *Springate'o S*, *Zmijewskio Z* ar *Groverio G* [17]. Šių modelių populiarumas literatūroje rodo, kad finansiniai rodikliai išlieka pagrindiniu bankroto prognozavimo instrumentu, tačiau jų prognostinė galia priklauso nuo ekonominio ir sektoriaus konteksto.

Svarbu pažymėti, kad finansinių rodiklių naudojimas neapsiriboja statinėmis reikšmėmis vienu laiko momentu. Literatūroje pabrėžiama, jog didelę prognozavimo vertę turi dinaminiai rodikliai, atspindintys finansinių rodiklių pokyčius laikui bėgant. Tokie rodikliai leidžia identifikuoti ankstyvuosius finansinių sunkumų signalus ir geriau prognozuoti bankroto riziką nei statiniai rodikliai [18]. Tai ypač aktualu ekonominio ciklo svyravimų kontekste, kai įmonių finansinė būklė gali greitai keistis.

Literatūroje galima rasti nemažai ekonomikos ciklą nusakančių indikatorių, kurie įtraukia tiek standartinius ekonominius rodiklius: tokius kaip BVP, nedarbo lygis, vartotojų kainų indeksas [19], tiek platesnį finansinių kintamųjų spektrą, apimantį akcijų rinkos rodiklius, bei nevienodo laikotarpio vyriausybės obligacijų palūkanų normų skirtumus [20]. BVP yra pagrindinis rodiklis, siekiant identifikuoti ekonomikos ciklus, bei augimo dinamiką [19]. Nedarbo lygis taip pat yra laikomas vienu pagrindiniu vėluojančiu (angl. *lagging*) rodikliu, kuris literatūroje yra dažnai siejamas su recesijomis [20]. Vartotojų kainų indekso rodiklio reikšmė yra naudojama stebėti infliacijai, kuri dažnai yra siejama su cikliniais pokyčiais, bei reakcija į taikomas pinigų politikos priemones [19]. Tokių rodiklių, kaip BVP ir VKI augimas reiškia ekonomikos atsigavimą, o nedarbo lygio didėjimas indikuoja recesijos periodą [21].

Tyrimai rodo, kad tokie rodikliai kaip BVP augimas, nedarbo lygis, palūkanų normos, infliacija, kreditų plėtra ar kainų indeksai gali reikšmingai paveikti įmonių bankroto riziką. Pavyzdžiui, Lietuvos statybos sektoriaus tyrimuose nustatyta, kad įmonių bankroto tikimybė yra susijusi su makroekonominiais veiksniais, tokiais kaip nedarbo lygis, statybos kainų indeksas ir būsto kainų pokyčiai [11]. Panašias išvadas pateikia ir [14], kurie parodė, kad makroekonominiai kintamieji yra svarbūs aiškinant bendrą bankrotų dinamiką, nors jų poveikis individualių įmonių bankroto prognozėms gali būti netiesioginis. Literatūroje sutinkamų makroekonominių rodiklių sąrašas pateikiamas žemiau esančioje lentelėje. Čia minimas kreditavimo augimo rodiklis, remiantis Ispanijos bankų sistemos pelningumo analize [22], gali būti siejamas su EURIBOR reikšme. Straipsnio autoriai teigia, kad esant neigiamai arba labai žemai tarpbankinės rinkos palūkanų normos reikšmei, bankai gali pradėti mažinti skolinimo standartus, tokiu būdu prisiimdami papildomą riziką. Siekis išlaikyti pelningumą gali padidinti bendrai suteikiamų paskolų kiekį.

**2 lentelė.** Literatūroje randamų makroekonominių rodiklių sąrašas

<b>Makroekonominis rodiklis</b>	<b>Autoriai</b>
BVP augimas / indeksas	Giriūnienė (2019), Kanapickienė (2023), Bauer (2016), Sigrist (2023)
Būsto kainų indeksas (HPI)	Giriūnienė (2019), Kanapickienė (2023), Sigrist (2023)

Infliacija, vartotojų kainų indeksas	Giriūnienė (2019), Kanapickienė (2023), Sigrisť (2023)
Palūkanų normos	Bauer (2016), Sigrisť (2023)
Kreditavimo augimas rinkoje	Bauer (2016), Sigrisť (2023)
Sektoriaus specifiniai rodikliai	Giriūnienė (2019), Kanapickienė (2023)
Nedarbo lygis	Kanapickienė (2023)

Prusako [18] apžvalgoje pateikiamas pavyzdys, kad privačių įmonių bankrotui prognozuoti buvo taikyta daugiamatė diskriminantinė analizė, remiantis finansiniais rodikliais. Tačiau tyrimuose pabrėžiama, kad vien finansinių rodiklių nepakanka siekiant tiksliai prognozuoti bankrotą. Dėl šios priežasties, vis dažniau į modelius įtraukiami nefinansiniai rodikliai, tokie kaip gautinų sumų elgsenos kintamieji, kokybinės įmonių charakteristikos, ar su kreditavimu susiję kintamieji, kurie leidžia geriau įvertinti įmonės veiklos riziką ir padidina modelių prognozavimo galią [18].

### 3 lentelė. Literatūroje randamų ne finansinių rodiklių sąrašas

Rodiklis	Autoriai
Įmonės amžius	Giriūnienė (2019), Kanapickienė (2023), Bauer (2016), Sigrisť (2023)
Vėlavimas pateikti ataskaitas	Giriūnienė (2019), Kanapickienė (2023)
Įmonės dydis	Onofrei (2014), Bauer (2016)
Eksporto požymis	Bauer (2016), Sigrisť (2023)
Užsienio kapitalas	Bauer (2016)
Sektoriaus požymis	Sigrisť (2023), Nehrebecka (2018)
Ryšiai su bankais	Sigrisť (2023)

Rodiklių parinkimo problematika yra glaudžiai susijusi su taikomais metodais. Tradiciniai statistiniai metodai, tokie kaip daugiamatė diskriminantinė analizė ir logistinė regresija, dažniausiai remiasi ribotu, teoriškai pagrįstu rodiklių rinkiniu [8]. Tuo tarpu pažangūs mašininio mokymosi metodai leidžia naudoti platesnį rodiklių spektrą ir identifikuoti sudėtingus netiesinius ryšius tarp kintamųjų [13]. Literatūroje minimi neuroniniai tinklai, atraminių vektorių metodas, sprendimų medžiai ir įvairūs ansambliniai metodai, tokie kaip „Boosting“, „Bagging“, „Stacking“, „AdaBoost“ ir „Gradient Boosting Machines“, kurie pasižymi aukštesniu prognozavimo tikslumu, tačiau dažnai praranda lengvo interpretavimo galimybę.

Siekiant užtikrinti modelių patikimumą, tyrimuose taip pat taikomi kintamųjų atrankos ir duomenų redukcijos metodai. Pagrindinių komponentių analizė ir kiti metodai leidžia sumažinti kintamųjų skaičių ir atrinkti statistiškai reikšmingiausius rodiklius, taip išvengiant

perteklinio modelių sudėtingumo ir multikolinearumo problemų. Tai ypač svarbu praktiniuose tyrimuose, kuriuose būtina suderinti modelio tikslumą ir jo interpretavimo galimybes.

Apibendrinant galima teigti, kad bankroto prognozavimo modeliuose naudojamų rodiklių pasirinkimas nėra atsitiktinis procesas, bet kompleksinis metodologinis sprendimas, priklausantis nuo ekonominio konteksto, sektoriaus specifikos ir taikomo metodo. Todėl tiek endogeninių, tiek egzogeninių rodiklių integravimas sudaro teorinį pagrindą praktinei tyrimo daliai, kurioje siekiama įvertinti skirtingų rodiklių reikšmę, prognozuojant įmonių bankroto tikimybę.

### **1.5. Duomenų imties ir struktūros įtaka bankroto prognozių tikslumui**

Vienas svarbiausių bankroto prognozavimo tyrimų aspektų yra naudojamų duomenų struktūra, kuri apima imties dydį, duomenų dažnumą, laikotarpį ir įmonių atrankos kriterijus. Literatūroje pastebima, kad tyrimų rezultatai ir modelių tikslumas dažnai priklauso ne tik nuo taikomų metodų ar pasirinktų rodiklių, bet ir nuo duomenų pobūdžio bei jų apimties. Todėl duomenų struktūra laikoma vienu esminių veiksnių, lemiančių bankroto prognozavimo modelių patikimumą.

Empiriniai tyrimai rodo, kad bankroto prognozavimo modeliai kuriami, remiantis labai skirtingo masto duomenų bazėmis – nuo nedidelių, detalių imčių iki didelio masto administracinių ar rinkos duomenų rinkinių. Vienas iš didelės apimties duomenų pavyzdžių yra [14] tyrimas, kuriame naudojama daugiau nei 1,5 mln. duomenų įrašų – įmonių metinių stebėjimų duomenys šešiolikos metų laikotarpiu. Tokia imtis leidžia analizuoti ne tik individualių įmonių bankroto riziką, bet ir jos dinamiką ilguoju laikotarpiu, identifikuoti struktūrinius pokyčius ir ekonomikos ciklo poveikį. Tačiau didelės apimties duomenys taip pat kelia metodologinių iššūkių, susijusių su duomenų heterogeniškumu ir modelių stabilumu.

Kita vertus, dalis tyrimų remiasi palyginti mažomis imtimis, kurios leidžia giliau analizuoti konkrečių įmonių finansinius rodiklius, tačiau riboja rezultatų apibendrinimą platesnei rinkai. Pavyzdžiui, [23] tyrime naudojama 100 įmonių imtis, papildoma 40 įmonių validacijos, bei apie 100 įmonių testavimo imtis, apimanti 2007–2011 m. laikotarpį. Nors tokia duomenų struktūra leidžia detaliai analizuoti finansinius rodiklius ir gauti aukštus prognozavimo tikslumo rodiklius, jų patikimumas turi būti vertinamas kritiškai, nes mažos imtys gali lemti modelių persimokymą ir ribotą rezultatų pritaikomumą platesniam ekonominiam kontekstui.

Tarpinio dydžio imtys būdingos didelei daliai tyrimų, orientuotų į atskirų sektorių analizę. Pavyzdžiui, [24] analizėje naudojama 2 266 įmonių imtis, iš kurių tik 153 yra bankrutavusios. Šis tyrimas išryškina vieną svarbiausių bankroto prognozavimo problemų – nesubalansuotą duomenų struktūrą. Autoriai parodo, kad tradiciniai statistiniai modeliai dažnai pasižymi ribotu jautrumu bankrutavusių įmonių klasei, todėl būtina taikyti duomenų balansavimo metodus ir pažangesnius algoritmus. Tokie rezultatai patvirtina, kad ne tik imties dydis, bet ir jos struktūra turi esminę įtaką modelių prognozavimo rezultatams.

Dar viena svarbi duomenų struktūros dimensija yra duomenų dažnumas. Dauguma tradicinių tyrimų remiasi metiniais finansiniais duomenimis, tačiau šiuolaikiniai tyrimai gali naudoti didesnio dažnumo duomenis. [15] tyrime naudojami mėnesiniai listinguojamų

įmonių duomenys, apimantys ilgą laikotarpį nuo 1961 iki 2022 m. Tokia duomenų struktūra leidžia modeliuoti daugiaperiodines bankroto tikimybes ir tiksliau įvertinti rizikos dinamiką laike. Tačiau, didelio dažnumo duomenys reikalauja sudėtingesnių modelių, didesnių skaičiavimo resursų ir kelia papildomų iššūkių dėl duomenų triukšmo bei nestabilumo.

Žemiau pateiktoje lentelėje galima matyti autorių naudotų duomenų apimtį, gautų modelių tikslumą, bei kurios šalies duomenimis remiantis buvo vykdomas tyrimas.

**4 lentelė.** Literatūroje gaunami įmonių bankroto prognozavimo modelių tikslumo rezultatai

<b>Straipsnis</b>	<b>Šalis</b>	<b>Duomenų kiekis</b>	<b>Gautas tikslumas</b>
X. Brédart and D. A. Correa-Mejía (2022) [24]	Belgija	7814 įmonių duomenys, tačiau 5548 atmetos dėl finansinės informacijos trūkumo, 153 įmonės iš likusios imties – bankrutavusios, 2017 m. duomenys	Bendras tikslumas: <i>boosting</i> algoritmas: 86,7 %, logistinė regresija: 93,2 %
Onofrei, Lupu (2014) [23]	Rumunija	100 įmonių: 50 bankrutavusių, 50 – veikiančių, papildomai 40 ir apie 100 įmonių validacijos bei testavimo imtims, duomenys apima 2007-2011 m. laikotarpį	Bendras tikslumas: daugiamatė diskriminantinė analizė – 95%, logistinė regresija – 82 %
Bauer & Endrész (2016) [14]	Vengrija	1585663 duomenų eilutės nuo 1996 iki 2012m. įmonės lygmenyje, įmonių skaičius neminimas	Panelinis probit modelis: AUC = 0,859
Sigrist & Leuenberger (2023) [15]	JAV	20235 įmonių duomenys kiekvienam mėnesiui, iš viso 2824412 eilučių, 1402 įmonės iš sąrašo bankrutavusios	1 metų horizontas: GBM: AUC = 0,933; atsitiktiniai miškai: AUC = 0,931
Kanapickienė, Kanapickas, Nečiūnas (2023) [12]	Lietuva	2007-2013 m. periodo duomenys, 321 tuo periodu bankrutavusi įmonė ir 260 įmonių, kurios gyvavo bent iki 2021 metų	Bendras tikslumas: daugiamatė diskriminantinė analizė: 68 % – 76 %, logistinė regresija: 71 % – 77 %, ML modeliai siekia 85 % tikslumą.

Apibendrinant galima teigti, kad bankroto prognozavimo tyrimuose egzistuoja kompromisas tarp duomenų apimtys ir jų detalumo. Didelės apimtys duomenys užtikrina statistinį patikimumą ir leidžia analizuoti sisteminius reiškinius, tačiau mažesnės imtys suteikia galimybę interpretuoti individualių įmonių finansinius procesus. Be to, duomenų dažnumas ir struktūra lemia ne tik modelių tikslumą, bet ir jų interpretavimo galimybes. Todėl duomenų pasirinkimas tampa ne techniniu, o metodologiniu sprendimu, kuris turi būti pagrįstas tyrimo tikslais ir ekonominiu kontekstu.

### 1.6. Įmonių bankroto prognostikos teorinių įžvalgų apibendrinimas

Atlikta mokslinės literatūros analizė parodė, kad įmonių bankroto prognozavimas yra kompleksinė ir dinamiška tyrimų sritis, kurioje nėra universalaus modelio, tinkamo visoms ekonominėms aplinkoms, sektoriams ar laikotarpiais. Dauguma autorių sutaria, kad

bankroto prognozavimo modelių efektyvumas priklauso ne tik nuo pasirinktų metodų, bet ir nuo ekonominio konteksto, duomenų struktūros bei naudojamų rodiklių rinkinio. Tai leidžia daryti išvadą, kad bankroto prognozavimas turi būti vertinamas ne kaip standartizuotas metodologinis sprendimas, bet kaip procesas, atsižvelgiantis į kontekstą, reikalaujantis modelių adaptacijos.

Literatūra atskleidė, kad tradiciniai statistiniai modeliai, tokie kaip daugiamatė diskriminantinė analizė ir logistinė regresija, išlieka svarbūs dėl savo lengvos interpretacijos ir paprastumo. Tačiau jų prognozavimo galimybės dažnai yra ribotos, ypač struktūrinių ekonomikos pokyčių, ar krizinių situacijų metu. Tuo tarpu, pažangūs mašininio mokymosi metodai, leidžia identifikuoti sudėtingesnius ryšius tarp kintamųjų ir dažnai pasižymi tikslesnėmis prognozėmis, tačiau jų taikymas reikalauja didesnių duomenų apimčių ir sudėtingesnių metodologinių sprendimų. Tai patvirtina, kad modelių pasirinkimas turi būti grindžiamas ne tik remiantis tikslumo kriterijais, bet ir atsižvelgiant į duomenų prieinamumą bei tyrimo tikslą.

Svarbi literatūros įžvalga – bankroto rizika negali būti paaiškinama vien tik įmonės vidiniais finansiniais rodikliais. Nors finansiniai rodikliai sudaro pagrindinį bankroto prognozavimo modelių pagrindą, daugelis tyrimų pabrėžia nefinansinių ir makroekonominių veiksnių svarbą. Endogeniniai rodikliai, tokie kaip pelningumas, likvidumas, įsiskolinimas ir veiklos efektyvumas, leidžia įvertinti įmonės finansinę būklę, tačiau egzogeniniai veiksniai – skirtingi ekonomikos ciklo etapai, sektoriaus dinamika, makroekonominiai rodikliai – dažnai lemia sisteminius rizikos pokyčius, kurių neįmanoma identifikuoti remiantis vien tik įmonės lygmens duomenimis. Todėl literatūroje vis dažniau pabrėžiamas integruotas požiūris, jungiantis mikro ir makro lygmens kintamuosius.

Analizuoti tyrimai taip pat parodė, kad duomenų struktūra yra vienas iš esminių veiksnių, lemiančių modelių rezultatus. Didelės apimties duomenų bazės leidžia pasiekti statistinį patikimumą ir analizuoti rizikos dinamiką ilguoju laikotarpiu, tačiau jos pasižymi heterogeniškumu ir struktūriniais pokyčiais. Tuo tarpu mažesnės imtys suteikia galimybę analizuoti individualias įmones, tačiau riboja rezultatų pritaikymą ateityje ir gali didinti persimokymo riziką. Be to, literatūra išryškino nesubalansuotų duomenų problemą, kai bankrutavusių įmonių dalis yra santykinai maža, o tai daro reikšmingą įtaką modelių tikslumui ir reikalauja specialių metodologinių sprendimų.

Svarbus aspektas, išryškėjęs literatūros analizėje, yra modelių jautrumas ekonominėms sąlygoms. Tyrimai rodo, kad modelių prognozių tikslumas kinta priklausomai nuo ekonomikos ciklo etapo, o kriziniais laikotarpiais net finansiškai stabilios įmonės gali tapti rizikingos. Tai leidžia teigti, kad statiniai modeliai yra nepakankami dinamiškoje ekonominėje aplinkoje, o bankroto prognozavimas turi būti grindžiamas laiko dimensija ir makroekonominį kontekstą įtraukiančiais metodais.

Apibendrinant galima teigti, kad literatūra patvirtina būtinybę taikyti kompleksinį požiūrį į bankroto prognozavimo problemą, integruojant skirtingus rodiklių tipus, duomenų struktūras ir metodologijas. Šios išvados sudaro teorinį pagrindą praktinei tyrimo daliai, kurioje bus siekiama empiriškai įvertinti pasirinktų endogeninių ir egzogeninių rodiklių reikšmę bei pagrįsti jų įtraukimą į kuriamą bankroto prognozavimo modelį.

## 2. Tyrimo metodologija ir duomenų analizės metodai

Šioje darbo dalyje aprašomi praktinėje darbo dalyje taikomi klasifikavimo metodai: regularizuota logistinė regresija, atsitiktiniai miškai, gradientinio stiprinimo metodai. Pateikiamos modelių tikslumo įvertinimo metrikos, tokios kaip ROC ir preciziškumo-atkuriamumo kreivės, bei sumaišymo matrica. Taip pat yra pateikiama trūkstumų duomenų užpildymo metodų apžvalga.

### 2.1. Logistinė regresija

Logistinė regresija yra vienas iš svarbiausių statistinių metodų, taikomų kintamojo, įgyjančio dvi reikšmes, analizei, kai siekiama įvertinti nepriklausomų kintamųjų poveikį įvykio tikimybei. Šis metodas yra plačiai taikomas ekonominiuose tyrimuose, jis leidžia modeliuoti tikimybę, kad priklausomas kintamasis įgis tam tikrą reikšmę [25].

Tarkime, kad turime nepriklausomų stebėjimų imtį  $(y_i, x_i)$ , čia  $y_i \in 0,1$  yra dvejetainis priklausomas kintamasis, o  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  –  $p$  - dimensinis nepriklausomų kintamųjų vektorius. Logistinės regresijos modelyje, tikimybė, kad priklausomas kintamasis įgis reikšmę 1, aprašoma logistine funkcija:

$$\pi(x_i) = \exp(x_i^T \beta) / (1 + \exp(x_i^T \beta)); \quad (1)$$

čia  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  yra nežinomų modelio parametrų vektorius [26].

Siekiant užtikrinti tiesinę priklausomybę tarp nepriklausomų kintamųjų ir transformuotos tikimybės, taikoma logaritminė transformacija:

$$\log(\pi(x_i)/(1 - \pi(x_i))) = x_i^T \beta; \quad (2)$$

modelio parametrų įverčiai gaunami maksimizuojant logaritmuotą tikėtimumo funkciją, kuri Bernulio skirstinio atveju išreiškiama:

$$l(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - \log(1 + \exp(x_i^T \beta))]. \quad (3)$$

Maksimalaus tikėtimumo metodas užtikrina statistiškai pagrįstus parametrų įverčius, tačiau didelių dimensijų duomenų atveju, šie įverčiai gali būti nestabilūs dėl multikolinearumo ir riboto stebėjimų skaičiaus. Be to, klasikinė logistinė regresija pasižymi ribotu gebėjimu kontroliuoti modelio sudėtingumą, todėl kyla persimokymo rizika [26].

Dėl šių priežasčių, logistinės regresijos modelis dažnai išplečiamas taikant regularizacijos metodus, kurie leidžia stabilizuoti parametrų įverčius ir atlikti kintamųjų atranką [27].

#### 2.1.1. Logistinė regresija su bauda

Logistinė regresija su bauda, grindžiama logaritminės tikėtimumo funkcijos modifikacija. Į ją įtraukiamas baudos narys, kuris riboja modelio sudėtingumą:

$$l_\lambda(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - \log(1 + \exp(x_i^T \beta))] - \lambda P(\beta); \quad (4)$$

čia  $\lambda > 0$  yra regularizacijos parametras, o  $P(\beta)$  – baudos funkcija [26].

Dažniausiai taikomos baudos yra *Ridge* (L2 norma) ir LASSO (L1 norma). *Ridge* regresija sumažina koeficientų dispersiją, tačiau neatlieka kintamųjų atrankos, tuo tarpu LASSO regresija leidžia atlikti kintamųjų atranką, tačiau gali būti nestabili, kai kintamieji yra tarpusavyje koreliuoti [27].

### 2.1.2. Logistinė regresija su regularizacija

*Elastic Net* regularizacija sujungia *Ridge* ir LASSO metodų privalumus, naudodama kombinuotą baudos funkciją. Tai leidžia vienu metu spręsti multikolinearumo ir kintamųjų atrankos problemas [27].

*Elastic Net* baudos funkcija apibrėžiama taip:

$$P_\alpha(\beta) = \sum_{j=1}^p [(\alpha/2)\beta_j^2 + (1 - \alpha)|\beta_j|]; \quad (5)$$

čia  $\alpha \in [0,1]$  yra parametras, nusakantis *Ridge* ir LASSO baudų santykį, o  $\beta_j$  – j-ojo nepriklausomo kintamojo koeficientą [26]. Atitinkamai *Elastic Net* regularizuotos logistinės regresijos tikslinė (angl. *objective*) funkcija išreiškiama:

$$l_\lambda(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - \log(1 + \exp(x_i^T \beta))] - \lambda \sum_{j=1}^p [(\alpha/2)\beta_j^2 + (1 - \alpha)|\beta_j|]; \quad (6)$$

Kai  $\alpha = 0$ , modelis sutampa su LASSO logistine regresija, o kai  $\alpha = 1$  – su *Ridge* logistine regresija. Tarpinės  $\alpha$  reikšmės leidžia pasiekti kompromisą tarp kintamųjų atrankos ir koeficientų stabilizavimo, ypač kai nepriklausomi kintamieji yra tarpusavyje koreliuoti [27]. *Elastic Net* regularizacija yra tinkama didelės dimensijos duomenų analizei, kadangi leidžia sumažinti modelio persimokymo riziką ir pagerinti prognozavimo tikslumą [26].

Siekiant įtraukti koreliacijos struktūrą tarp nepriklausomų kintamųjų, pasiūlytas *Adjusted Elastic Net* (AEN) metodas, kuris modifikuoja *Elastic Net* baudos narį, papildomai įtraukdamas koreliacijos koeficientus tarp kintamųjų [27].

AEN tikslinė (angl. *objective*) funkcija logistinėje regresijoje apibrėžiama taip:

$$l_{AEN}(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - \log(1 + \exp(x_i^T \beta))] - \lambda_1 \sum_{j=1}^p |\beta_j| - \lambda_2 \sum_{j < k} (\beta_j - r_{jk} \beta_k)^2; \quad (7)$$

čia:  $r_{jk}$  – empirinis koreliacijos koeficientas tarp  $j$  ir  $k$  kintamųjų,  $\lambda_1$  ir  $\lambda_2$  – regularizacijos parametrai [27].

Dar labiau išplėstas metodas yra *Adaptive Adjusted Elastic Net* (AAEN), kuriame naudojami adaptaciniai svoriai, leidžiantys pagerinti kintamųjų atrankos stabilumą ir prognozinės savybes [27]. AAEN tikslinė (angl. *objective*) funkcija apibrėžiama taip:

$$l_{AAEN}(\beta) = \sum_{i=1}^n [y_i(x_i^T \beta) - \log(1 + \exp(x_i^T \beta))]$$

$$-\lambda_1 \sum_{j=1}^p w_j |\beta_j| - \lambda_2 \sum_{j < k} (\beta_j - r_j k \beta_k)^2 ; \quad (8)$$

$$\text{čia adaptaciniai svoriai apibrėžiami: } w_j = 1/|\beta_j(AEN)|^\gamma ; \quad (9)$$

čia  $\beta_j(AEN)$  – AEN metodo gautas koeficiento įvertis, o  $\gamma > 0$  – adaptacijos parametras (dažniausiai  $\gamma = 1$ ) [27].

### 2.1.3. Praktinis logistinės regresijos taikymas

Tyrimo logistinės regresijos modeliui optimizuoti buvo naudojamas R paketas *glmnet*. Šis paketas taiko baudos didžiausio tikėtimumo funkciją. Toliau aprašomi parametrai, nusakantys logistinės regresijos modelį.

Parametras  $\alpha$  kontroliuoja balansą tarp LASSO ir *Ridge* baudų, jis gali įgyti reikšmes nuo 0 iki 1. Tyrimo metu optimaliai  $\alpha$  parametro reikšmei rasti buvo atlikta tinklelio paieška, kurios metu  $\alpha$  reikšmėms leista keistis nuo 0 iki 1 su žingsniu 0,1. Skirtingoms  $\alpha$  reikšmėms vertinti, kryžminės patikros metu buvo naudoti du kriterijai:

- *type.measure* = „*auc*“ – plotas po ROC kreive. Ši metrika skaičiuojama optimizuojant modelį pagal jo gebėjimą teisingai atskirti dviejų klasių elementus;
- *type.measure* = „*deviance*“ – binominė paklaida. Ši metrika veikia kaip modelio klaidos funkcija. Ji matuoja kaip tiksliai modelio prognozuojamos tikimybės atitinka faktines reikšmes ir už užtikrintus klaidingus sprendimus modelį baudžia labiau.

Parametras  $\lambda$  nusako bendrą regularizacijos stiprumą. Siekiant parinkti geriausią  $\lambda$  reikšmę, kiekvienam modeliui, buvo atliktas 5 dalių kryžminė validacija, jos metu kiekvienam modeliui buvo siekiama surasti dvi  $\lambda$  reikšmes:

- *lambda.min* – tai baudos reikšmė, prie kurios yra pasiekama minimali kryžminės patikros klaida;
- *lambda.1se* – regularizuoto modelio baudos reikšmė, kurios klaida patenka į vienos standartinės paklaidos rėžius nuo minimalios klaidos dydžio gauto su *lambda.min*. Šis kintamasis dažnai pasirenkamas, siekiant sudaryti paprastesnį, mažiau kintamųjų turintį modelį [28].

Kadangi duomenyse fiksuotas stiprus klasių disbalansas, į funkciją perduotas papildomas svorio parametras *weights*. Jis užtikrina stebėjimų svorių perskaičiavimą: sveikos įmonės vertinamos įprastai (svoris lygus 1), o bankrutuojančių įmonių stebėjimai gauna padidintą svorio reikšmę (*ratio\_weight*). Norint palyginti svorių įtaką, vienas iš modelių (*fit\_no\_weights*) buvo apmokytas nenaudojant šio parametro.

## 2.2. Atsitiktiniai miškai

Atsitiktinių miškų (angl. *random forest*) RF metodas yra vienas efektyviausių ansamblinio mokymosi algoritmų, pagrįstas sprendimų medžių (angl. *decision trees*) visuma. Remiantis [29], šis metodas yra klasifikavimo ir regresijos medžių (angl. *classification and regression trees*) CART metodikos tęsinys, skirtas pagerinti modelio prognozavimo tikslumą bei sumažinti pavieniams medžiams būdingą didelę dispersiją.

### 2.2.1. Atsitiktinių miškų algoritmas

Atsitiktinių miškų algoritmas remiasi duomenų agregavimo (angl. *bagging* arba *bootstrap aggregating*) principu. Kiekvienas individualus sprendimų medis miške formuojamas naudojant atskirą atsitiktinę duomenų imtį su gražinimu (angl. *bootstrap sample*), kurią paprastai sudaro apie du trečdaliai originalių stebėjimų. Likusi duomenų dalis (angl. *Out-Of-Bag*) OOB, naudojama modelio tikslumui vertinti.

Esminis atsitiktinių miškų skirtumas nuo standartinio duomenų agregavimo (angl. *bagging*) metodo yra atsitiktinis nuspėjamųjų kintamųjų parinkimas kiekviename medžio mazgo skaidymo etape. Užuovertinus visus  $p$  kintamuosius, algoritmas atsitiktinai atrenka kintamųjų poaibį (*mtry*), iš kurio parenkamas geriausias skaidymo kriterijus. Ši strategija leidžia mažinti miško medžių koreliaciją, taip sumažinant galutinę modelio dispersiją ir užtikrinant geresnį modelio pritaikymą naujiems duomenims.

### 2.2.2. Atsitiktinių miškų klasifikavimo procesas

RF klasifikavimo metodas veikia „daugumos balso“ principu. Procesą galima apibrėžti šiais etapais:

- iš turimų duomenų sudaroma  $B$  skirtingų duomenų agregavimo (angl. *bootstrap*) imčių;
- kiekvienai imčiai sukonstruojamas atskiras sprendimų medis;
- kiekvienas medis pateikia savo prognozę: (pvz: įmonė „bankrutavo“ arba „nebankrutavo“);
- galutinis modelio sprendimas priimamas susumavus visų medžių rezultatus – įmonei priskiriama ta klasė, kuri pasikartojo daugumos medžių rezultatuose.

Matematiškai klasifikavimo taisyklė išreiškiama taip:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x); \quad (10)$$

čia  $\hat{f}_b(x)$  yra  $b$  – ojo medžio prognozuojama klasė [30].

### 2.2.3. Atsitiktinių miškų metodo privalumai

Remiantis [29], pagrindiniai šio klasifikavimo metodo privalumai yra:

- netiesinių sąveikų identifikavimas: algoritmas aptinka sudėtingus ryšius tarp rodiklių be išankstinių prielaidų;
- kintamųjų svarbos vertinimas: leidžia nustatyti, kurie kintamieji turi didžiausią įtaką teisingai klasifikuojant;
- atsparumas modelio persimokymui: didėjant medžių skaičiui, modelis tampa stabilesnis ir geriau apibendrina naujus duomenis.

## 2.2.4. Praktinis atsitiktinių miškų metodo taikymas

Tyrime atsitiktinių miškų modeliui optimizuoti buvo naudojamas R paketas *randomForest*. Toliau aprašomi parametrai, nusakantys atsitiktinių miškų klasifikavimo modelį.

Parametras *n tree* nurodo, kiek sprendimų medžių patenka į kiekvieną mišką. Remiantis dokumentacija [31], šis skaičius neturėtų būti per mažas, kad užtikrintų stabilų kintamųjų svarbos įvertinimą ir paklaidos stabilizavimą.

Parametras *m try* nusako kintamųjų skaičių, kuris yra atsitiktinai atrenkamas ir vertinamas medžio šakojimosi momentu. Tai yra pagrindinis parametras, skirtas reguliuoti medžių tarpusavio koreliaciją miško viduje. Dokumentacijoje nurodyta [31], kad klasifikavimo uždaviniams standartiškai taikoma kvadratinės šaknies iš bendro kintamųjų skaičiaus reikšmė. Tyrimo metu išbandytos dvi šio parametro strategijos:

- standartinė reikšmė – toks nustatymas užtikrina didesnę medžių įvairovę, kadangi kiekvienas šakojimasis mato nedidelę požymių dalį;
- padidinta reikšmė – vertinamų kintamųjų skaičius padidintas iki trečdaliao visų kintamųjų.

Dokumentacijoje *classwt* apibrėžiamas kaip klasių svorio parametras, naudojamas asimetrinei klaidų kainai nustatyti. Parametro reikšmė *ratio\_weight*, modelį veikia taip, kad skaičiuojant Gini indekso reikšmę, kiekvienas praleistas bankroto atvejis modeliui suteiktų didesnę baudą, nei už klaidingai suklasifikuotą veikiančią įmonę. Šio rodiklio reikšmė apskaičiuojama kaip veikiančių ir bankrutavusių įmonių skaičiaus santykis apmokymo imtyje.

Parametras *sampsiz*e nurodo stebėjimų skaičių, taikant atsitiktinę imtį su gražinimu, atrenkamą kiekvieno medžio auginimui. Klasifikavimo uždaviniuose šis parametras gali būti perduodamas kaip vektorius [31], nurodantis tikslų kiekvienos klasės stebėjimų skaičių. Viename iš modelių buvo pritaikyta subalansavimo strategija, čia nurodyta, kad į kiekvieną medį patektų vienodas kiekis bankrutavusių ir veikiančių įmonių.

## 2.3. Gradientinio stiprinimo algoritmai

Remiantis [32], gradientinio stiprinimo metodas (GBM) yra viena galingiausių mašininio mokymosi klasifikavimo technikų. Autoriai šį algoritmą apibrėžia per funkcinio optimizavimo prizmę – modelis yra konstruojamas iš silpnų klasifikatorių (dažniausiai sprendimo medžių). Algoritmas veikia iteratyviai – pradedama nuo priminės prognozės, tada kiekviename žingsnyje naujas sprendimų medis yra pridodamas, siekiant minimizuoti bendrą modelio klaidą.

### 2.3.1. GBM matematinė išraiška ir sąsaja su R funkcijomis

Praktinis modelio realizavimas tyrime atliktas naudojant R *gbm* paketą. Šis procesas tiesiogiai remiasi [32] apibrėžta teorija.

Autoriai nurodo, kad algoritmo tikslas yra surasti tokią prognozavimo funkciją  $F$ , kuri pagal pasirinktą nuostolių funkciją  $L$  minimizuotų riziką  $C_n(F)$ :

$$C_n(F) = \frac{1}{n} \sum_{i=1}^n L(F(X_i), Y_i). \quad (11)$$

Kadangi uždavinys yra binarinė įmonių bankroto klasifikacija, pasirinktas Bernulio skirstinys. Tai algoritmui nurodo optimizuoti binominę paklaidos funkciją, kur  $Y_i \in \{0, 1\}$  žymi faktinę, o  $X_i$  – modelio prognozuojamą įmonės būseną.

Remiantis [32] metodologija, šiame tyrime netaikoma ankstyvojo sustabdymo strategija – algoritmas vykdomas iki galo, kol pasiekia nustatytą iteracijų skaičių  $M$ . R kode bendrą nuoseklių medžių skaičių nustato parametras *n.trees*, o kintamųjų sąveikos lygį kontroliuoja parametras *interaction.depth*.

Užauginus medį  $h_m$ , esamas modelis atnaujinamas pritaikant adityviają taisyklę [32]:

$$F_m = F_{m-1} + \lambda h_m; \quad (12)$$

čia  $\lambda$  yra mokymosi greitis (*shrinkage* parametras), veikiantis kaip statistinis regularizatorius. Jis sumažina kiekvieno naujo medžio įtaką – tokiu būdu užtikrinamas tolygus nuostolių funkcijos minimizavimas, bei modelis apsaugomas nuo persimokymo.

Taip pat praktinėje dalyje naudojamas parametras *bag.fraction*, kuris nurodo, kokią dalį imties stebėjimų modelis pasirenka prieš kiekvieną iteraciją. Parametro reikšmė lygi 0,5 reiškia, kad prieš kiekvieną iteraciją algoritmas atsitiktine tvarka atrenka tik pusę mokymo imties stebėjimų. Remiantis straipsnio autoriais, mokinimasis iš skirtingų poabių padidina modelio stabilumą bei jo galimybes pritaikyti prie naujų duomenų.

R *gbm* paketas taip pat turi galimybę įvertinti sąlyginę koeficientų įtaką (angl. *relative influence*) modeliui. Šis metodas visiems medžiams sumuoja kvadratinės paklaidos mažėjimą ties kiekvienu medžio šakojimosi tašku. Gauti rezultatai normalizuoti taip, kad bendra visų kintamųjų svarbos suma sudarytų 100 %.

## 2.4. Statistinės metrikos

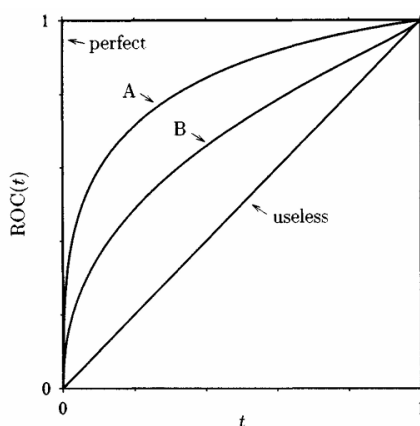
Šioje darbo dalyje aprašomos praktinėje darbo dalyje klasifikavimo metodų tikslumui vertinti naudojamos statistinės metrikos.

### 2.4.1. ROC kreivė

Veiklos charakteristikos kreivė (angl. *receiver operating characteristic curve*) ROC yra vienas svarbiausių statistinių įrankių, naudojamų klasifikavimo modelių diagnostiniam tikslumui vertinti. Istoriskai šis metodas kilo iš signalų detekcijos teorijos [33]. Toliau aprašomi pagrindiniai ROC kreivės apibrėžimai, bei tikslumo matavimo aspektai.

ROC kreivė parodo ryšį tarp teisingų teigiamos grupės rezultatų (angl. *true positive fraction*) TPF, dar vadinamos jautrumu ir teisingų neigiamos grupės rezultatų (angl. *false positive fraction*) FPF, kurie yra įvardijami kaip iš vieneto atimta specifiškumo reikšmė. ROC kreivė apibūdina šį ryšį skirtingiems klasifikavimo slenksčiams. Pagal nusistovėjusią tvarką daroma prielaida, kad didesnės testo reikšmės  $Y$  rodo didesnę tiriamos būsenos (pvz., bankroto) tikimybę. Klasifikavimo taisyklė apibrėžiama per slenkstį  $c$ : rezultatas laikomas teigiamu, jei  $Y \geq c$ , ir neigiamu, jei  $Y < c$ . Pačią ROC kreivę sudaro visa aibė galimų (FPF, TPF) taškų porų, gaunamų slenksčio  $c$  reikšmei varijuojant nuo  $-\infty$  iki  $+\infty$ . Kadangi didėjant slenksčiui abi grupės dėsningai mažėja, ROC kreivė matematiškai apibrėžiama kaip visada monotoniškai didėjanti funkcija teigiamame vienetiniame kvadrato.

Modelio tikslumo vertinimas remiantis ROC kreive, gali būti vizualiai nustatomas, remiantis kreivės padėtimi grafike. Tokia, ROC kreivė, kuri sutampa su 45 laipsnių įstrižaine ( $ROC(t)=t$ ) yra neinformatyvi, o tobulo testo, visiškai atskiriančio grupes, kreivė eina kairiuoju ir viršutiniu vienetinio kvadrato kraštais. Apibendrinant kreivę, dažniausiai naudojamas ploto po kreive rodiklis (AUC). Šio rodiklio reikšmė yra interpretuojama kaip tikimybė, kad atsitiktinai parinktas subjektas iš teigiamos grupės turės didesnę klasifikatoriaus priskirtą reikšmę nei atsitiktinis subjektas iš neigiamos grupės [1].

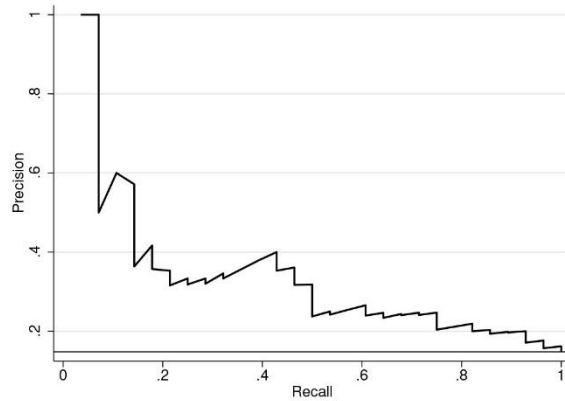


1 pav. ROC kreivių palyginimas [1]

#### 2.4.2. Preciziškumo-atkuriamumo kreivė

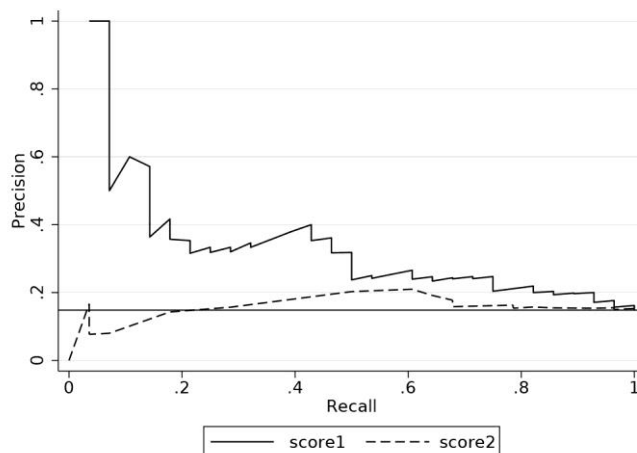
Preciziškumo-atkuriamumo (angl. *precision-recall*) PR kreivė yra diagnostinė priemonė, kuri vaizduoja atkuriamumo preciziškumą, kiekvienai galimai slenksčio reikšmei. Ši kreivė dažniausiai naudojama modelio veiksmingumui vertinti, kai duomenyse yra pastebimas klasių disbalansas – teigiamoji klasė sąlyginai retai aptinkama duomenyse. Todėl PR kreivių analizės metu yra tinkama nagrinėti preciziškumo ir atkuriamumo rodiklius [34].

PR kreivė parodo ryšį, tarp teisingai priskirtų teigiamos klasės elementų santykio, su visomis klasifikatoriaus priskirtomis teigiamomis reikšmėmis (preciziškumo), ir teisingų teigiamos grupės rezultatų TPF, dar vadinamos atkuriamumu. Grafiko koordinatinių plokštumos x ašyje yra atkuriamumo rodiklis, parodantis teisingai suklasifikuotų teigiamos klasės elementų dalį tarp visų elementų, originaliai duomenyse priskiriamų teigiamai klasei. Vertikaliojoje (y ašyje) yra preciziškumo rodiklis, matuojantis teisingai suklasifikuotų teigiamos klasės elementų dalį tarp visų (teisingai ir neteisingai) elementų klasifikavimo metu priskirtų teigiamai klasei. Slenksčio riba yra ribinė vertė, lemianti viršijančių ją modelio priskirtų balų patekimą į teigiamą klasę ir mažesnių balų priskyrimą neigiamai klasei. Slenksčio reikšmei keičiantis, kartu su ja keičiasi preciziškumo ir atkuriamumo metrikų vertės, šių porų atvaizdavimas visoms įmanomoms slenksčio reikšmėms sudaro grafike matomą kreivę.



2 pav. PR kreivė [34]

Modelio tikslumo vertinimas naudojant PR kreivę yra svarbus duomenims, kuriuose yra didelis klasių disbalansas. Ši kreivė neleidžia dideliame neigiamos klasės stebėjimų skaičiui paslėpti modelio daromų klaidų, kadangi net ir nedidelis klaidų skaičius gali reikšti staigų kreivės kritimą. Lyginant su ROC kreive, ji tam pačiam klasifikavimo metodui gali rodyti žymiai geresnius rezultatus nei PR kreivė, kadangi ROC kreivė ne taip gerai įvertina retus įvykius. Todėl esant dideliame duomenų disbalansui, straipsnio autoriai siūlo kreipti daugiau dėmesio į PR kreivę, bei F1 rodiklio reikšmę [34].



3 pav. PR kreivių palyginimas [34]

### 2.4.3. Sumaišymo matrica

Sumaišymo matrica – tai struktūrizuota lentelė, skirta įvertinti klasifikavimo uždavinio rezultatus. Joje tikrosios duomenų rinkinyje esančios klasės yra lyginamos su klasifikatoriaus priskirtomis klasėmis. Dviejų klasių klasifikavimo atveju duomenų rinkinys suskirstomas į dvi kategorijas: teigiamąją ir neigiamąją klases. Matricoje pateikiami keturi galimi rezultatai, priklausomai nuo to, ar klasifikatoriaus prognozė buvo teisinga, ar klaidinga. Siekiant priimti sprendimą dėl klasės priskyrimo yra naudojama slenksčio reikšmė. Kiekvienam elementui klasifikatorius priskiria balą, kuris gali būti didesnis, arba mažesnis už nustatytą slenksčio ribą. Tokios metrikos kaip AUC yra tiesiogiai susijusios su nustatyta slenksčio reikšme.

Pagrindiniai sumaišymo matricos komponentai yra:

- teisingai teigiami (angl. *true positive*) TP: klasifikavimo metu elementams yra teisingai priskiriama teigiama klasė;
- teisingai neigiami (angl. *true negative*) TN: klasifikavimo metu elementams yra teisingai priskiriama neigiama klasė;
- klaidingai teigiami (angl. *false positive*) FP: klasifikavimo metu elementams yra neteisingai priskiriama teigiama klasė, dar vadinama pirmojo tipo klaida;
- klaidingai neigiami (angl. *false negative*) FN: klasifikavimo metu elementams yra neteisingai priskiriama neigiama klasė, dar vadinama antrojo tipo klaida.

		Predicted Class		Instances
		P	N	
Actual Class	P	TP $\lambda_{PP}m_P$	FN $(1 - \lambda_{PP})m_P$	$m_P$
	N	FP $(1 - \lambda_{NN})m_N$	TN $\lambda_{NN}m_N$	$m_N$
Estimations		$e_P$	$e_N$	$m$

4 pav. Sumaišymo matrica [35]

Remiantis pagrindiniais sumaišymo matricos komponentais, yra išskaičiuojamos papildomos metrikos, padedančios įvertinti modelių patikimumą. Jautrumo rodiklis parodo teigiamų elementų dalį, kurią klasifikatorius teisingai priskiria teigiamai klasei. Specifiškumo rodiklis parodo neigiamų elementų dalį, kurią klasifikatorius teisingai priskyrė neigiamai klasei. Preciziškumo rodiklis nusako, kokia dalis klasifikatoriaus rodiklių priskirtų teigiamai klasei, iš tiesų jai priklauso. Straipsnio [35] autoriai pabrėžia, kad šį rodiklį gali stipriai veikti klasių disbalansas. Tikslumo – bendro teisingumo rodiklis – parodo bendrą teisingai suklasifikuotų elementų santykį su visais elementais. Ši metrika naudojama dažnai, tačiau yra jautri klasių disbalansui duomenyse [35]. F1 metrikos reikšmė yra naudojama, siekiant subalansuotų preciziškumo ir jautrumo rodiklių, tačiau straipsnyje [35] yra pabrėžiama, kad klasių disbalansas duomenyse gali šią reikšmę iškreipti. Kito straipsnio autoriai, esant klasių disbalansui, labiau fokusuojasi į atkuriamumo ir preciziškumo rodiklių reikšmes. Jie pabrėžia, kad skirtingos klaidos įmonių prognozavimo kontekste turi skirtingą kainą: bankrutuojančios įmonės priskyrimas nebankrutuojančiųjų klasei (FN) gali kainuoti žymiai daugiau, nei veikiančios įmonės priskyrimas bankrutavusiųjų klasei (FP) [36].

**5 lentelė.** Išvestinės sumaišymo matricos metrikos

Simbolis	Metrika	Formulė
SNS	Jautrumas (angl. <i>sensitivity</i> ), atkuriamumas (angl. <i>recall</i> )	$\frac{TP}{TP + FN}$
SPC	Specifiškumas (angl. <i>specificity</i> )	$\frac{TN}{TN + FP}$
PRC	Preciziškumas (angl. <i>precision</i> )	$\frac{TP}{TP + FP}$
ACC	Tikslumas	$\frac{TP + TN}{TP + FN + TN + FP}$
F1	F1 reikšmė	$2 \frac{PRC \cdot SNS}{PRC + SNS}$

## 2.5. Kintamųjų svarba

Spearmano ranginės koreliacijos koeficientas yra naudojamas nustatyti dviejų kintamųjų tarpusavio santykį, čia yra naudojama ne koreliacija tarp pačių reikšmių, tačiau tarp sudarytų rangų. Spearmano koeficiento reikšmės gali svyruoti tarp minus ir plus vieneto. Teigiama rodiklio reikšmė indikuoja, kad abu rodikliai vienu metu arba kyla, arba leidžiasi, o neigiama rodiklio reikšmė indikuoja, kad vienam rodikliui kylant, kitas rodiklis leidžiasi žemyn. Spearmano koreliacijos koeficiento reikšmė lygi nuliui reiškia ryšio tarp dviejų kintamųjų nebuvimą. Šios koreliacijos koeficientų interpretavimas priklauso nuo duomenų kiekio: jei duomenų kiekis yra pakankamai didelis, gali būti nagrinėjami ir mažesni koeficientai, o rodiklio patikimumas mažame duomenų rinkinyje sumažėja, kadangi gali neatspindėti tikros situacijos [37].

## 2.6. Metodai ir duomenys

Tyrime naudojamuose duomenyse pastebėtas didelis klasių disbalansas, todėl kilo poreikis užpildyti mažąją duomenų rinkinio dalį, sudarančios klasės įmonių, trūkstamus finansinius duomenis.

### 2.6.1. Tradiciniai vienos reikšmės užpildymo metodai

Remiantis straipsniu, kuriame yra aprašomi duomenų užpildymo metodai [38], paprasčiausi iš jų yra pildymas žemyn (angl. *last observation carried forward*) LOCF arba aukštyn (angl. *next observation carried backward*) NOCB. Abipusis pildymas yra traktuojamas kaip geras

būdas, siekiant išvengti duomenų praradimo sekos pradžioje. Tokie metodai skaičiavimo prasme yra greitai ir paprasti, todėl teisingai grupuojant duomenis, galima užtikrinti, kad pildymas vyktų tik tos pačios objekto klasės ribose. Metodas labiausiai tinka tada, kai duomenų trūksta atsitiktinėse vietose, nėra kelių iš eilės praleistų reikšmių. Pagrindinis pabrėžiamas šio metodo trūkumas yra šališkumas, kadangi vėliau naudojamas klasifikavimo modelis yra apmokomas ant užpildytų duomenų ir gali užfiksuoti didesnį duomenų stabilumą nei jis yra iš tiesų [38].

### 2.6.2. MICE trūkstamų reikšmių užpildymo metodas

MICE trūkstamų reikšmių užpildymo metodas veikia iteraciniu principu – čia kiekvienas kintamasis, turintis trūkstamų reikšmių yra užpildomas atskirai. Šis pildymo metodas leidžia išlaikyti sudėtingus ryšius tarp duomenyse esančių rodiklių. Pagrindinis šio proceso privalumas yra galimybė pildymo metu naudoti sudėtingesnius algoritmus, tokius kaip atsitiktiniai miškai, kurie neturi išankstinės normaliojo pasiskirstymo nuostatos. Toks metodų derinys leidžia fiksuoti netiesinius ryšius ir kintamųjų sąveikas jų nespacificuojant. Straipsnyje [39] nurodoma, kad MICE kombinacija su atsitiktiniais miškais pasižymi mažesniu šališkumu ir geresniu neapibrėžtumo įvertinimu, ypač tada, kai duomenyse randami sudėtingi ryšiai.

### 2.6.3. Duomenų aprašymas

Duomenyse [40] esančius kintamuosius galima suskirstyti į keturias pagrindines grupes: bendrinė informacija apie įmonę, pagrindiniai finansiniai rodikliai, santykiniai finansiniai rodikliai, bei bankroto indikatoriai. Remiantis stulpelių „finance\_year\_start\_date“ ir „finance\_year\_end\_date“ reikšmėmis, galima teigti, kad duomenys apie įmones yra pateikiami nuo 2004 metų sausio pirmos dienos iki 2022 metų rugpjūčio 31 d. Duomenų analizės etape, naudojant pagrindinius finansinius rodiklius, reikia atkreipti dėmesį į valiutos pokytį, kadangi 2015 metų sausio pirmą dieną Lietuvoje valstybinę lito valiutą pakeitė euro valiuta. Duomenis iš viso sudaro 805176 eilutės ir 24 stulpeliai, skirtingų unikalų įmonių šiame duomenų rinkinyje yra 161115, kadangi viena įmonė gali turėti kelis įrašus skirtingiems finansiniams metams. Iš šių įmonių 46591 (263781 duomenų eilutės) turi pilną duomenų rinkinį be trūkstamų duomenų ir daugiau nei 70% (114524) įmonių duomenyse yra bent viena trūkstama reikšmė. Bendrinę įmonės informaciją nusakantys, bei bankroto indikatorių stulpeliai duomenys visada turi ne tuščią reikšmę. Svarbu paminėti, kad tokie stulpeliai kaip „registration\_name“, „legal\_form\_name“, „status\_name“ bėgant metams tai pačiai įmonei („code“) duomenyse nekinta.

### 2.6.4. Kintamųjų sąrašas

Žemiau matomoje lentelėje yra pateikiama bendrinė informacija apie įmonę. Įmonės kodas, įregistravimo data, teisinė forma, įmonės statusas. Kintantys stulpeliai yra finansinių metų pradžios ir pabaigos datos, bei dokumentų registravimo data – viena įmonė duomenyse dažnai turi daugiau nei vienerių finansinių metų informaciją.

**6 lentelė.** Bendrinė informacija apie įmonę pateikiama duomenyse

Kintamojo pavadinimas	Paaiškinimas
-----------------------	--------------

code	Įmonės identifikavimo kodas
registration_date	Įmonės įregistravimo data
legal_form_name	Įmonės teisinė forma
status_name	Dabartinis įmonės statusas
finance_year_start_date	Finansinių metų pradžios data
finance_year_end_date	Finansinių metų pabaigos data
document_registration_date	Finansinės atskaitomybės dokumentų reistravimo data už finansinius metus

Diskrečių baigtinių reikšmių sąrašą įgyjančių stulpelių galimos reikšmės:

- legal\_form\_name: "Private Limited Liability Company", "Public Limited Liability Company", "Budget Institution", "Cooperative Society (Cooperative)", "Association", "Public Institution", "State Enterprise", "Chamber of Commerce, Industry and Craft", "Assoc. of Lithuanian Chambers of Commerce Industry and Crafts", "Small Partnership", "Agricultural Company", "Branch of the Foreign Enterprise", "Repr office of the Foreign Enterprise", "Individual Enterprise", "General Partnership", "Limited Partnership", "Charity and Sponsorship Fund", "Municipal Enterprise", "Community", "European Company";
- status\_name: "Under Liquidation", "Liquidation being initiated", "Removed", "No legal proceedings", "Bankrupt", "Liquidation due to bankruptcy", "Under Restructuring", "Partic. in Reorganization", "Partic. in division", "Going Bankrupt", "Under Reorganization", "Under Reformation", "Cross-border merging of a legal entity", "Legal entity partic. in cross-border merging".

Pagrindiniai finansiniai rodikliai pateikiami žemiau matomoje lentelėje, čia galima matyti, kad tokie finansiniai rodikliai kaip nuosavas kapitalas, bei ilgalaikis ir trumpalaikis turtas ataskaitose yra pildomi geriausiai.

**7 lentelė.** Pagrindiniai įmonės finansiniai rodikliai pateikiami duomenyse

Kintamojo pavadinimas	Paaiškinimas	Trūkstantų reikšmių kiekis
equity	Nuosavas kapitalas	3207
liabilities	Įsipareigojimai	89468
long_term_assets	Ilgalaikis turtas	6327
short_term_assets	Trumpalaikis turtas	6891
profit_before_taxes	Pelnas iki mokesčių	259132
net_profit	Grynasis pelnas	108665
sales_revenue	Pardavimo pajamos	182584

Santykiniai finansiniai rodikliai yra išskaičiuojami iš pagrindinių finansinių rodiklių, todėl logiška čia matyti didesnį trūkstumų reikšmių kiekį.

**8 lentelė.** Santykiniai įmonės finansiniai rodikliai pateikiami duomenyse

Kintamojo pavadinimas	Paaiškinimas	Trūkstumų reikšmių kiekis
idx_debt_to_equity_ratio	Skolos ir nuosavo kapitalo santykis	139993
idx_sh_as_lia_ratio	Trumpalaikio turto ir visų įsipareigojimų santykis	181393
idx_equity_ratio	Nuosavo kapitalo koeficientas	40257
idx_lia_as_ratio	Bendras skolos rodiklis	117078
idx_n_prof_as_ratio	Investicijų grąža	128916
idx_s_rev_as_ratio	Turto apyvartumo rodiklis	189814
idx_prof_tax_eq_ratio	Nuosavo kapitalo grąža	270604
idx_prof_tax_as_ratio	Turto pelningumo rodiklis	274112

Toliau pateikiami įmonės bankroto indikatoriai – kintamasis „is\_bankrupted\_last\_year“ yra klasifikavimo uždavinio priklausomo kintamojo „is\_bankrupted\_next\_year“ sukūrimo šaltinis.

**9 lentelė.** Bankroto indikacijos stulpelių reikšmės pateikiamos duomenyse

Kintamojo pavadinimas	Paaiškinimas
is_bankrupted_last_year	Parodo, ar įmonei buvo iškelta bankroto byla paskutiniuose duomenyse turimais metais (vieneto reikšmė matoma tik paskutiniuose duomenyse esančiais metais)
is_bankrupted_all_years	Parodo, ar įmonė per visą savo istoriją turėjo bankroto statusą (vieneto reikšmė matoma visais duomenyse esančiais metais)

### 2.6.5. Žvalgomoji analizė

Unikaliam įmonės kodui filtruojant pirmus duomenyse esančius finansinius metus (angl. *financial year*) FY, galima matyti, kad 59,58 % įmonių finansinių metų prasideda sausio mėnesį. Unikaliam įmonės kodui filtruojant paskutinius duomenyse esančius finansinius metus, kurie yra aktualiausi, galima matyti, kad 90,67 % įmonių paskutiniai finansiniai metai yra sausio mėnuo. Pirmi finansiniai metai kartais gali būti artimi įmonės įkūrimo datai, kuri nebūtinai yra kalendorinių metų pradžia, bet toliau įmonė įprastai finansinius metus pradeda sausio mėnesį. Galima pastebėti, kad bendras įmonių skaičius tarp dviejų paskutinių lentelės stulpelių skiriasi, taip yra todėl, kad 22666 įmonės duomenų rinkinyje turi informaciją tik apie vienus finansinius metus. Ši informacija leidžia dar tiksliau atfiltruoti įmonių finansinių metų pradžią, pagal paskutinius įmonės duomenų rinkinyje esančius finansinius metus, su sąlyga, kad įmonė turi duomenų daugiau nei vieneriems finansiniams metams. Pasidaro aišku, kad 99,36 % įmonių, turinčių duomenis daugiau nei vieneriems metams, finansinių metų pradžia yra sausio mėnuo. Kadangi tik 0,64 % įmonių finansiniai metai prasideda ne

sausio mėnesį, makroekonominis rodiklius galima naudoti apskaičiuotus kalendoriniams metams.

**10 lentelė.** Įmonių pasiskirstymas pagal finansinių metų pradžios mėnesį

<b>Mėnuo</b>	<b>Įmonių pirmieji FY</b>	<b>Įmonių paskutiniai FY</b>	<b>Įmonių paskutiniai FY įmonėms duomenyse turinčioms daugiau nei 1 FY</b>
Sausis	96002	146086	137564
Vasaris	6054	1364	34
Kovas	6723	1721	50
Balandis	6181	1560	107
Gegužė	5886	1458	94
Birželis	5712	1285	42
Liepa	5772	1449	258
Rugpjūtis	5609	1209	27
Rugsėjis	6342	1348	84
Spalis	6531	1419	107
Lapkritis	5878	1277	18
Gruodis	4425	939	64
Suma	161115	161115	138449

Toliau tikrinama įmonių istorija – kiek skirtingų finansinių metų pradžios mėnesių gali turėti viena įmonė. Galima matyti, kad 68,42 % įmonių finansinių metų pradžios mėnesio nekeičia, arba keičia jį tik vieną kartą (31,5 %). Todėl pridamas dar vienas lentelės stulpelis, siekiant parodyti, kad toks pokytis dažniausiai įvyksta po pirmųjų finansinių metų pabaigos, filtruojant finansinių metų pradžią, neįskaitant pirmųjų finansinių metų pradžios. Čia matomas dar reikšmingesnis skirtumas – 99,6 % įmonių per visą savo gyvavimo laikotarpį matomą duomenyse, nuo antrųjų finansinių metų, turi pastovią pradžią.

**11 lentelė.** Įmonių pasiskirstymas pagal finansinių metų pradžios mėnesio keitimosi dažnį

<b>Skirtingų FY mėnesių skaičius vienai įmonei</b>	<b>Įmonių skaičius</b>	<b>Įmonių skaičius, neįskaitant pirmųjų FY pradžios</b>
1	110233	137900
2	50759	543
3	120	4
4	3	2
Suma	161115	138449

Toliau yra nagrinėjamos įmonės, kurios duomenų rinkinyje turi informaciją tik apie vienerius finansinius metus. Čia galima pastebėti, kad 78,69 % šių įmonių nebankrutavo, nebuvo likviduota ir nevyko reorganizacija, todėl tolesnių duomenų nebuvimas nereiškia, kad įmonė bankrutavo. Iš 22666 įmonių tik 181 įmonė turi bankroto indikaciją stulpeliuose „is\_bankrupted\_last\_year“ ir „is\_bankrupted\_all\_years“.

**12 lentelė.** Įmonių, turinčių vieną įrašą, pasiskirstymas pagal teisinį statusą

<b>Įmonės statusas</b>	<b>Įmonių skaičius</b>
No legal proceedings	17836
Removed	3328
Liquidation being initiated	994
Under Liquidation	298
Liquidation due to bankruptcy	93
Going Bankrupt	62
Bankrupt	26
Under Reformation	13
Under Reorganization	13
Partic. in Reorganization	2
Partic. in division	1
Suma	22666

Iš viso duomenyse galima rasti 1635 bankrutavusias įmones, toliau pateikiamas bankrutavusių įmonių pasiskirstymas pagal įmonės statusą.

**13 lentelė.** Bankrutavusių įmonių pasiskirstymas pagal teisinį statusą

<b>Įmonės statusas</b>	<b>Įmonių skaičius</b>
Liquidation due to bankruptcy	851
Going Bankrupt	523
Bankrupt	209
Under Restructuring	52

Pateikiamas bankrutavusių įmonių pasiskirstymas pagal tai, keliems metams įmonės duomenys yra randami duomenų rinkinyje. Remiantis šiais duomenimis negalima padaryti išvadų, kad įmonės bankrotas yra susijęs su jos gyvavimo amžiumi, kadangi turimuose duomenyse bankrutavusių įmonių procentas reikšmingai nesiskiria tarp skirtingų finansinių metų skaičiaus.

**14 lentelė.** Bankrutavusių įmonių dalis pagal įrašų skaičių

<b>Skirtingų FY kiekis</b>	<b>Įmonių skaičius</b>	<b>Bankrutavusių įmonių skaičius</b>	<b>Bankrutavusių įmonių dalis (proc.)</b>
1	22666	181	0,80 %
2	18609	179	0,96 %
3	15980	188	1,18 %
4	14215	224	1,58 %
5	13629	207	1,52 %
6	11958	167	1,40 %
7	17623	248	1,41 %
8	43208	214	0,50 %
9	1919	13	0,68 %
10	713	7	0,98 %
11	282	3	1,06 %
12	148	0	0,00 %
13	87	1	1,15 %
14	42	2	4,76 %
15	15	0	0,00 %
16	12	1	8,33 %
17	3	0	0,00 %
18	6	0	0,00 %

### 3. Empirinis Lietuvos įmonių bankroto prognozavimo tyrimas

Šioje dalyje yra pateikiamos vykdytos duomenų tvarkymo procedūros, transformacijos. Čia aprašomi išbandyti klasifikavimo modeliai: reguliarizuota logistinė regresija, atsitiktiniai miškai, bei GBM. Pateikiami šių bankroto prognozavimo modelių rezultatai, bei jų prasmė.

#### 3.1. Duomenų analizė

Toliau nagrinėjamos tokios įmonės, kurios duomenyse turi informacijos mažiausiai dviems finansiniams metams – jų duomenyse yra 138499.

Literatūros apžvalgos metu buvo pastebėta įmonių bankroto koreliacija su ataskaitų teikimo vėlavimu, todėl sukuriamas papildomas stulpelis: „reporting\_delay“ = „document\_registration\_date“ – „finance\_year\_end\_date“, kuris gali padėti indikuoti prastėjančią įmonės padėtį – kai vėluojama pateikti finansines ataskaitas. Šio stulpelio reikšmių pasiskirstymas pateikiamas lentelėje (reikšmės pateikiamos dienomis).

**15 lentelė.** Vėlavimo pateikti ataskaitas (dienomis) pasiskirstymas duomenyse

Minimumas	1 kvartilis	Mediana	Vidurkis	3 kvartilis	Maksimumas
0	124	149	224	179	6123

Analizės metu buvo pastebėta, kad yra tokių įmonių, kurios nėra bankrutavusios, tačiau jų rodikliai primena bankrutavusių įmonių rodiklius.

**16 lentelė.** Įmonių teisinio statuso reikšmės

status_name	Įmonė bankrutavusi	Įmonė nebankrutavusi
Bankrupt	878	0
Cross-border merging of a legal entity	0	19
Going Bankrupt	2462	0
Legal entity partic. in cross-border merging	0	4
Liquidation being initiated	0	4933
Liquidation due to bankruptcy	3853	0
No legal proceedings	0	728842
Partic. in Reorganizatio	0	303
Partic. in division	0	213
Removed	0	32327
Under Liquidation	0	7273
Under Reformation	0	330
Under Reorganization	0	686

Under Restructuring	387	0
---------------------	-----	---

Kadangi lieka nemaža dalis įmonių, kurių statusas yra „Removed“ arba „Under Liquidation“, šių bendrovių finansiniai duomenys gali sukelti triukšmą. Klasifikavimo uždavinio metu daroma prielaida, kad visos įmonės be bankroto indikacijos yra sėkmingai veikiančių įmonių pavyzdžiai, todėl reikia pašalinti visas įmones, kurių šio stulpelio reikšmė nėra lygi „No legal proceedings“. Po šių veiksmų duomenyse lieka 126513 įmonės ir 736422 duomenų eilutės.

Klasifikavimo uždavinyje nėra tikslinga palikti paskutinius – įmonės bankroto metus, kadangi siekis yra prognozuoti vienerius metus į priekį, todėl sukuriamas naujas stulpelis: „is\_bankrupted\_next\_year“ – šis stulpelis įgyja reikšmę lygią vienetui vieneri metai iki bankroto fakto. Po naujo stulpelio sukūrimo iš visų įmonių duomenų pašalinti paskutiniai duomenyse turimi metai ir panaikinamas stulpelis „is\_bankrupted\_last\_year“, kadangi jis netenka prasmės. Duomenyse lieka 609909 eilutės. Šioje vietoje taip pat reikia sutvarkyti duomenyse esančių pagrindinių finansinių rodiklių informaciją – visas šias reikšmes iki 2014 finansinių metų imtinai reikia padalinti iš paskutinio fiksuoto euro ir lito kurso: 3,4528.

Toliau duomenų analizės metu įmonės padalinamos į dvi dalis – įmones kurios turi bankroto indikaciją (stulpelio reikšmė „is\_bankrupted\_all\_years“ lygi vienetui) ir įmones, kurios bankroto indikacijos neturi (stulpelio reikšmė „is\_bankrupted\_all\_years“ lygi nuliui). Iš viso duomenyse yra 1454 bankrutavusios įmonės (6126 duomenų eilutės), iš kurių 779 turi trūkstamų duomenų. Kadangi duomenyse yra didelis klasių disbalansas, tokiu atveju negalima prarasti daugiau nei pusės bankrutavusių įmonių duomenų. Duomenų užpildymas vykdomas trimis etapais, pirmiausia pildomos pagrindinių finansinių rodiklių to paties stulpelio reikšmės pagal turimas istorines reikšmes – jei tokių reikšmių nėra, toliau duomenys pildomi MICE metodu, pasitelkiant atsitiktinius miškus. Po šio etapo pagal formules, esančias 1 priede, naudojantis papildytais pagrindiniais finansiniais rodikliais, yra išskaičiuojami santykiniai finansiniai rodikliai. Visi trūkstamų reikšmių užpildymo veiksmai vykdomi atskyrus bankrutavusias įmones į atskirą duomenų rinkinį, kad MICE metodas sugebėtų atrinkti į kitų bankrutavusių įmonių panašias rodiklių reikšmes. Nebankrutavusioms įmonėms, iš duomenų pašalinamos visos įmonės, kurios turi bent vieną trūkstamą reikšmę – tai leidžia turėti tikslesnius duomenis [24] ir sumažina klasių disbalansą. Po tokių veiksmų lieka 39743 nebankrutavusios įmonės ir 217293 duomenų eilutės. Taip pat, yra 140 bankrutavusių ir 695 nebankrutavusios įmonės, kurių duomenys nėra pateikti nuosekliai kiekvieniems metams. Siekis yra apskaičiuoti kuo tikslesnius skirtumus tarp einamųjų ir praeitų finansinių metų, todėl 695 nebankrutavusios įmonės yra pašalinamos iš duomenų rinkinio, lieka 39048 nebankrutavusios įmonės, bei 213740 duomenų eilučių.

Duomenų tvarkymo etapas baigtas: pašalinti visų įmonių, turėjusių jų tik vieneriems metams, duomenys, taip pat visų nebankrutavusių įmonių duomenys, turėję trūkstamų reikšmių, ar nenuosekliai teikusią ataskaitas. Taip pat, pašalinti duomenys nebankrutavusioms įmonėms, kurios turėjo teisinį statusą kitokį nei „no legal proceedings“. Papildomai duomenims buvo pridėti stulpeliai: „is\_bankrupted\_next\_year“ ir „reporting\_delay“, taip pat, remiantis literatūros apžvalga, pateikta 2 priede, sukurtas viso turto stulpelis (sudėjus trumpalaikį ir ilgalaikį turtą) bei panaikintas stulpelis „is\_bankrupted\_last\_year“ – duomenyse yra 26 stulpeliai ir 219866 eilutės.

Sutvarkius duomenis galima pastebėti, kad naujai sukurtame stulpelyje „reporting\_delay“ matoma koreliacija su įmonės bankroto statusu. Nors mediana yra gana panaši abiejose grupėse, tačiau beveik šimtu dienų didesnė vidurkio reikšmė reiškia didesnį išskirčių kiekį, didesnį vidurkį paaiškina bankrutavusioms įmonėms beveik du kartus didesnė 95 procentilio reikšmė.

**17 lentelė.** Vėlavimo pateikti ataskaitas (dienomis) pasiskirstymo duomenyse palyginimas

	<b>Veikiančios</b>	<b>Bankrutavusios</b>
<b>Minimumas</b>	0	4
<b>5 procentilis</b>	76	85
<b>1 kvartilis</b>	127	137
<b>Mediana</b>	147	152
<b>Vidurkis</b>	183	278
<b>3 kvartilis</b>	159	241
<b>95 procentilis</b>	483	929
<b>Maksimumas</b>	5528	5214

Toliau nagrinėjamas medianos stulpelis žemiau pateiktoje lentelėje, remiantis [7] įvardintomis rodiklių interpretacijomis. Šioje lentelėje antrame stulpelyje matomų trumpinių reikšmės: veikiančios – V, bankrutavusios – B. Stulpelių pavadinimuose taip pat yra naudojami trumpiniai: minimali reikšmė (min.), maksimali reikšmė (maks.), pirmas kvartilis, trečias kvartilis ir mediana, atitinkamai – Q1, Q3, Md. Pirmasis duomenyse esantis rodiklis yra skolos ir nuosavo kapitalo santykis, didesnė jo reikšmė indikuoja didesnį finansinį svertą, kuris didina riziką, tačiau mažesnė už vienetą reikšmė reiškia, kad rizika yra kontroliuojama. Šio rodiklio interpretacija taip pat priklauso nuo to, kokie yra įmonės pelningumo rodikliai – kai jie yra neigiami, tada didesnė reikšmė indikuoja didesnę riziką, o jei įmonės pelningumo rodikliai yra geri – į šiek tiek didesnes rodiklio reikšmes galima žiūrėti atlidžiau. Šiuo atveju medianos stulpelyje tiek veikiančioms, tiek bankrutavusioms įmonėms yra matoma gana panaši reikšmė, artima pusei, o veikiančioms įmonėms reikšmė yra netgi šiek tiek didesnė. Vidurkio stulpelyje galima matyti didesnes rodiklio reikšmes – čia bankrutavusių įmonių rizika yra apie 3 kartus didesnė nei veikiančiųjų – tai lemia grupėje esantis didesnis išskirčių skaičius. Toliau – trumpalaikio turto ir įsipareigojimų santykis, čia matuojamas įmonės sugebėjimas trumpalaikiu turtu padengti visus įsipareigojimus – didesnė reikšmė indikuoja finansinį saugumą, tačiau per didelė reikšmė taip pat gali indikuoti neefektyvų kapitalo valdymą. Matomas didesnis finansinis saugumas veikiančių įmonių grupėje, tačiau atkreipus dėmesį į vidurkį – matomas antros teorijos dalies patvirtinimas apie kapitalo valdymo neefektyvumą – bankrutavusioms įmonėms rodiklio reikšmė yra dvigubai didesnė. Nuosavo kapitalo reikšmė yra apie 4 kartus didesnė veikiančioms nei bankrutavusioms įmonėms – tai indikuoja, kad didesnė įmonės turto dalis yra finansuojama savininkų lėšomis – ši reikšmė tiesiogiai atspindi įmonės finansinį stabilumą, kuris bankrutavusioms įmonėms yra matomas mažesnis. Toliau lentelėje matomas bendrasis skolos rodiklis – jo reikšmė, tiek medianos, tiek vidurkio stulpelyje indikuoja, kad finansinis svertas yra pastebimai didesnis bankrutavusioms įmonėms, nei veikiančioms. Grynojo pelno ir turto santykio palyginime medianos stulpelyje matoma nulinė reikšmė bankrutavusioms įmonėms, bei apie 4 % dydžio reikšmė veikiančioms įmonėms. Bendrai peržvelgus visus stulpelius galima pastebėti dėsningumą, kad veikiančioms įmonėms šio rodiklio reikšmė yra didesnė – rodiklis rodo didesnę galutinę grąžą investuotojui. Toliau matomas turto apyvartumas, kaip teigiama [7],

yra lengviausiai interpretuojamas kai nagrinėjamos to paties sektoriaus įmonės, tačiau šiame duomenų rinkinyje tokios informacijos apie įmones pateikta nėra, todėl būtų sunku nuspėti, ar bankrutavusios įmonės savo sudėtyje turi daugiau vieno tipo įmonių, ar veikiančios turi daugiau kitokio tipo įmonių. Didesnė rodiklio reikšmė įprastai vertinama palankiai, kadangi indikuoja efektyvų išteklių naudojimą. Šiuo atveju matomas didesnis turto apyvartumas veikiančiose įmonėse pirmojo kvartilio bei medianos reikšmėse, bet kaip jau minėta anksčiau, šio rodiklio interpretacija yra gana sudėtinga kai duomenyse skirtingų sektorių įmonės. ROE yra vienas pagrindinių rodiklių, parodančių investuotų lėšų grąžą, sugeneruojamą savininkams ir investuotojams, čia didesnė reikšmė vėlgi matoma nebankrutavusioms įmonėms. Panaši situacija matoma su rodikliu ROA, jo reikšmė, indikuojanti įmonės gebėjimą uždirbti pelną, naudojant turimą turtą, yra didesnė nebankrutavusioms įmonėms – tai reiškia, kad šios įmonės efektyviau įdarbina turimas likvidžias lėšas.

**18 lentelė.** Santykinų finansinių rodiklių reikšmių pasiskirstymo duomenyse palyginimas

Rodiklis	Grupė	Min.	Q1	Vidurkis	Md	Q3	Maks.
idx_debt_to_equity_ratio	V	-21263,6364	0,0733	3,6656	0,4992	1,7048	140818,583
idx_debt_to_equity_ratio	B	-60173	-1,4612	10,4225	0,3878	2,7614	42799
idx_sh_as_lia_ratio	V	-109391	0,6902	11,8655	1,3201	3,0731	85287,4222
idx_sh_as_lia_ratio	B	-906,1053	0,31	22,6048	0,8033	1,266	40001
idx_equity_ratio	V	-64412	0,1618	-2,5891	0,4836	0,7749	23540
idx_equity_ratio	B	-1582404	-0,1449	-528,4306	0,1386	0,4692	5561,1429
idx_lia_as_ratio	V	-23572	0,2192	3,5986	0,5069	0,8339	64413
idx_lia_as_ratio	B	-6392,1429	0,4601	529,4268	0,8187	1,1684	1582405
idx_n_prof_as_ratio	V	-15748,6667	-0,0153	-0,3578	0,0439	0,1652	16273
idx_n_prof_as_ratio	B	-11676,25	-0,1525	-6,072	0	0,0556	5204
idx_s_rev_as_ratio	V	-45224	0,7294	3,6702	1,6092	2,8916	129900
idx_s_rev_as_ratio	B	-22,9784	0,3964	221,5815	1,5581	3,5665	327707,368
idx_prof_tax_eq_ratio	V	-10377	0,0078	0,0576	0,1548	0,4667	1934,5
idx_prof_tax_eq_ratio	B	-1716,3333	-0,0113	0,0135	0,1101	0,6711	1955,4524
idx_prof_tax_as_ratio	V	-15748,6667	-0,0149	-0,3458	0,0491	0,1832	16273
idx_prof_tax_as_ratio	B	-102673,772	-0,1404	-39,2301	0	0,0668	718,4

Toliau duomenys yra praturtinami papildomais stulpeliais, sukuriami stulpeliai su atitinkama pavadinimo galūne: „\_lag1“ stulpelyje yra praeito periodo to rodiklio (pvz: įsipareigojimų) reikšmė, „\_diff“ reiškia, kad yra apskaičiuojamas skirtumas tarp turimos ir „\_lag1“ vertės, o „\_chng“ reiškia, kad yra apskaičiuojamas santykis tarp dabartinės ir „\_lag1“ vertės.

**19 lentelė.** Papildomi finansinių rodiklių stulpeliai

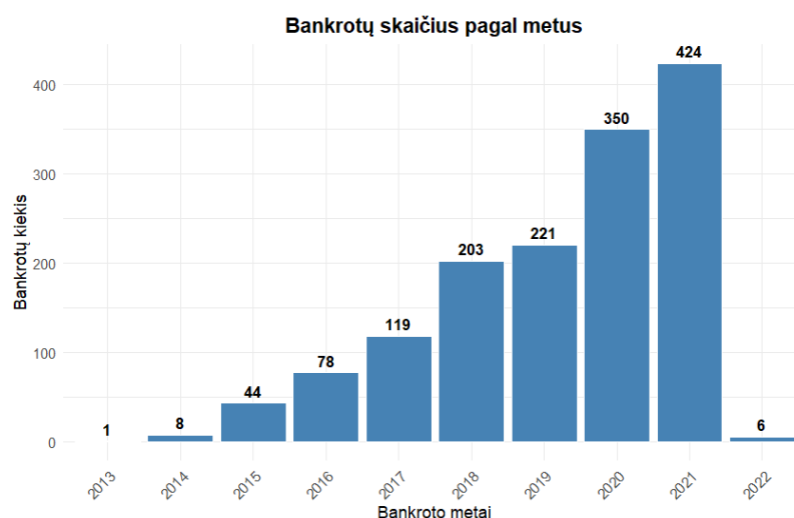
Originalus stulpelis	_lag1	_diff	_chng
A	A_lag1	A - A_lag1	A / A_lag1

Toliau atliekant analizę, galima pastebėti, kad nors patys duomenys yra nuo 2004 metų, tačiau bankrutavusios įmonės duomenų rinkinyje atsirado nuo 2013 metų.

**20 lentelė.** Bankrutavusių įmonių skaičius skirtinguose laiko intervaluose

Periodas	Eilučių skaičius	Unikalios įmonės	Bankrotų skaičius
2004-2012	1010	586	0
2013-2022	218856	40465	1454

Kadangi duomenyse iki 2013 m. nėra nė vienos bankroto indikacijos ir tokių duomenų eilučių yra nedaug, įmonės, kurios neturi tęstinių duomenų 2013m. ir vėliau yra pašalinamos, kad šis laikotarpis nedarytų įtakos modeliui. Siekiant sumažinti makroekonominių rodiklių įtaką iki šios datos, naudojami įmonių duomenis nuo 2013 metų. Duomenyse lieka 1454 bankrutavusios ir 39011 nebankrutavusios įmonės. Analizuojamu laikotarpiu bankrutavusių įmonių dalis yra nedidelė, tačiau nuo 2015 m. iki 2021 m., kasmet augo ir sudarė nuo 0,15 % iki 1,37 % visų duomenyse esančių įmonių.



**5 pav.** Bankrutavusių įmonių skaičius laike

Toliau būtų naudinga turėti indikatorius priklausantį nuo įmonės teisinės formos (stulpelis legal\_form\_name), tačiau 96,2 % įmonių duomenyse yra uždarnosios akcinės bendrovė (UAB). Duomenyse galima rasti 38934 UAB ir tik 6 asociacijas, toks skirtumas neleidžia šio stulpelio naudoti kaip kategorinio kintamojo klasifikavimo uždavinyje.

Paskutinis žingsnis prieš sprendžiant klasifikavimo uždavinį yra padalinti gautus duomenis į mokymo ir testavimo imtis: duomenys nuo 2013 iki 2020 metų patenka į apmokymo imtį (joje yra 1024 bankroto atvejai), o 2021-2022 metų intervalas patenka į testavimo imtį (kurioje yra 430 bankroto atvejų). Kadangi šiuose laiko intervaluose buvo labai skirtingos makroekonominės sąlygos labai svarbu, kad į duomenis, prieš jų padalinimą, taip pat būtų įtraukiami makroekonominiai rodikliai.

Remiantis literatūros apžvalgos dalimi, buvo pasirinkti šie makroekonominiai rodikliai: BVP metinis augimas, EURIBOR metinis vidurkis [41], vartotojų kainų indekso (VKI) metinis augimas ir nedarbo lygio metinis vidurkis [42], čia procentai duomenyse išreiškiami dešimtainėmis trupmenomis.

Makroekonominiai duomenys buvo praturtinti papildomais stulpeliais, sukuriama stulpeliai su atitinkama pavadinimo galūne: „\_lag1“ stulpelyje yra praeito periodo rodiklio (pvz: BVP metinio augimo) reikšmė, „\_diff“ reiškia, kad yra apskaičiuojamas skirtumas tarp turimos ir „\_lag1“ vertės, o „\_ratio“ reiškia, kad yra apskaičiuojamas santykis tarp dabartinės ir „\_lag1“ vertės.

**21 lentelė.** Papildomi makroekonominių rodiklių stulpeliai

Originalus stulpelis	_lag1	_diff	_ratio
A	A_lag1	A - A_lag1	A / A_lag1

Sutvarkyti metiniai makroekonominiai duomenys yra sujungiami su finansiniais įmonių duomenimis pagal įmonių finansinius metus ir padalinami į apmokymo ir testavimo imtis. Padalinus duomenis atliekamas paskutinis žingsnis: ištrinami unikalūs identifikavimo stulpeliai, arba tokie stulpeliai, kurie gali indikuoti bankroto faktą: ("code", "registration\_date", "legal\_form\_name", "status\_name", "finance\_year\_start\_date", "finance\_year\_end\_date", "document\_registration\_date", "is\_bankrupted\_all\_years"). Duomenų rinkinyje lieka 81 nepriklausomas kintamasis ir vienas priklausomas kintamasis „is\_bankrupted\_next\_year“.

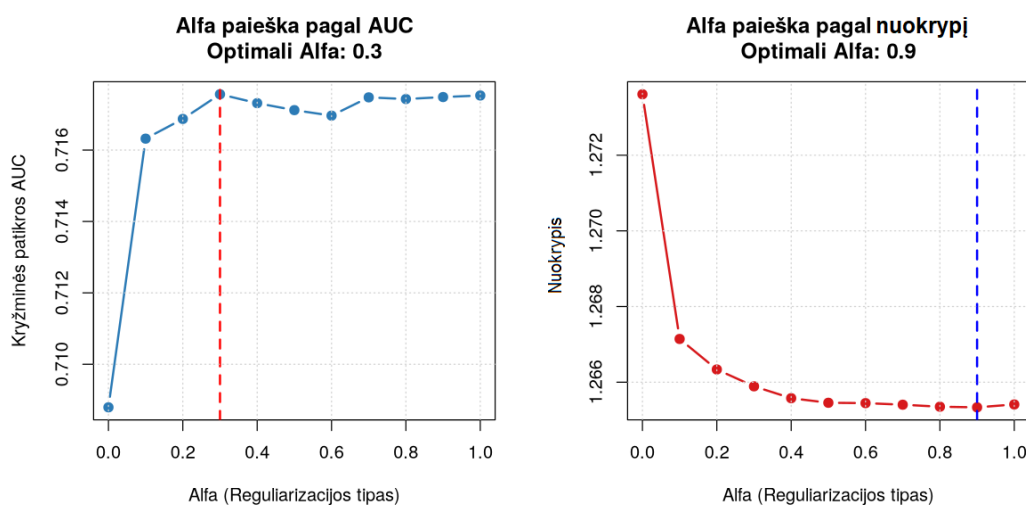
### 3.2. Rezultatai

Tyrimo metu buvo išbandyti 3 klasifikavimo metodai: reguliarizuota logistinė regresija, atsitiktiniai miškai, bei gradientinio stiprinimo metodai. Tyrimo priklausomas kintamasis yra stulpelis „is\_bankrupted\_next\_year“, jis sukurtas tokiu būdu, kad indikuotų įmonės bankrotą po vienerių metų, o įmonės bankroto metai iš duomenų yra ištrinti – tokiu būdu konstruojami modeliai leidžia prognozuoti įmonių bankrotą vienerius metus į priekį. Nepriklausomų kintamųjų duomenyse yra 81 – juos sudaro tiek finansiniai, tiek makroekonominiai rodikliai. Visi skaičiavimai atlikti naudojant R programavimo kalbą.

#### 3.2.1. Logistinė regresija

Pirmiausia šiame tyrime buvo pritaikyta logistinė regresija, kuri palyginta su logistinės regresijos rezultatais, siekiant parodyti reguliarizacijos svarbą. Vienas pirmųjų žingsnių, kuriant reguliarizuotas logistinės regresijos modelį, buvo nustatyti optimaliausia alfa parametras, kuris parodo balansą tarp LASSO ir ridge reguliarizacijų. Optimalios alfa

reikšmės buvo ieškoma vertinant dvi metrikas: plotą po ROC kreive (AUC), bei modelio nuokrypį. Remiantis kryžminės patikros rezultatais nustatyta, kad didžiausia AUC reikšmė yra pasiekama, kai alfa yra lygi 0,3, o mažiausias modelio nuokrypis yra fiksuojamas kai alfa reikšmė yra 0,9.



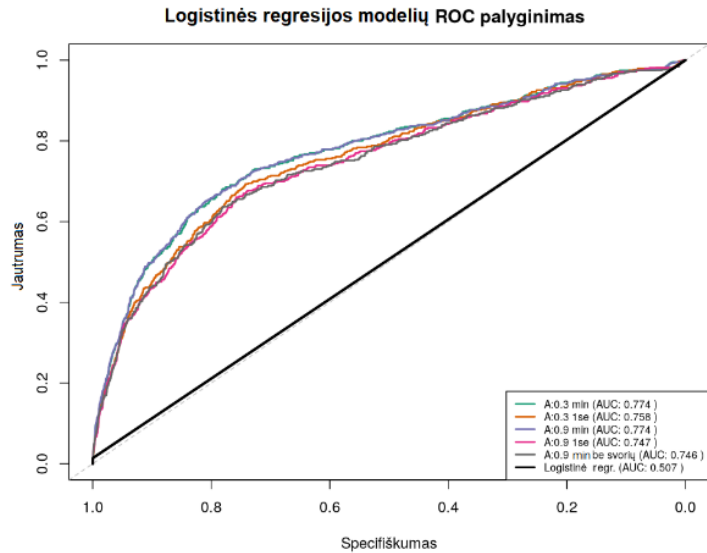
**6 pav.** Optimalios alfa reikšmės paieškos grafikas

Buvo atliktas modelių apmokymas su gautais alfa parametrais bei skirtingomis lambda parametro reikšmėmis (minimalia *min* ir remiantis standartinio nuokrypio taisykle *1se*), gauti rezultatai pateikiami lentelėje.

**22 lentelė.** Logistinės regresijos modelių parametrai

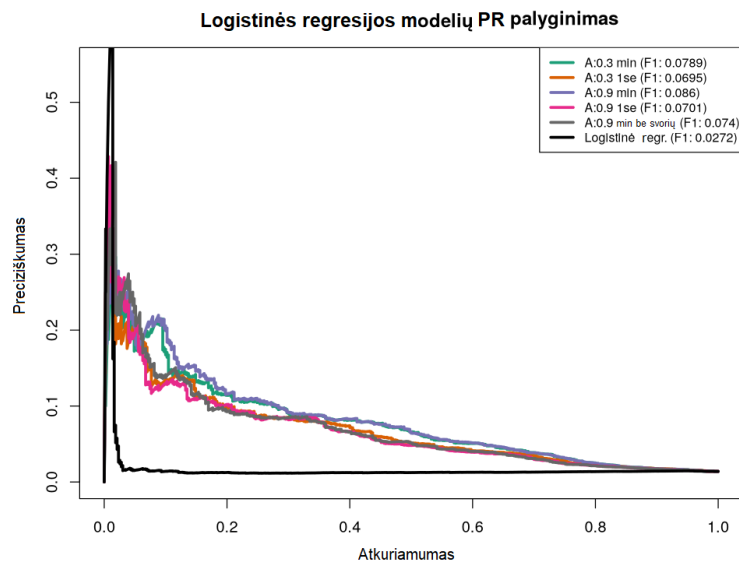
Modelis	AUC	Prec.	Sens.	F1	Slenkstis	Kintamieji	Alfa	Lambda
A:0,3 (min)	0,7742	0,0419	0,6814	0,0789	0,7746	66	0,3	min
A:0,3 (1se)	0,7578	0,0366	0,693	0,0695	0,7686	34	0,3	1se
A:0,9 (min)	0,7742	0,046	0,6512	0,086	0,8248	60	0,9	min
A:0,9 (1se)	0,7468	0,037	0,6605	0,0701	0,7678	20	0,9	1se
A:0,9 (min) be svorių	0,7464	0,0393	0,6372	0,074	0,0247	46	0,9	min
Logistinė regresija	0,5069	0,5455	0,014	0,0272	0,5	81	-	-

ROC grafike galima matyti, kad visos reguliarizuotos logistinės regresijos kreivės yra išsidėsčiusios gerokai aukščiau už paprastos logistinės regresijos kreivę, kuri beveik susilieja su įstrižaine, o tai vizualiai patvirtina reguliarizacijos efektyvumą. Kreivės atitinkančius modelius su lambda parametru *min* ir svoriais yra labiausiai nutolusios į kairį viršutinį kampą – tai patvirtina didžiausias jų AUC reikšmes matomas lentelėje (0,7742).



**7 pav.** Logistinės regresijos modelių ROC palyginimas

PR grafikas rodo, kad preciziškumas sparčiai krenta didėjant atkuriamumui. Tai yra būdinga duomenų rinkiniams su dideliu klasių disbalansu – norint identifikuoti daugiau bankrotų, negalima išvengti didėjančio klaidingų bankroto nustatymo atvejų skaičiaus. Plotas po reguliarizuotos logistinės regresijos modelių kreivėmis yra žymiai didesnis nei standartinės logistinės regresijos (juoda kreivė). Tai parodo, kad reguliarizuoti modeliai tam pačiam atkuriamumo lygiui išlaiko didesnę tikslumą.



**8 pav.** Logistinės regresijos modelių PR palyginimas

Remiantis gautais rezultatais galima teigti, kad logistinės regresijos su reguliarizacija tikslumas (AUC > 0,74) stipriai viršija paprastos logistinės regresijos rezultatus, kurios AUC reikšmė yra nežymiai didesnė už 0,5 – atsitiktinio spėjimo reikšmę. Galima pastebėti, kad modeliai su lambda *min* reikšme pasiekia didesnę AUC reikšmę, tačiau naudoja didesnę kintamųjų kiekį (60 – 66). Naudojant vieno standartinio nuokrypio taisyklę, kintamųjų skaičius sumažėja atitinkamai tris arba du kartus, tačiau prognozės tikslumas lieka panašus

(AUC = 0,74 – 0,75). Taip pat buvo išbandytas modelis netaikant svorių, naudojamų klasių disbalanso problemai spręsti. Čia galima matyti itin žemą slenksčio reikšmę (0,0247), kuri rodo, kad į modelį neįtraukus svorių, jis neefektyviai atskiria klases.

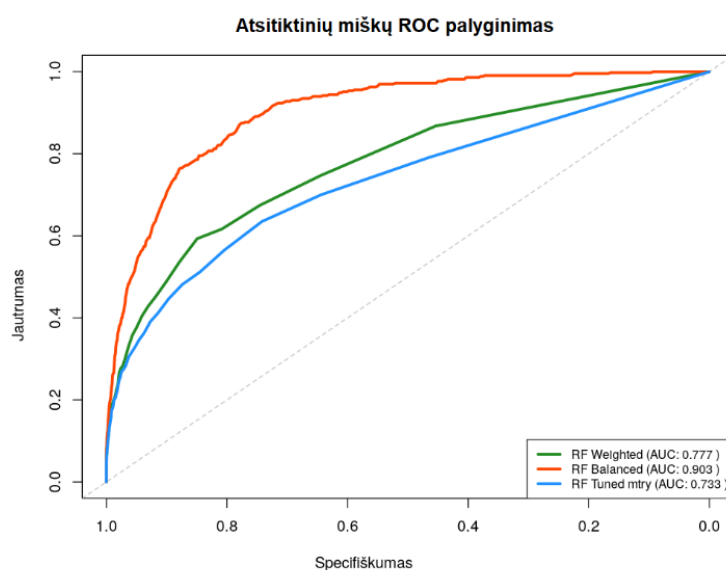
### 3.2.2. Atsitiktiniai miškai

Tolesniame tyrimo etape bankrotui prognozuoti buvo panaudotas atsitiktinių miškų klasifikavimo metodas. Jis pasirinktas norint įvertinti, ar šis metodas gali pagerinti klasifikavimo rezultatus, lyginant su anksčiau nagrinėta reguliarizuota logistine regresija. Buvo sudaryti trys skirtingi atsitiktinių miškų modelių variantai, kurių rezultatai pateikiami lentelėje.

23 lentelė. Atsitiktinių miškų modelių parametrai

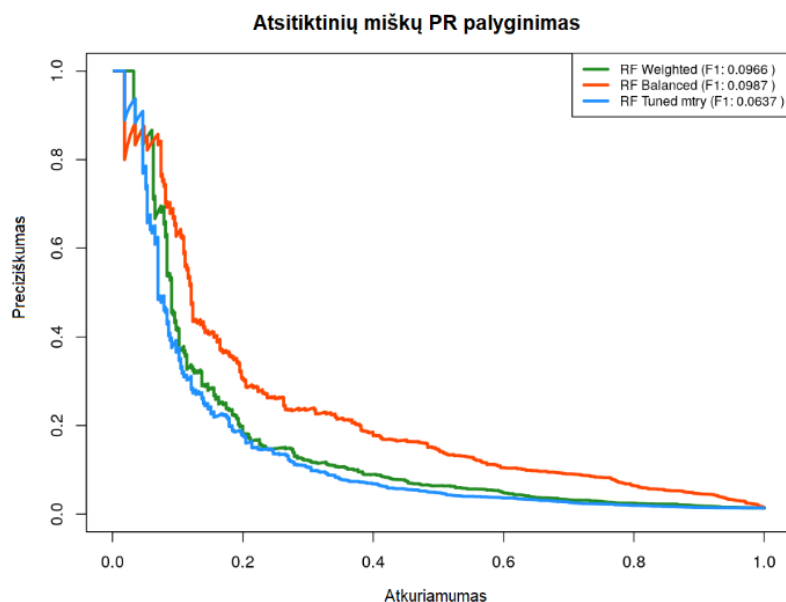
Modelis	AUC	Prec.	Sens.	F1	Slenkstis	Kintamieji	Medžiai	Svoriai
RF (weighted)	0,7773	0,0526	0,593	0,0966	0,009	81	500	ratio_weight
RF (balanced)	0,9027	0,0523	0,8744	0,0987	0,469	81	500	n_bankrupt
RF (tuned mtry)	0,7326	0,0335	0,6349	0,0637	0,005	81	500	ratio_weight

ROC grafike raudona kreivė reprezentuoja subalansuoto atsitiktinių miškų modelio pranašumą prieš kitus atsitiktinių miškų modelius, kadangi jis yra arčiausiai viršutinio kairiojo grafiko kampo. Čia matoma aukščiausia pasiekta AUC reikšmė (0,903), kuri indikuoja gerą modelio gebėjimą atskirti bankrutuojančias įmones. Lyginant kitus atsitiktinių miškų modelius, galima matyti, kad *RF weighted* pasiekia ganėtinai aukštą AUC (0,777), o kintamųjų didinimas viename medyje (*RF tuned mtry* - mėlyna linija) pademonstravo silpniausią rezultatą (0,7326).



9 pav. Atsitiktinių miškų modelių ROC palyginimas

PR grafiko kreivės yra ypatingai svarbios, kadangi duomenų rinkinys pasižymi stipriu klasių disbalansu. Subalansuotas atsitiktinių miškų modelis (raudona spalva) šiame grafike taip pat yra matomas kaip aiškus lyderis. Tokia kreivės pozicija taip pat koreliuoja su aukščiausia F1 rodiklio reikšme (0,0987).



**10 pav.** Atsitiktinių miškų modelių PR palyginimas

Geriausius rezultatus pademonstravo subalansuotas atsitiktinių miškų modelis *RF (balanced)*. Jo AUC reikšmė pasiekė 0,9027, o modelio specifiškumas lygus 0,8744. Tai yra reikšmingas skirtumas lyginant šį modelį tiek su kitais atsitiktinių miškų modeliais, tiek su reguliarizuota logistine regresija. Šio modelio sėkmė slypi svorio kintamajame, galima matyti, kad kiekvienas miškas buvo formuojamas iš subalansuoto duomenų poaibio, taip buvo užtikrinamas subalansuotas modelių apmokymas abiejų klasių požymiams. Taip pat galima atkreipti dėmesį į slenksčio reikšmę, ji kitų dviejų atsitiktinių miškų modelių atveju yra itin žema. Tuo tarpu subalansuotas atsitiktinių miškų modelis pasiekė slenksčių lygų 0,469, kuris yra artimas teoriniam 0,5 vidurkiui. Tai parodo, kad modelis yra stabilesnis ir jo prognozės remiasi ne tik prisitaikymu prie duomenyse dominuojančios klasės, bet yra realiai pagrįstos bankroto tikimybe. Modelis *RF (tuned mtry)*, kuriame buvo padidintas vienu metu vertinamų kintamųjų skaičius, pasirodė prasčiausiai (AUC = 0,7326). Tai gali reikšti, kad leidžiant modeliui rinktis iš per didelio kintamųjų skaičiaus kiekviename išsišakojime, medžiai tampa per daug panašūs vienas į kitą, arba padidėja „triukšmas“, mažinantis klasifikacijos modelio tikslumą.

### 3.2.3. Gradientinio stiprinimo metodai

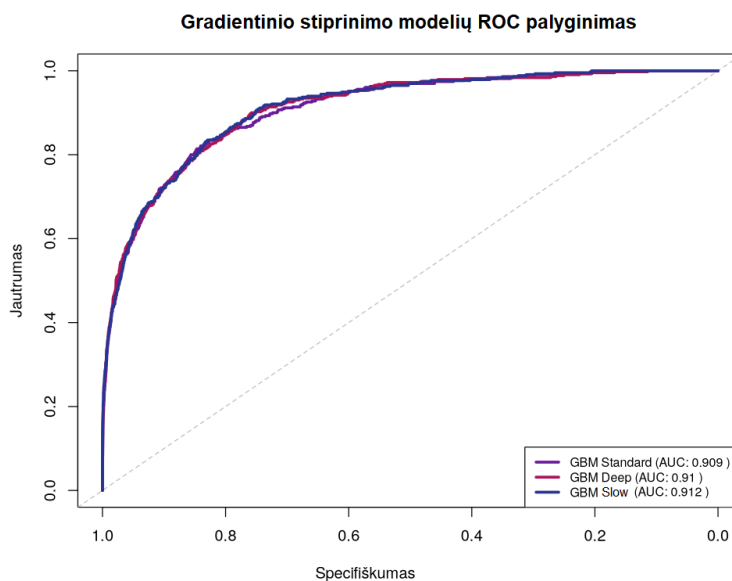
Paskutiniame tyrimo etape bankrotui prognozuoti buvo pritaikytas gradientinio stiprinimo metodas. Buvo išbandytos trys GBM variacijos: standartinė, gilioji ir lėtoji. Analizuojant lentelės duomenis galima matyti, kad visi GBM modeliai pasiekė didesnę nei 0,9 rodiklio AUC, bei didesnę nei 0,09 F1 rodiklio reikšmę. Lėtasis GBM modelis pasižymi didžiausiu plotu po kreive (0,9115), tačiau praktiniu požiūriu gilusis GBM modelis yra pranašesnis, kadangi pasižymi aukščiausiu atkuriamumo rodikliu (0,9). Jis teisingai klasifikuoja 387

bankroto atvejus, praleidžia 43 – tai yra mažiausias praleistų bankrotų (FN) skaičius tarp visų modelių.

**24 lentelė.** GBM modelių parametrai

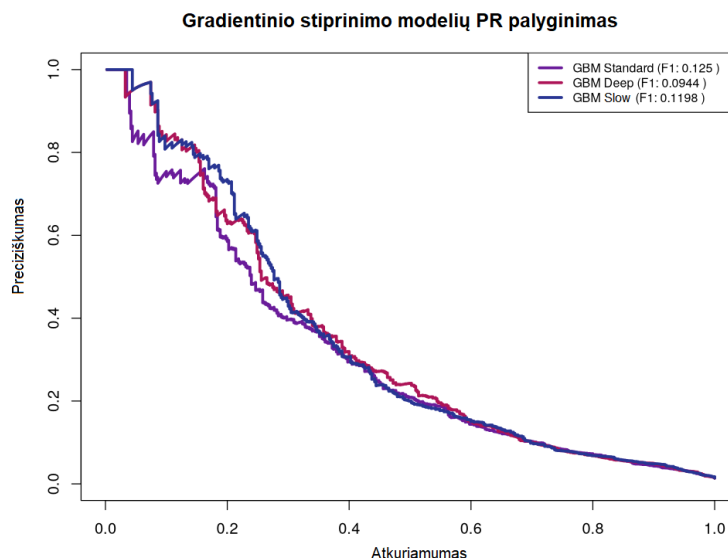
Modelis	AUC	Prec.	Sens.	F1	Slenkstis	Kintamieji	medžiai	gylis
GBM (slow)	0,9115	0,0645	0,8349	0,1198	0,5559	81	200	4
GBM (deep)	0,9104	0,0498	0,9	0,0944	0,4061	81	100	8
GBM (standard)	0,9085	0,0677	0,8209	0,125	0,5601	81	100	4

Vizualiai modelių klasifikavimo gebėjimai yra matomi ROC palyginimo grafike. Visų trijų modelių kreivės beveik persidengia bei yra arti viršutiniojo kairiojo grafiko kampo, todėl reiškia aukštą klasifikavimo stabilumą. Nors skaitiniai AUC skirtumai tarp modelių nėra dideli, lėtojo modelio (mėlyna linija) pranašumas ties vidutinėmis specifiškumo reikšmėmis rodo, kad lėtesnis mokymasis (*shrinkage* = 0.05) padeda modeliui geriau atpažinti dėsningumus.



**11 pav.** GBM modelių ROC palyginimas

PR grafikas geriau atskleidžia modelių veikimą, kai duomenyse yra didelis klasių disbalansas. Galima matyti, kad standartinis GBM modelis pasiekia šiek tiek didesnę preciziškumą, esant mažoms atkuriamumo reikšmėms, tai atsispindi rodiklio F1 reikšmėje (0,125). Tačiau giliojo GBM modelio kreivė išlieka stabili, net ir siekiant didesnio atkuriamumo. Tai patvirtina modelio gebėjimą identifikuoti bankroto požymius tada, kai kiti modeliai šioms įmonėms priskiria nebankrutavusių įmonių klasę. Tai parodo, kad sudėtingesnės kintamųjų sąveikos leidžia modeliui sudaryti tikslesnius klasifikavimo dėsningumus.

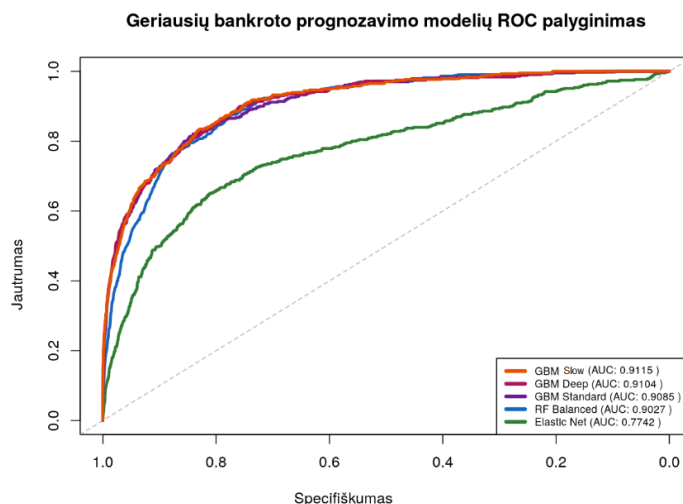


12 pav. GBM modelių PR palyginimas

Apibendrinant GBM modelių rezultatus, galima teigti, kad gradientinio stiprinimo algoritmai yra šio tyrimo lyderiai, remiantis pateiktomis rezultatų matavimo metrikomis. GBM geba efektyviai išnaudoti kintamuosius, nustatydamas sudėtingus jų tarpusavio ryšius. Nors standartinis modelis pasižymi didžiausiu balansu (aukščiausia F1 reikšme), tačiau gilusis modelis parodo mažiausią antro tipo (FN) klaidų, bankroto prognozavimo kontekste turinčių didesnę įtaką, tikimybę.

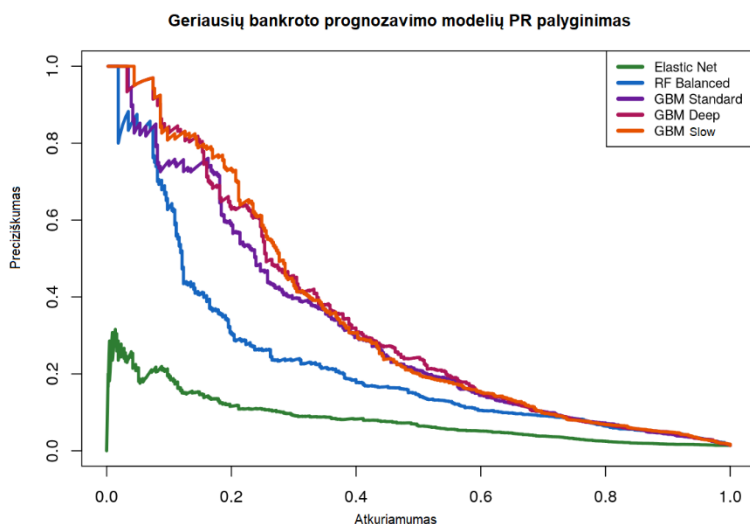
### 3.2.4. Modelių rezultatų palyginimas

Paskutiniame žingsnyje tikslas yra išrinkti modelį, gebantį geriausiai klasifikuoti įmones į tas, kurios per vienerius metus bankrutuos, arba į tas, kurios šiuo laikotarpiu nebankrutuos. Žemiau pateiktame grafike galima matyti ROC kreives šiems modeliams: GBM gilusis, subalansuotas atsitiktinių miškų modelis, bei reguliarizuotos logistinės regresijos modelis, turėjęs didžiausias AUC ir F1 reikšmes. Akivaizdu, kad medžių ansambliniai metodai (AUC = 0,91) turi pranašumą prieš reguliarizuotą logistinę regresiją (AUC = 0,77).



13 pav. Geriausių bankroto prognozavimo modelių ROC palyginimas

Preciziškumo – atkuriamumo (PR) kreivių palyginimas patvirtina, kad GBM gilusis (Atkuriamumas = 0,9) ir atsitiktinių miškų subalansuotas (Atkuriamumas = 0,8744) modeliai geba išlaikyti aukštą modelio specifiškumą, siekiant kuo didesnio bankrotų atpažinimo lygio. Galima pastebėti, kad logistinė regresija šiame grafike demonstruoja silpniausius rezultatus, tai rodo, kad logistinė regresija, net ir naudojant regularizaciją, nepakankamai efektyviai išskiria bankrutuojančių įmonių dalį iš bendros populiacijos.



**14 pav.** Geriausių bankroto prognozavimo modelių PR palyginimas

Atlikus išsamią trijų skirtingų metodų analizę, galima daryti išvadą, kad geriausius bankroto prognozavimo rezultatus vienerių metų horizonte Lietuvos įmonių duomenų imtyje demonstruoja gilusis GBM modelis. Nors standartinis GBM modelis pasiekė šiek tiek aukštesnį F1 rodiklį, tačiau gilusis modelis pasirenkamas kaip optimalus sprendimas dėl jo gebėjimo minimizuoti antro tipo klaidų (FN) kiekį.

**25 lentelė.** Geriausių klasifikavimo modelių rezultatai

Modelis	RF (balanced)	GBM (slow)	GBM (deep)	GBM (standard)
TP	376	359	387	353
TN	23682	25293	23116	25633
FP	6816	5205	7382	4865
FN	54	71	<b>43</b>	77
Sens.	0,8744	0,8349	0,9	0,8209
Spec.	0,7765	0,8293	0,7580	0,8405
Prec.	0,0523	0,0645	0,0498	0,0677
F1	0,0987	0,1198	0,0944	0,125

Atsižvelgiant į tai, kad praktikoje nepastebėtas bankrotas (FN) gali sukelti didesnius finansinius nuostolius nei klaidingas veikiančios įmonės priskyrimas bankrutuojančiųjų

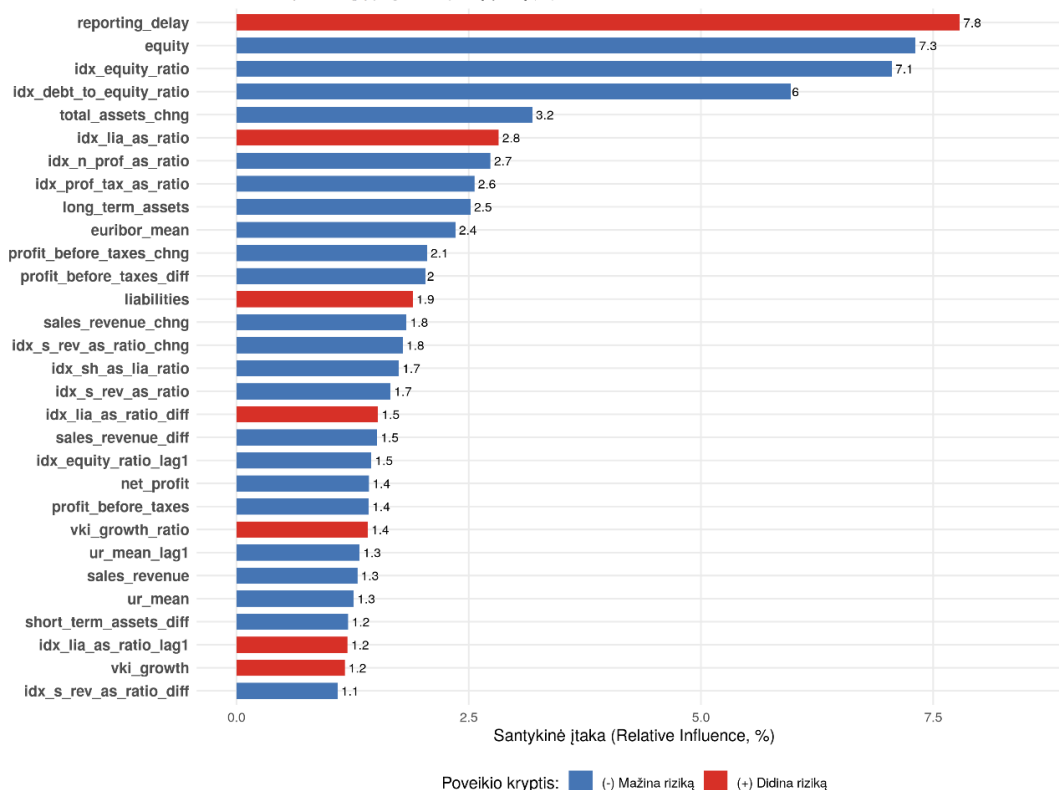
įmonių klasei [36], giliojo GBM modelis užtikrinamas atkuriamumas (0,9) yra didelis privalumas. Tyrimas patvirtina, kad ansambliniai medžių metodai suteikia esminį pranašumą prognozuojant įmonių bankroto tikimybę vienerių metų horizonte.

### 3.2.5. Ekonominė rezultatų prasmė

Modelio kintamųjų svarbos analizei buvo pasitelktas kombinuotas metodas – pirmiausia kintamųjų reikšmingumas buvo nustatytas naudojant vidinį *gbm* algoritmą, o jų poveikio kryptis įvertinta skaičiuojant *Spearmano* ranginės koreliacijos koeficientus. Interpretuojant giliojo GBM modelio rezultatus, žemiau pateiktame grafike raudona spalva reiškia didesnės absoliučios rodiklio reikšmės daromą teigiamą įtaką didesnei bankroto tikimybei, o mėlyna spalva reiškia didesnės absoliučios rodiklio reikšmės daromą neigiamą įtaką didesnei bankroto tikimybei. Grafike galima pastebėti, kad stipriausią įtaką rezultatams turi naujai sukurtas stulpelis „reporting\_delay“, indikuojantis įmonės vėlavimą pateikti finansines ataskaitas – kuo ilgiau įmonė nepateikia ataskaitų – tuo didesnė tikimybė įmonei bankrotuoti. Antrą – ketvirtą vietas svarbumo skalėje užima rodikliai susiję su įmonės nuosavu kapitalu: absoliuti nuosavo kapitalo reikšmė, nuosavo kapitalo koeficientas, bei skolos ir nuosavo kapitalo santykis – didesnės šių rodiklių reikšmės modelyje reiškia mažesnę tikimybę bendrovę priskirti prie bankrutavusių įmonių klasės. Didesnės nuosavo kapitalo, bei nuosavo kapitalo koeficiento reikšmės reiškia didesnę įmonės stabilumą, todėl tokios modelio padarytos išvados yra logiškos. Remiantis modelio rezultatais, vien tik didesnė skolos ir nuosavo kapitalo santykio reikšmė, indikuojanti didesnę riziką, nereiškia bankroto, o kaip tik, koreliuodama su kitais finansiniais rodikliais (gerais pelningumo rodikliais) rodo mažesnę bankroto tikimybę. Kiti penki grafike matomi rodikliai yra susiję su įmonės turtu: didesnė santykinio šio ir praeito periodo viso turto rodiklio reikšmė reiškia mažesnę tikimybę įmonei bankrotuoti per artimiausius metus, taip pat mažesnę tikimybę bankrotuoti rodo didesnė investicijų grąža, didesnis turto apyvartumas, bei didesnė ilgalaikio turto reikšmė. Didėjanti bendrojo skolos rodiklio reikšmė – priešingai – didina įmonės bankroto tikimybę. Šio rodiklio ne viena modifikacija, tiek vėluojanti reikšmė, tiek skirtumas tarp dviejų periodų pasirodo vėliau grafike su šiek tiek mažesne įtaka modelio rezultatams. Šie su įmonės turtu susiję rodikliai modelio rezultatus veikia teisingai, vadovaujantis teorija. Toliau grafike matomi du – iš pagrindinių finansinių rodiklių apskaičiuoti papildomi rodikliai – dabartinio laikotarpio pelno iki mokesčių ir prieš tai buvusios šio rodiklio reikšmės santykis ir skirtumas. Jei per paskutinius metus didėjo tiek absoliuti pelno iki mokesčių, tiek santykinė reikšmė, tai modelyje reiškia mažėjančią bankroto tikimybę – nors 18 lentelėje galima matyti išskirčių, tačiau modelis teisingai užfiksuoja rodiklio įtaką, remiantis teorija. Toliau galima matyti pagrindinius finansinius rodiklius: didėjantys įmonės įsipareigojimai indikuoja didesnę bankroto riziką, o didesni grynojo pelno, pelno iki mokesčių, bei pardavimo pajamų rodikliai bankroto tikimybę mažina. Didesnės veiklos apimtys rodiklių – procentinio, bei absoliutaus pardavimo pajamų procentinio pokyčio reikšmės mažina bankroto riziką, įdomu pastebėti, kad procentinis pokytis užima svarbesnę vietą nei absoliutus. Turto apyvartumo rodiklio reikšmė, jo pokytis tarp periodų (tiek atimant, tiek dalinant reikšmes) yra įtrauktas į svarbiausių 30 rodiklių grafiką.

### Svarbiausi bankroto indikatoriai (Deep GBM modelis)

30 svarbiausių kintamųjų pagal santykinę įtaką (%)



15 pav. Giliojo GBM modelio 30 svarbiausių indikatorių

Svarbu įvertinti makroekonominių rodiklių įtaką modeliui, pirmasis, daugiausiai įtakos turintis rodiklis – EURIBOR metinis vidurkis. Kintamojo „euribor\_mean“ neigiamą įtaką galima paaikškinti specifine 2013-2022 m. laikotarpio dinamika. Junttila ir Tauriainen [43] pastebėjo, kad palūkanų normų poveikis finansiniams rodikliams stipriai priklauso nuo makroekonominės aplinkos. 2015-2021 m. periodu buvo fiksuotos neigiamos EURIBOR reikšmės, šiuo laikotarpiu ekonominė veikla pasižymėjo žemesniu aktyvumu. Kadangi modelis buvo apmokytas 2013-2020 metų periodo duomenimis, o prognozė vykdoma 2021-2022 metų periodui, remiantis [22] įžvalgomis, galima teigti, kad mažesnė tarpbankinės rinkos palūkanų norma taip pat lemia mažesnius bankų taikomus skolinimo standartus. Grafike yra matoma teigiama koreliacija tarp mažesnės EURIBOR reikšmės, bei tikimybės įmonei bankrutuoti per artimiausius metus. Modelis, mokydamasis iš 2013-2020 m. duomenų, tokią tendenciją priskyrė didesnei rizikai, o 2022 m. įvykęs staigus šuolis buvo interpretuotas kaip grįžimas į sveiką ekonomikos būseną. Taip pat, verta pabrėžti, kad 2022 metų bankroto prognozė remiasi neigiamomis EURIBOR 2021 metų reikšmės ir 2022 m. pabaigoje stebimas palūkanų normos šuolis, dar neatsispindi tų metų bankrotų statistikoje. Įdomu atkreipti dėmesį į ekonomikos ciklą indikuojančius rodiklius – VKI ir nedarbo lygis. Kaip minėta literatūrinėje dalyje, nedarbo lygio rodiklis yra vėluojantis, todėl logiška matyti vėluojančią šio rodiklio reikšmę, kuri yra papildoma dabartine rodiklio reikšme, taip pat yra matoma VKI rodiklio augimo pagreičio, bei paties augimo reikšmė. Tokios vėluojančios arba santykinės rodiklių reikšmės reiškia, kad jos yra matuojamos du metai iki įmonės bankroto fakto. Todėl, pavyzdžiui, jei bankroto faktas yra 2020, šios įmonės bankrotui prognozuoti yra naudojami 2019 m. finansiniai rodikliai, bei makroekonominių rodiklių augimo reikšmės tarp 2018 ir 2019 metų („vki\_growth\_ratio“ ir „vki\_growth“), bei metinės rodiklio vidurkio reikšmės

(„ur\_mean\_lag1“, „ur\_mean“). Remiantis Liu [44], makroekonominiai rodikliai modeliuose, skirtuose prognozuoti bankrotui, yra sisteminės rizikos indikatoriai. Čia iškeliami hipotezė apie infliacijos sukeltą įmonių sąnaudų padidėjimą, lemiantį didesnę bankroto riziką, yra pagrindžiama vėlesniame tyrime. Jame Hol [45] parodo VKI augimo įtaką didėjančiai įmonių bankroto rizikai, tokia įtaka yra matoma modelio, kuris buvo apmokytas laikotarpyje, stebint nuolat augančią VKI reikšmę, rezultatuose. Didžiojoje nagrinėjamo periodo dalyje yra stebimas vidutinis arba didelis nedarbo lygis, 2013 m. siekiantis netgi 11 %. Modelis rodo didesnio nedarbo lygio neigiamą koreliaciją su didesne bankroto tikimybe ateityje. Toks rezultatas yra pagrindžiamas Hol [45] įžvalga, jog nedarbo lygis yra vėluojantis rodiklis, o interpretacija tiesiogiai priklauso nuo nagrinėjamo verslo ciklo etapo. 2013-2020 m. Lietuvoje ekonomikos augimo fazėje nedarbo lygis mažėjo [42], o modelis mažą nedarbo lygį sugretino su didėjančia riziką. Tokią koreliaciją galima paaiškinti padidėjusiu darbo rinkos perkaitimu ir atlyginimų augimu, šie veiksniai tiesiogiai veikė įmonių patiriamus kaštus. Tokie makroekonominiai rodikliai kaip VKI ir nedarbo lygis, nagrinėjamu laikotarpiu sudarytame modelyje, indikuoja įmonių sąnaudų didėjimą – brangstančią darbo jėgą ir kylančias žaliavų kainas. Remiantis 26 lentele ir 15 paveikslu, galima pamatyti, kad 2021 m. bankroto prognozė remiasi neigiama 2020 metų EURIBOR reikšme, taip pat 2019 ir 2020 m. matomu didėjančiu nedarbo lygiu, bei 2019-2020 m. augančia VKI reikšme.

**26 lentelė.** Makroekonominių rodiklių reikšmės modelio testavimo periode [42]

	2018 m.	2019 m.	2020 m.	2021 m.
euribor_mean	- 0,18 %	- 0,22 %	- 0,3 %	- 0,5 %
vki_growth_ratio	0,7	0,9	0,5	3,9
ur_mean_lag1	7,1 %	6,19 %	6,28 %	8,53 %
ur_mean	6,19 %	6,28 %	8,53 %	7,14 %
vki_growth	2,7 %	2,33 %	1,20 %	4,7 %

Analizuojant svarbiausius bankroto indikatorius išryškėja kelios pagrindinės tendencijos. Atskaitomybės vėlavimas yra matomas kaip kritinis rizikos signalas, indikuojantis vidines įmonės problemas. Antra – didelę įtaką daro įmonės nuosavas kapitalas, svarbus tiek absoliutus jo dydis, tiek jo santykis su turtu, kuris reiškia tiek didesnę įmonės stabilumą, finansuojant turtą nuosavu kapitalu, tiek ir didesnę įmonės savininkų įsitraukimą į įmonės veiklą. Trečias – svarbūs yra bankroto riziką didinantys rodikliai, tokie kaip bendras įsipareigojimų dydis, įsipareigojimų ir turto santykis, jo vėluojanti reikšmė, bei šio rodiklio dviejų laikotarpių skirtumas. Taip pat galima matyti, kad gana svarbią vietą užima makroekonominiai rodikliai – toks rezultatas leidžia daryti išvadą, kad įmonių bankrotui prognozuoti yra svarbūs ne tik vidiniai įmonės rodikliai, bet ir makroekonominė aplinka.

### 3.3. Apibendrinimas ir rekomendacijos

Apibendrinant rezultatus, galima teigti, kad makroekonominių rodiklių integracija į bankroto prognozavimo modelius, suteikia jiems analitinį pranašumą, vertinant sisteminę riziką. Tradiciniai modeliai, tokie kaip standartinė logistinė regresija yra neefektyvūs, kai duomenyse yra pastebimas didelis klasių disbalansas. Sprendimų medžių ansamblių

klasifikavimo metodai, tokie kaip atsitiktiniai miškai ir GBM, reikšmingai pranoksta logistinės regresijos modelius prognozuojant įmonių bankrotą vieneriems metams į priekį. Giliojo gradientinio stiprinimo modelio rezultatai parodė aukščiausią atkuriamumo rodiklio reikšmę (0,9), tačiau nuo jo nežymiai atsilieka subalansuotas atsitiktinių miškų modelis (0,87). Šiame atsitiktinių miškų modelyje medžiams formuoti buvo panaudotas poaibių išlyginimas pagal bankrutavusių įmonių skaičių.

Šiame tyrime viena pagrindinių problemų buvo klasių disbalansas, kadangi norint kokybiškai apmokyti mašininio mokymosi metodus geriausia yra turėti panašų kiekį abiejų klasių duomenų. Problema buvo sprendžiama dviem skirtingais būdais: subalansuotas atsitiktinių miškų modelis pasižymi didesniu stabilumu, kadangi kiekvienas medis buvo formuojamas iš subalansuoto poaibio, kitiems modeliams buvo naudojamas bendras klasių santykio metodas. Kadangi modelio apmokymui naudoti realūs 2013-2020 m. Lietuvos įmonių bei makroekonominiai rodikliai, algoritmai sugebėjo identifikuoti specifiškai šiam verslo ciklo etapui būdingas rizikas. Čia ekonomikos atsigavimo laikotarpio metu yra fiksuojami išaugusi rizika dėl darbo rinkos perkaitimo, bei didėjančių žaliavų kainų.

Tęsiant tyrimą, galima būtų plačiau ištirti duomenų disbalanso problemą, ir nustatyti kokius balansavimo metodus geriausia taikyti. Šio darbo metu išbandyti du metodai – abiejų klasių narių skaičiaus suvienodinimas kiekviename sprendimo medyje, bei didesnės svarbos priskyrimas mažesniajai klasei. Taip pat, galima būtų testuoti suformuotus mašininio mokymosi modelius, įtraukiant vėlesnį ekonomikos ciklo laikotarpį (2023-2025 m.). Tyrimo metu modeliai buvo apmokyti neigiamos tarpbankinės palūkanų normos laikotarpiu, buvo stebimas nedarbo lygio mažėjimas ir VKI augimas, o prognozė vykdyta 2021-2022 m. ekonominių šokų fone. Modelio prognozavimo stabilumą galima būtų įvertinti, išbandant jį naujesnei duomenų imčiai, apimančiai periodą po EURIBOR šuolio. Tai leistų įvertinti, kaip modelis prisitaiko prie kardinaliai pasikeitusių kredito rinkos sąlygų.

## Išvados

1. Remiantis mokslinės literatūros analize, įmonių bankrotas yra apibrėžiamas kaip kompleksinė ekonominė ir vadybinė problema. Ji sukelia neigiamas pasekmes įmonės savininkams, darbuotojams, kreditoriams, bei daro įtaką šalies finansiniam stabilumui. Literatūroje bankrotas yra suprantamas ne tik kaip teisinis procesas, bet ir kaip vidinių vadybinių klaidų, bei nepalankių išorinių veiksnių visuma. Atlikta teorinė apžvalga atskleidė, kad įmonių bankrotą aiškinančios teorijos, prasidėjusios tradicinėmis finansinių rodiklių analizėmis, dabar yra papildomos daugiafunkciais modeliais. Nustatyta, kad šiuolaikinių mokslinių darbų praktikoje, tradiciniai statistiniai metodai (tokie kaip logistinė regresija ar diskriminantinė analizė) yra keičiami mašininio mokymosi algoritmais, gebančiais identifikuoti netiesinius ryšius. Mokslinių straipsnių analizė leido išskirti esminius bankrotui prognozuoti naudojamus rodiklius. Tyrimuose dažniausiai naudojami finansiniai rodikliai apima įmonių pelningumo (ROA, ROE), kapitalo struktūros, likvidumo ir veiklos efektyvumo kintamuosius. Taip pat, siekiant įvertinti sisteminę verslo ciklo įtaką, literatūroje identifikuojama makroekonominių veiksnių (BVP augimo, VKI, nedarbo lygio) įtraukimo į modelius svarba;
2. suformuotas galutinis tyrimo duomenų rinkinys, apimantis realius 2013-2022 m. veikiančių, bei bankrutavusių Lietuvos įmonių finansinius, bei makroekonominius rodiklius. Kadangi duomenyse iki 2013 m. duomenų rinkinyje bankroto faktų nebuvo fiksuota, periodas iki šių metų buvo pašalintas. Galutinį analizuojamą rinkinį sudarė 1454 bankrutavusios ir 39011 nebankrutavusių įmonių. Duomenys padalinti, remiantis finansiniais metais: 2013-2020 m. laikotarpis (1024 bankrotai) skirtas modeliui apmokyti, o 2021-2022 m. intervalas (430 bankrotų) – testuoti;
3. atliktas duomenų paruošimas, suvaldant trūkstamas reikšmes, išskirtis, bei neinformatyvius kategorinius kintamuosius (teisinės formos stulpelis pašalintas dėl UAB formos dominavimo duomenyse – 96,2 %). Siekiant užtikrinti galimybę stebėti vėluojančią rodiklių įtaką, makroekonominiai duomenys buvo praturtinti išvestiniais vėlavimo („\_lag1“), skirtumo („\_diff“) ir santykio („\_ratio“) kintamaisiais. Galutinį duomenų rinkinį sudarė 81 nepriklausomas kintamasis ir vienas priklausomas kintamasis („is\_bankrupted\_next\_year“), leidžiantis prognozuoti bankrotą vienerius metus į priekį;
4. įvertinus sudarytų modelių tikslumą ir patikimumą testavimo imtyje, nustatyta, kad standartinės logistinės regresijos modelio rezultatai beveik nesiskiria nuo atsitiktinio spėjimo (AUC = 0,5069). Reguliarizuotos logistinės regresijos taikymas leido pastebimai pagerinti prognozavimo kokybę (AUC = 0,7742). Vis dėlto, ansambliniai medžių metodai pademonstravo pranašumą prieš logistinės regresijos modelius, buvo pasiekti AUC įverčiai viršijantys 0,9. Šių modelių formavimo etape buvo suvaldyta duomenyse esanti klasių disbalanso problema;
5. palyginus modelių rezultatus, tiksliausiu bankroto prognozavimo modeliu Lietuvos įmonių duomenims vienerių metų horizonte, išrinktas gilusis GBM modelis. Šis modelis pasiekė aukštą ploto po ROC kreive reikšmę (AUC = 0,9104), bei aukščiausią atkuriamumo rodiklį (Sens. = 0,9). Tokie rezultatai leido modeliui teisingai identifikuoti 387 bankroto atvejus iš 430. Čia matomas mažiausias antrojo tipo (FN) klaidų, kurios praktikoje gali sukelti didžiausius finansinius nuostolius, kiekis. Ekonominė kintamųjų svarbos analizė parodė, kad geriausiai bankroto riziką indikuoja vėlavimas pateikti ataskaitas, nuosavo kapitalo rodikliai, taip pat svarbus yra makroekonominis kontekstas. Sudarytą modelį galima naudoti, kaip papildomą

įrankį vertinant verslų rizikingumą, verslo partnerių patikimumą, bei stebint investicinę aplinką.

## Literatūros sąrašas

1. PEPE, M.S. *The Statistical Evaluation of Medical Tests for Classification and Prediction - EBSCO* [interaktyvus]. . [s.l.]: Oxford Statistical Science Series, 2004. .
2. NSAIF, A.A. - ABD, D.H. Political Post Classification based on Firefly and XG Boost. In *Journal of Algebraic Statistics* . 2022. Vol. 13, no. 3, p. 4079–4101. .
3. [s.l.]: Audito, apskaitos, turto vertinimo ir nemokumo valdymo tarnyba, 2016. .
4. [s.l.]: Audito, apskaitos, turto vertinimo ir nemokumo valdymo tarnyba, 2016. .
5. PARKER, R.H. *Macmillan Dictionary of Accounting* . 1992. .
6. MARTINKUS, B. ir kt. *Aiškinamasis įmonės vadybos terminų žodynas* . 2000. .
7. [s.l.]: Corporate Finance Institute, 2020. .
8. BĀRBUŤĀ-MIŠU, N. - MADALENO, M. Assessment of Bankruptcy Risk of Large Companies: European Countries Evolution Analysis. In *Journal of Risk and Financial Management* . 2020. Vol. 13, no. 3, p. 58. .
9. ONAKOYA, A.B. - OLOTU, A.E. Bankruptcy and Insolvency: An Exploration of Relevant Theories. In *International Journal of Economics and Financial Issues* . 2017. Vol. 7, no. 3, p. 706–712. .
10. HART, O. [interaktyvus]. .[s.l.]: National Bureau of Economic Research, 2000. DOI: 10.3386/w7921. Prieiga per internetą: <<https://www.nber.org/papers/w7921>>.
11. GIRIŪNIENE, G. ir kt. A Comparison on Leading Methodologies for Bankruptcy Prediction: The Case of the Construction Sector in Lithuania. In *Economies* . 2019. Vol. 7, no. 3, p. 82. .
12. KANAPICKIENĖ, R. ir kt. Bankruptcy Prediction for Micro and Small Enterprises Using Financial, Non-Financial, Business Sector and Macroeconomic Variables: The Case of the Lithuanian Construction Sector. - EBSCO. In [interaktyvus]. 2023. Prieiga per internetą: <<https://research-ebSCO-com.ezproxy.ktu.edu/c/7a7sn3/viewer/html/bvqle67icn>>.
13. KRISTÓF, T. - VIRÁG, M. A Comprehensive Review of Corporate Bankruptcy Prediction in Hungary. In *Journal of Risk and Financial Management* . 2020. Vol. 13, no. 2, p. 35. .
14. BAUER, P. - ENDRÉSZ, M. Modelling bankruptcy using Hungarian firm-level data. In *MNB Occasional Papers* . 2016. no. 122, p. 1–36. .
15. SIGRIST, F. - LEUENBERGER, N. Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. In *European Journal of Operational Research* . 2023. Vol. 305, no. 3, p. 1390–1406. .

16. NEHREBECKA, N. PREDICTING THE DEFAULT RISK OF COMPANIES. COMPARISON OF CREDIT SCORING MODELS: LOGIT VS SUPPORT VECTOR MACHINES. In *ECONOMETRICS* . 2018. Vol. 22, no. 2, p. 54–73. .
17. DEVI, P.K. - HEMAVATHY, Dr.R. Assessing Bankruptcy Vulnerability In Manufacturing Firm Using Altman's Z'-Score. In *International Journal of Research Publication and Reviews* . 2025. Vol. 6, no. 4, p. 13082–13086. .
18. PRUSAK, B.Sud. *Corporate Bankruptcy Prediction: International Trends and Local Experience*. . Basel, Switzerland: MDPI - Multidisciplinary Digital Publishing Institute, 2020. 1 p. ISBN 978-3-03928-911-0.
19. ESPINOZA, R. ir kt. The Role of Financial Variables in predicting economic activity. In *Journal of Forecasting* . 2012. Vol. 31, no. 1, p. 15–46. .
20. KUFENKO, V. - GEIGER, N. Business cycles in the economy and in economics: an econometric analysis. In *Scientometrics* . 2016. Vol. 107, no. 1, p. 43–69. .
21. BORMOTOV, M. Economic cycles: historical evidence, classification and explication. In [interaktyvus]. 2009. Prieiga per internetą: <<https://mpira.ub.uni-muenchen.de/19660/>>.
22. BOTO-GARCÍA, D. ir kt. The effect of Euribor on banking profitability: Evidence from the Spanish banking system. In *European Journal of Government and Economics (EJGE)* . 2021. Vol. 10, no. 1, p. 5–29. .
23. ONOFREI, M. - LUPU, D. THE MODELING OF FORECASTING THE BANKRUPTCY RISK IN ROMANIA. In *Economic Computation & Economic Cybernetics Studies & Research* . 2014. Vol. 48, no. 3, p. 249–267. .
24. BRÉDART, X. - CORREA-MEJÍA, D.A. Bankruptcy prediction modeling in real-world conditions: A contrast of boosting algorithm and logistic regression. - EBSCO. In [interaktyvus]. 2022. Prieiga per internetą: <<https://research-ebSCO-com.ezproxy.ktu.edu/c/7a7sn3/viewer/html/jtpsx7mvlv>>.
25. FREUND, R.J. - WILSON, W.J. Statistical Methods - EBSCO. In [interaktyvus]. 2003. [žiūrėta 2026-01-10]. Prieiga per internetą: <[https://research-ebSCO-com.ezproxy.ktu.edu/c/7a7sn3/ebook-viewer/pdf/fe6umojia5/page/pp\\_540?location=https%25253A%25252F%25252Fresearch-ebSCO-com.ezproxy.ktu.edu%25252Fc%25252F7a7sn3%25252Fsearch%25252Fdetails%25252Ffe6umojia5%25253Fdb%25253De000xww](https://research-ebSCO-com.ezproxy.ktu.edu/c/7a7sn3/ebook-viewer/pdf/fe6umojia5/page/pp_540?location=https%25253A%25252F%25252Fresearch-ebSCO-com.ezproxy.ktu.edu%25252Fc%25252F7a7sn3%25252Fsearch%25252Fdetails%25252Ffe6umojia5%25253Fdb%25253De000xww)>.
26. PARK, H. - KONISHI, S. Robust logistic regression modelling via the elastic net-type regularization and tuning parameter selection. In *Journal of Statistical Computation and Simulation* . 2016. Vol. 86, no. 7, p. 1450–1461. .

27. ALGAMAL, Z.Y. - LEE, M.H. High Dimensional Logistic Regression Model using Adjusted Elastic Net Penalty. In *Pakistan Journal of Statistics & Operation Research* . 2015. Vol. 11, no. 4, p. 667–676. .
28. HASTIE, T. ir kt. An Introduction to glmnet. In . .
29. LÓPEZ, O.A.M. ir kt. Multivariate Statistical Machine Learning Methods for Genomic Prediction - EBSCO. In *EBSCO* [interaktyvus]. 2022. Prieiga per internetą: <<https://research-ebSCO-com.ezproxy.ktu.edu/c/7a7sn3/ebook-viewer/pdf/ajlwobpsan?location=https%25253A%25252F%25252Fresearch-ebSCO-com.ezproxy.ktu.edu%25252Fc%25252F7a7sn3%25252Fsearch%25252Fdetails%25252Fajlwobpsan%25253Fdb%25253De001mww>>.
30. AU, T.C. Random Forests, Decision Trees, and Categorical Predictors: The “Absent Levels” Problem. In . .
31. BREIMAN, L. ir kt. [interaktyvus]. .2002. Institution: Comprehensive R Archive Network. Prieiga per internetą: <<https://CRAN.R-project.org/package=randomForest>>.
32. BIAU, G. - CADRE, B. Optimization by Gradient Boosting. In DAOUIA, A. - RUIZ-GAZEN, A.Sud. *Advances in Contemporary Statistics and Econometrics* [interaktyvus]. Cham: Springer International Publishing, 2021. p. 23–44. [žiūrėta 2026-05-21]. ISBN 978-3-030-73248-6Prieiga per internetą: <[https://link.springer.com/10.1007/978-3-030-73249-3\\_2](https://link.springer.com/10.1007/978-3-030-73249-3_2)>.
33. GREEN, D.M. - SWETS, J.A. *Signal Detection Theory and Psychophysics*. . [s.l.]: Wiley, 1966. .
34. COOK, J. - RAMADAS, V. When to consult precision-recall curves. In *The Stata Journal* . 2020. Vol. 20, no. 1, p. 131–148. .
35. LUQUE, A. ir kt. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. In *Pattern Recognition* . 2019. Vol. 91, p. 216–231. .
36. LOMBARDO, G. ir kt. Natural Language Processing and Deep Learning for Bankruptcy Prediction: An End-to-End Architecture. In *IEEE Access* . 2024. Vol. 12, p. 151075–151091. .
37. ALI ABD AL-HAMEED, K. Spearman’s correlation coefficient in statistical analysis. In *International Journal of Nonlinear Analysis and Applications* . 2022. Vol. 13, no. 1, p. 3249–3255. .
38. OSMAN, M.S. ir kt. A Survey on Data Imputation Techniques: Water Distribution System as a Use Case. In *IEEE Access* . 2018. Vol. 6, p. 63279–63291. .
39. Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study | American Journal of Epidemiology | Oxford

- Academic. In [interaktyvus]. Prieiga per internetą: <<https://academic.oup.com/aje/article/179/6/764/107562?guestAccessKey=>>.
40. GERON, A. [interaktyvus]. .2026. Prieiga per internetą: <<https://github.com/ageron/handson-ml2>>.
41. FM.M.U2.EUR.RT.MM.EURIBOR1YD\_.HSTA | ECB Data Portal. In *ECB Data Portal* [interaktyvus]. Prieiga per internetą: <<https://data.ecb.europa.eu/>>.
42. Pradžia - Oficialiosios statistikos portalas. In *Oficialiosios statistikos portalas* [interaktyvus]. Prieiga per internetą: <<https://osp.stat.gov.lt/>>.
43. TAURIAINEN, J.-M. - JUNTTILA, J.-P. Commercial Paper Market as a Forward-Looking Component for EURIBOR Interest Rates During the COVID-19 Pandemic. In *SSRN Electronic Journal* [interaktyvus]. 2021. Prieiga per internetą: <<https://www.ssrn.com/abstract=3979728>>.
44. LIU, J. Macroeconomic determinants of corporate failures: evidence from the UK. In *Applied Economics* . 2004. Vol. 36, no. 9, p. 939–945. .
45. HOL, S. The influence of the business cycle on bankruptcy probability. In *International Transactions in Operational Research* . 2007. Vol. 14, no. 1, p. 75–90. .

## Priedai

### 1 priedas

#### Santykiniai finansiniai rodikliai [7]

Pavadinimas	Formulė	Paiškinimas	Reikšmė
Skolos ir nuosavo kapitalo santykis	$\frac{\text{Įsipareigojimai}}{\text{Nuosavas kapitalas}}$	Finansinio sveto rodiklis, padedantis įvertinti įmonės kapitalo struktūrą lyginant visus įsipareigojimus su akcininkų nuosavu kapitalu. Šis santykis parodo, kurią įmonės veiklos finansavimo dalį sudaro skolintos lėšos ir kurią – savininkų investicijos. Santykio reikšmė parodo vienam kapitalo eurui tenkantį įsipareigojimų kiekį	Didesnė rodiklio reikšmė indikuoja didesnę finansinį svertą, kuris didina riziką, todėl toks finansavimo modelis yra palankesnis stabilius pinigų srautus turinčioms bendrovėms
Trumpalaikio turto ir visų įsipareigojimų santykis	$\frac{\text{Trumpalaikis turtas}}{\text{Įsipareigojimai}}$	Rodiklis parodo įmonės gebėjimą turimais pinigais ir gautinomis sumomis iš klientų, padengti visus turimus įsipareigojimus	Jei rodiklio reikšmė yra didesnė, tai indikuoja ypatingą finansinį saugumą, tačiau aukštas lygis taip pat gali indikuoti neefektyvų kapitalo valdymą
Nuosavo kapitalo koeficientas	$\frac{\text{Nuosavas kapitalas}}{\text{Turtas}}$	Finansinio sveto rodiklis, rodantis nuosavo kapitalo ir įmonės turto proporciją	Reiškia kokia dalis įmonės turto yra finansuojama įmonės savininkų lėšomis, o tai tiesiogiai atspindi įmonės finansinį stabilumą
Bendrasis skolos rodiklis	$\frac{\text{Įsipareigojimai}}{\text{Turtas}}$	Finansinio sveto rodiklis, kuris parodo kokia įmonės turto dalis yra finansuojama skolintomis lėšomis, o jo interpretavimas tiesiogiai atspindi įmonės finansinę riziką: kuo šis rodiklis aukštesnis, ypač jei pastebimas jo nuoseklus augimas, tuo didesnis įmonės svetas ir potencialus nemokumo pavojus	Rodiklio reikšmė lygi vienetui indikuoja, kad įmonės įsipareigojimai yra lygūs turtui – tai rodo aukštą finansinį svertą, reikšmė didesnė už vienetą reiškia, kad įmonės skolos viršija turimą turtą, todėl veikla laikoma rizikinga
Investicijų grąža	$\frac{\text{Grynasis pelnas}}{\text{Turtas}}$	Rodiklis matuoja galutinę grąžą investuotojui	Jo reikšmė tiesiogiai priklauso nuo mokestinės sistemos, todėl labiausiai tinka tarpusavyje lyginant tos pačios valstybės įmonės
Turto apyvartumo rodiklis	$\frac{\text{Pardavimo pajamos}}{\text{Turtas}}$	Rodiklis matuoja įmonės veiklos efektyvumą naudojant turimą turtą generuojant pardavimo pajamas. Jis parodo kokia pardavimų suma tenka	Aukštesnė rodiklio reikšmė paprastai vertinama palankiai, nes indikuoja efektyvų išteklių naudojimą, tuo tarpu žemas rodiklis gali

		kiekvienam įmonės valdomo turto eurui, o skaičiuojant dažniausiai naudojamas vidutinis metinis turtas arba turtas laikotarpio pabaigoje	signalizuoti neefektyvų išteklių naudojimą, vėluojančius atsiskaitymus ar neefektyvų atsargų valdymą. Šis rodiklis yra naudingiausias lyginant to paties sektoriaus įmones, kadangi teigiamai vertinamas apyvartumo lygis tiesiogiai priklauso nuo sektoriaus specifikos
ROE	$\frac{\text{Pelnas prieš mokesčius}}{\text{Nuosavas kapitalas}}$	ROE yra vienas svarbiausių pelningumo rodiklių, matuojantis įmonės metinį grynąjį pelną, tenkantį akcininkų nuosavo kapitalo vertei, paprastai rezultatas išreiškiamas procentais. Šis rodiklis parodo investuotų lėšų grąžą, kurią įmonė sugeneruoja savininkams bei investuotojams. Nuosavo kapitalo grąža yra laikoma esminiu rodikliu vertinant investicinį patrauklumą, kadangi sujungia tris svarbius aspektus: turto naudojimo efektyvumą, bendrą pelningumą ir finansinį svertą	Lyginant įmonės su pramonės šakos vidurkiu, galima nustatyti jos konkurencingumą.
ROA	$\frac{\text{Pelnas prieš mokesčius}}{\text{Turtas}}$	Rodiklis matuoja įmonės gebėjimą uždirbti pelną naudojant turimą turtą	Aukštesnė rodiklio reikšmė rodo, kad įmonė efektyviai įdarbina savo likvidžias lėšas

## Straipsniuose naudojami metodai ir rodikliai

Straipsnis	Naudojami finansiniai rodikliai	Naudojami ne finansiniai rodikliai	Naudojami makroekonominiai rodikliai	Metodai
Brédart & Correa-Mejía (2022). <i>Boosting algorithms for bankruptcy prediction</i>	Pelningumo rodikliai: (angl. <i>earnings before interest and tax, depreciation, amortization</i> ) EBITDA, ROA, ROE, grynoji pridėtinė vertė. Likvidumo rodikliai: trumpalaikio turto ir trumpalaikių įsipareigojimų santykis, (angl. <i>free cash flow</i> ) FCF. Įsipareigojimai: bendras skolos rodiklis, finansinis svertas, trumpalaikių ir visų įsipareigojimų santykis	Klasifikavimo kintamieji, balansavimo indikatoriai		Logistinė regresija, „XGBoost“, duomenų balansavimas
Bauer & Endrész (2016). <i>Modelling bankruptcy using Hungarian firm-level data</i>	Pardavimų augimas, ROA, finansinis svertas (skolos ir nuosavo kapitalo santykis), likvidumas, logaritmuota bendrojo turto reikšmė	Logaritmuota darbuotojų skaičiaus reikšmė, įmonės amžius, eksporto požymis, užsienio kapitalo požymis	BVP augimas, užsienio šalių BVP augimas, kreditavimo augimas rinkoje, šalies palūkanų norma	„Probit“ modelis, paneliniai duomenys
Giriūnienė, Giriūnas, Morkūnas, Brucaitė (2019). <i>A Comparison on Leading Methodologies for Bankruptcy Prediction: The Case of the Construction Sector in Lithuania</i>	Veiklos pelno marža, bendrojo pelno ir viso turto santykis, ROA, gautinų sumų ir visų įsipareigojimų santykis, skubaus likvidumo koeficientas, bendrasis skolos rodiklis, trumpalaikio turto dalis visame turte, turto apyvartumas, pardavimų ir įsipareigojimų santykis, pardavimo savikainos ir pardavimų santykis	Įmonės amžius, įrašų skaičius juridinių asmenų registre, vėlavimas pateikti finansines ataskaitas	Specifiniai statybų sektoriaus rodikliai, BVP indeksas, suderintas vartotojų kainų indeksas, vidutinė metinė infliacija, būsto kainų indeksas	Daugiamatė diskriminan tinė analizė (angl. <i>multivariate discriminant analysis</i> ) (MDA), logistinė regresija, klasikinių modelių palyginimas
Kanapickienė, Kanapickas, Nečiūnas (2023)	Likvidumo rodikliai: Apyvartinio kapitalo ir turto	Įmonės amžius ir dydis, įrašų skaičius juridinių	Specifiniai statybų sektoriaus rodikliai, BVP, būsto kainų	Logistinė regresija,

<p><i>Bankruptcy Prediction for Micro and Small Enterprises Using Financial, Non-Financial, Business Sector and Macroeconomic Variables: The Case of the Lithuanian Construction Sector</i></p>	<p>santykis, trumpalaikio turto ir trumpalaikių įsipareigojimų santykis, skubaus likvidumo koeficientas, grynujų pinigų ir trumpalaikių įsipareigojimų santykis. Pelningumo rodikliai: pelno ir turto santykis, ROA, grynoji pelno marža, palūkanų padengimo koeficientas. Kapitalo struktūros: nuosavo kapitalo rinkos vertės ir skolos santykis, nuosavo kapitalo buhalterinės vertės ir skolos santykis, bendrasis skolos rodiklis. Veiklos efektyvumo rodikliai: turto apyvartumas, apyvartinio kapitalo apyvartumas, nuosavo kapitalo apyvartumas, ilgalaikio turto apyvartumas. Pinigų srautų rodikliai: pinigų srautų ir visos skolos santykis, pinigų srautų ir trumpalaikių skolų santykis, pinigų srautų ir turto santykis. Mokumo rodikliai: įsiskolinimo rodiklis</p>	<p>asmenų registre, vėlavimas pateikti finansines ataskaitas, indikatorius, ar įmonė turi daugiau nei vieną akcininką</p>	<p>indeksas, vartotojų kainų indeksas, nedarbo lygis</p>	<p>integruotas modelis</p>
<p>Onofrei, Lupu (2014) <i>The Modelling of Forecasting the Bankruptcy Risk in Romania</i></p>	<p>Grynoji pelno marža, ROA, ROE, pelnas vienam darbuotojui, bendrasis likvidumas, skubaus likvidumo koeficientas, skolos ir nuosavo kapitalo santykis, atsargų apyvartumas, gautinų sumų apyvartumas, turto apyvartumas, turto kiekis, tenkantis</p>	<p>Įmonės dydis</p>	<p>-</p>	<p>MDA, logistinė regresija</p>

	vienam darbuotojui, veiklos pajamos vienam darbuotojui			
Sigrist & Leuenberger (2023). <i>Machine learning for corporate default prediction</i>	<p>Pelningumo rodikliai: ROA, ROE, (angl. <i>earnings before interest and tax</i>) EBIT, grynoji pelno marža. Likvidumo ir apyvartinio kapitalo: trumpalaikio likvidumo koeficientas, apyvartinio kapitalo ir viso turto santykis, pinigų ir turto santykis. Mokumo: bendrasis skolos rodiklis, nepaskirstyto pelno ir viso turto santykis, finansinis svertas (skolos ir nuosavo kapitalo santykis). Veiklos efektyvumo: turto apyvartumas, darbo našumas, atsiskaitymas su tiekėjais (laikotarpis dienomis)</p>	Įmonės amžius, sektoriaus požymis, ryšiai su bankais, eksporto požymis, laiko dimensijos kintamieji, latentinis rizikos veiksnys	BVP augimas, vietinės ir užsienio palūkanų normos, kreditavimo augimas rinkoje, infliacijos lygis, būsto kainų indeksas	GBM, atsitiktiniai miškai, neuroniniai tinklai, <i>hazard</i> modeliai
Bärbutė-Mišu & Madaleno (2020). <i>Assessment of Bankruptcy Risk of Large Companies</i>	<p>Bankroto tikimybės indeksai (angl. <i>Z-scores</i>), apyvartinio kapitalo ir turto santykis, nepaskirstyto pelno ir turto santykis, nuosavo kapitalo ir visų skolų buhalterinės vertės santykis, ilgalaikis kapitalas padalintas iš visų skolų, veiklos efektyvumo rodikliai: EBIT padalinta iš turto, turto apyvartumas, personalų išlaidų dalis pridėtinėje vertėje, finansinių išlaidų svoris pajamose. Likvidumo rodiklis (skolos ir nuosavo kapitalo santykis),</p>	-	-	<i>Altman, Zmijewski, Springate, Tafler, Conan-Holder</i> modeliai, pagrindinių komponentų analizė (angl. <i>principal component analysis</i> ) PCA

	grynųjų pinigų ciklo rodiklis, EBIT			
Devi & Hemavathy (2025). <i>Assessing Bankruptcy Vulnerability Using Altman's Z'-Score</i>	Grynojo apyvartinio kapitalo ir turto santykis, nepaskirstytojo pelno ir turto santykis, ROA, nuosavo kapitalo rinkos vertės ir įsipareigojimų santykis (finansinis svertas), pardavimo pajamų ir turto santykis (turto apyvartumas)	-	-	Altmano Z' modelis
Nehrebecka (2018). <i>Logit vs SVM</i>	ROA, investicijų grąža (angl. <i>return on capital employed</i> ) ROCE, EBIT, Apmokėjimo laikotarpis, mokumo koeficientas	Statistiškai atrinkti kintamieji	Rodiklis, nusakantis įmonės sektoriaus rizikingumą	Logistinė regresija, SVM
Prusak (2020). <i>Corporate Bankruptcy Prediction: International Trends and Local Experience</i>	Visas turtas, ilgalaikis turtas, nematerialusis turtas, trumpalaikis turtas, atsargos, finansinis turtas, trumpalaikės gautinos sumos, visi įsipareigojimai, nuosavas kapitalas, rezervų fondai, ankstesnių metų pelnas/nuostolis			Literatūros apžvalga; MDA, logistinė regresija, SVM, neuroniniai tinklai, PCA
Kristóf & Virág (2020). <i>A Comprehensive Review of Corporate Bankruptcy Prediction in Hungary</i>	Finansiniai rodikliai	Ne finansiniai kintamieji	Makroekonominiai rodikliai	Literatūros apžvalga; MDA, logistinė regresija, neuroniniai tinklai, SVM, sprendimų medžiai

## Geriausių modelių kintamųjų svarba visai duomenų imčiai

Rodiklis	GBM_Deep	GBM_Standard	GBM_Slow	Poveikis
reporting_delay	7,785336172	9,868280946	8,923203578	(+) Didina riziką
equity	7,309327724	7,609630074	7,020580982	(-) Mažina riziką
idx_equity_ratio	7,054679825	10,63014019	10,66220071	(-) Mažina riziką
idx_debt_to_equity_ratio	5,964204715	3,760666288	5,668294018	(-) Mažina riziką
total_assets_chng	3,184543576	3,411811103	2,622453213	(-) Mažina riziką
idx_lia_as_ratio	2,822087831	1,647893683	1,759172559	(+) Didina riziką
idx_n_prof_as_ratio	2,733520985	2,190694848	1,19566919	(-) Mažina riziką
idx_prof_tax_as_ratio	2,561582642	2,889165774	4,53783482	(-) Mažina riziką
long_term_assets	2,518569115	2,925790136	2,841050779	(-) Mažina riziką
euribor_mean	2,356104639	3,0797763	3,339741187	(-) Mažina riziką
profit_before_taxes_chng	2,050416623	2,965402949	2,319048828	(-) Mažina riziką
profit_before_taxes_diff	2,03640486	1,896734851	1,800844102	(-) Mažina riziką
liabilities	1,898039575	2,682585177	1,844086454	(+) Didina riziką
sales_revenue_chng	1,831557631	1,021621919	1,613794678	(-) Mažina riziką
idx_s_rev_as_ratio_chng	1,790018656	1,715492276	1,916891197	(-) Mažina riziką
idx_sh_as_lia_ratio	1,746845862	1,777286168	2,091146524	(-) Mažina riziką
idx_s_rev_as_ratio	1,656864588	0,980140793	1,219308107	(-) Mažina riziką
idx_lia_as_ratio_diff	1,524058717	1,398289028	0,764582135	(+) Didina riziką
sales_revenue_diff	1,513599459	0,956207251	1,07619458	(-) Mažina riziką
idx_equity_ratio_lag1	1,451584703	1,837560296	1,811745684	(-) Mažina riziką
net_profit	1,424967214	4,573639033	3,209372653	(-) Mažina riziką
profit_before_taxes	1,422967006	1,023027115	0,870238271	(-) Mažina riziką
vki_growth_ratio	1,411785954	1,30599464	1,782288647	(+) Didina riziką
ur_mean_lag1	1,324269889	0	1,505536903	(-) Mažina riziką
sales_revenue	1,300726426	1,057755316	0,964016975	(-) Mažina riziką
ur_mean	1,261566929	3,171656009	1,693536256	(-) Mažina riziką
short_term_assets_diff	1,202135709	0,481836609	0,82435026	(-) Mažina riziką
idx_lia_as_ratio_lag1	1,195044349	1,389232314	0,826842146	(+) Didina riziką

<b>vki_growth</b>	1,166063742	3,394988556	2,130257071	(+) Didina riziką
idx_s_rev_as_ratio_diff	1,087588544	0,379062362	0,302057573	(-) Mažina riziką
long_term_assets_lag1	0,984699947	0,679768818	1,237862208	(-) Mažina riziką
equity_chng	0,968498149	0,669840134	0,796893375	(-) Mažina riziką
<b>euribor_mean_ratio</b>	0,953018034	0,828204627	1,526007174	(+) Didina riziką
idx_s_rev_as_ratio_lag1	0,92118957	0,943900699	0,803781364	(-) Mažina riziką
short_term_assets	0,918094899	1,050626547	0,778481622	(-) Mažina riziką
idx_debt_to_equity_ratio_diff	0,909359597	0,731977177	1,329557409	(+) Didina riziką
net_profit_lag1	0,90724678	1,199074042	1,280347322	(-) Mažina riziką
profit_before_taxes_lag1	0,89232959	0,508863108	0,825220863	(-) Mažina riziką
net_profit_diff	0,890259794	0,51109795	0,638673633	(-) Mažina riziką
long_term_assets_chng	0,845232004	0,726872874	0,769258781	(-) Mažina riziką
idx_prof_tax_eq_ratio_lag1	0,788413655	0,748656514	0,544804325	(+) Didina riziką
<b>vki_growth_lag1</b>	0,763480975	0,082811687	0	(+) Didina riziką
idx_prof_tax_as_ratio_diff	0,754296791	0,124764139	0,276714179	(-) Mažina riziką
equity_lag1	0,746871769	0,437717801	0,50605862	(-) Mažina riziką
liabilities_chng	0,709395098	0,393556611	0,658966672	(+) Didina riziką
short_term_assets_chng	0,707684921	0,332400018	0,629604666	(-) Mažina riziką
net_profit_chng	0,677101795	0,47672075	0,745669221	(-) Mažina riziką
idx_prof_tax_eq_ratio_chng	0,658397976	0,347894192	0,155823389	(-) Mažina riziką
long_term_assets_diff	0,655219915	0,529673383	0,386612354	(-) Mažina riziką
idx_sh_as_lia_ratio_diff	0,619619321	0,072846337	0,250596179	(-) Mažina riziką
<b>ur_mean_ratio</b>	0,59027955	0	0,038518292	(+) Didina riziką
idx_prof_tax_eq_ratio_diff	0,588094321	0,581343972	0,322962401	(-) Mažina riziką
sales_revenue_lag1	0,577904661	0,501457452	0,660008929	(-) Mažina riziką
idx_prof_tax_as_ratio_lag1	0,572796514	0,212802577	0,241853832	(-) Mažina riziką
idx_debt_to_equity_ratio_chng	0,55806222	0,229986688	0,41008696	(-) Mažina riziką
idx_debt_to_equity_ratio_lag1	0,520497012	0,169615287	0,352595394	(-) Mažina riziką
idx_n_prof_as_ratio_lag1	0,489282721	0,82739156	0,876899215	(-) Mažina riziką
idx_n_prof_as_ratio_diff	0,476149512	0,396851856	0,235437405	(-) Mažina riziką

equity_diff	0,45338731	0,284591977	0,26340588	(-) Mažina riziką
idx_equity_ratio_diff	0,430149358	0,405461493	0,175594933	(-) Mažina riziką
idx_prof_tax_eq_ratio	0,417112098	0,28437484	0,47298876	(-) Mažina riziką
liabilities_diff	0,402094759	0,719683453	0,342440596	(+) Didina riziką
idx_equity_ratio_chng	0,380902255	0,60939584	0,314994655	(-) Mažina riziką
idx_lia_as_ratio_chng	0,35009484	0,100588634	0,058804963	(+) Didina riziką
idx_sh_as_lia_ratio_lag1	0,336928938	0,116963484	0,185765696	(-) Mažina riziką
idx_sh_as_lia_ratio_chng	0,31558665	0,327287746	0,179127246	(-) Mažina riziką
total_assets_diff	0,310254261	0,426528635	0,355034901	(-) Mažina riziką
short_term_assets_lag1	0,262093685	0,156697432	0,253346157	(-) Mažina riziką
idx_prof_tax_as_ratio_chng	0,249565239	0	0,087735321	(-) Mažina riziką
liabilities_lag1	0,233781043	0	0,148227038	(+) Didina riziką
total_assets_lag1	0,212010391	0	0,261332912	(-) Mažina riziką
<b>ur_mean_diff</b>	0,156889573	0	0	(+) Didina riziką
idx_n_prof_as_ratio_chng	0,142210977	0,126288899	0,078973959	(-) Mažina riziką
total_assets	0,11699787	0,103058763	0,318710683	(-) Mažina riziką
bvp_growth	0	0	0	(-) Mažina riziką
bvp_growth_lag1	0	0	0	(+) Didina riziką
euribor_mean_lag1	0	0	0	(-) Mažina riziką
<b>bvp_growth_diff</b>	0	0	0,039628964	(-) Mažina riziką
euribor_mean_diff	0	0	0	(+) Didina riziką
vki_growth_diff	0	0	0	(-) Mažina riziką
<b>bvp_growth_ratio</b>	0	0	0,048251472	(-) Mažina riziką