



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

**Aplinkosauginių, socialinės atsakomybės ir valdysenos  
rodiklių įtaka įmonių finansinei rizikai prognozuoti taikant  
mašininio mokymosi  
metodus**

Baigiamasis magistro studijų projektas

---

**Matas Povilonis**

Projekto autorius

**doc. dr. Šviesa Leitonienė**

Vadovė

**doc. dr. Loreta Saunorienė**

Vadovė

**Kaunas, 2026**



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

**Aplinkosauginių, socialinės atsakomybės ir valdysenos  
rodiklių įtaka įmonių finansinei rizikai prognozuoti taikant  
mašininio mokymosi  
metodus**

Baigiamasis magistro studijų projektas  
Didžiųjų verslo duomenų analitika (6213AX001)

---

**Matas Povilonis**

Projekto autorius

**doc. dr. Šviesa Leitonienė**

Vadovė

**doc. dr. Loreta Saunorienė**

Vadovė

**prof. dr. Lina Dagilienė**

Recenzentė

**vyresn. lekt. Vilma Petrauskienė**

Recenzentė

---

**Kaunas, 2026**



**Kauno technologijos universitetas**

Matematikos ir gamtos mokslų fakultetas

Matas Povilonis

## **Aplinkosauginių, socialinės atsakomybės ir valdysenos rodiklių įtaka įmonių finansinei rizikai prognozuoti taikant mašininio mokymosi metodus**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autorius ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Matas Povilonis

*Patvirtinta elektroniniu būdu*

Povilonis, Matas. Aplinkosauginių, socialinės atsakomybės ir valdysenos rodiklių įtaka įmonių finansinei rizikai prognozuoti taikant mašininio mokymosi metodus. Magistro studijų baigiamasis projektas / projektui vadovavo doc. dr. Šviesa Leitonienė ir doc. dr. Loreta Saunorienė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

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

Reikšminiai žodžiai: ESG, finansinė rizika, mašininis mokymasis, prognozavimas, modeliai, interpretavimas.

Kaunas, 2026. 77 p.

### **Santrauka**

Augantis dėmesys įmonių tvarumui ir griežtėjantys reguliaciniai reikalavimai lemia poreikį finansinės rizikos vertinimą papildyti nefinansiniais – aplinkosaugos, socialinės atsakomybės ir valdysenos, rodikliais. Nors mokslinėje literatūroje pastebima, kad aukštesni tvarumo rodikliai gali būti susiję su geresniais įmonių finansiniais rezultatais ir mažėjančia rizika, vis dar trūksta tyrimų, kurie sistemingai įvertintų šių veiksnių įtaką finansinės rizikos prognozavimui taikant mašininio mokymosi metodus. Baigiamasis magistro projektas yra skirtas iširti aplinkosauginių, socialinės atsakomybės ir valdysenos rodiklių įtaką prognozuojant įmonių finansinę riziką. Literatūros apžvalgoje pristatomas finansinės rizikos apibrėžimas ir jos vertinimo metodai, apžvelgiami moksliniai darbai tiriantys mašininio mokymosi metodų taikymą finansinės rizikos prognozavimo uždaviniuose. Mokslinės literatūros apžvalgoje išskiriamos dvi pagrindinės šios tyrimų srities problemos: finansinės rizikos apibrėžimo nevienareikšmiškumas ir tvarumo rodiklių duomenų kokybė. Dėl šios priežasties, darbe tiriami penki skirtingi finansinės rizikos apibrėžimai ir kiekvienam iš jų sudaromi mašininio mokymosi modeliai. Kadangi duomenų imtyje įmonių, patiriančių finansinius sunkumus, yra mažai (apie 10-15% visų įmonių), šiame darbe buvo pritaikytas papildomas metodas spręsti duomenų disbalanso problemą. Atliekamas tyrimas į finansinės rizikos prognozavimą įtraukia ne tik finansinius įmonių duomenis, bet ir agreguotus aplinkosaugos, socialinės atsakomybės bei valdysenos balus. Tyrime lyginami mašininio mokymosi modelių rezultatai dviem duomenų imtims: įtraukiant aplinkosauginius, socialinius ir valdysenos balus bei jų neįtraukiant. Nepriklausomai nuo pasirinkto finansinės rizikos apibrėžimo (tikslų kintamojo) nėra statistiškai reikšmingo skirtumo tarp modelių, kuriuose įtraukiami šie tvarumo rodikliai, ir modelių, kuriuose yra tik finansiniai įmonės duomenys. Mašininio mokymosi modelių prognozavimo tikslumas skiriasi, priklausomai nuo tikslo kintamojo apibrėžimo, tačiau tarp geriausių modelių visada išlieka aukštesnis nei 87%. Tyrime pritaikyti mašininio mokymosi modelių interpretavimo metodai parodė, kad rinkos kapitalizacija, grynasis pelnas, veiklos pelnas prieš mokesčius ir palūkanas yra pagrindiniai modelio prognozę nulemiantys veiksniai. Gauti interpretavimo metodų rezultatai yra reikšmingi tiek vienos įmonės lygiu, tiek bendrame ekonominiame, socialiniame ir verslo vadybos kontekste, nes sudaro galimybę priimti labiau pagrįstus sprendimus, vertinant įmonių finansinę riziką.

Povilonis, Matas. Environmental, Social Responsibility and Governance Indicators' Impact on Corporate Financial Risk Prediction Applying Machine Learning Methods. Master's Final Degree Project/ supervisor doc. dr. Šviesa Leitonienė and doc. dr. Loreta Saunorienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

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

Keywords: ESG, financial risk, machine learning, prediction, models, interpretation.

Kaunas, 2026. 77.

### **Summary**

The growing attention to corporate sustainability and tightening regulatory requirements create a need to supplement financial risk assessment with non-financial indicators – environmental, social responsibility, and governance metrics. Although the scientific literature suggests that higher sustainability scores may be associated with better financial performance and lower risk, there is still a lack of studies that systematically evaluate the impact of these factors on financial risk prediction using machine learning methods. The master's thesis is dedicated to investigating the impact of environmental, social responsibility, and governance indicators on predicting corporate financial risk. The literature review presents the definition of financial risk and its assessment methods, and reviews scientific studies examining the application of machine learning methods in financial risk prediction tasks. The scientific literature review identifies two main problems in this research field: the ambiguity of the definition of financial risk and the quality of sustainability indicator data. For this reason, the thesis examines five different definitions of financial risk and develops machine learning models for each of them. Since the data sample contains few companies experiencing financial difficulties (about 10–15% of all companies), an additional method was applied in this thesis to address the problem of data imbalance. The conducted research into financial risk prediction includes not only companies' financial data, but also aggregated environmental, social responsibility, and governance scores. The study compares the results of machine learning models using two data samples: one including environmental, social, and governance scores and one excluding them. Regardless of the chosen definition of financial risk (target variable), there is no statistically significant difference between models that include these sustainability indicators and models that contain only company financial data. The predictive accuracy of the machine learning models differs depending on the definition of the target variable; however, among the best models, it always remains higher than 87%. The machine learning model interpretation methods applied in the study showed that market capitalization, net profit, and operating profit before taxes and interest are the main factors determining the model's predictions. The results obtained from the interpretation methods are significant both at the individual company level and in the broader economic, social, and business management context, as they provide the opportunity to make more informed decisions when assessing corporate financial risk.

## Turinys

<b>Lentelių sąrašas.....</b>	<b>8</b>
<b>Paveikslų sąrašas .....</b>	<b>9</b>
<b>Santrumpų ir terminų sąrašas.....</b>	<b>10</b>
<b>Įvadas.....</b>	<b>11</b>
<b>1. Literatūros analizė.....</b>	<b>12</b>
1.1. ESG rodiklių samprata ir reikšmė vertinant finansinę riziką .....	12
1.1.1. Aplinkosauginiai veiksniai .....	12
1.1.2. Socialiniai veiksniai.....	13
1.1.3. Valdymo veiksniai.....	13
1.2. ESG įtaka įmonių finansiniams rezultatams.....	14
1.2.1. ESG įtaka prognozuojant įmonių finansinius rodiklius.....	17
1.3. Finansinės rizikos prognozavimas.....	18
1.3.1. Mašininis mokymasis prognozuojant finansinę riziką .....	20
1.3.2. ESG įtraukimas į mašininio mokymosi modelius .....	22
1.3.3. Finansinės rizikos prognozavimas įtraukiant ESG rodiklius.....	23
1.4. Darbo pagrindimas .....	24
<b>2. Tyrimo metodai .....</b>	<b>26</b>
2.1. Duomenys.....	26
2.2. Finansinės rizikos (priklausomo kintamojo) apibrėžimas .....	26
2.3. Duomenų disbalanso problema ir jos sprendimo būdai .....	27
2.4. Mašininio mokymosi metodai .....	28
2.4.1. Logistinė regresija .....	28
2.4.2. Atraminių vektorių metodas .....	28
2.4.3. XGBoost metodas.....	30
2.4.4. LightGBM metodas .....	31
2.4.5. Atsitiktiniai miškai .....	32
2.4.6. EasyEnsemble metodas .....	32
2.5. Mašininio mokymosi modelių gerumo vertinimas.....	33
2.5.1. Statistiniai gerumo matai .....	33
2.5.2. Modelių rezultatų kreivės .....	35
2.6. Kintamųjų atranka .....	35
2.7. Mašininio mokymosi modelių interpretavimas .....	35
2.7.1. SHAP metodas.....	36
2.7.2. DiCE metodas.....	38
2.8. Statistinis modelių palyginimas.....	38
2.9. Tyrimo eiga .....	39
<b>3. Tyrimo rezultatai.....</b>	<b>40</b>
3.1. Žvalgomoji analizė .....	40
3.2. Finansinių sunkumų prognozavimas .....	41
3.2.1. Hiperparametrų optimizavimas .....	42
3.2.2. Finansinių sunkumų prognozavimas, kai tikslo kintamas paremtas Altman'o reikšmėmis..	44
3.2.3. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas grynuoju pelnu ...	50
3.2.4. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą.....	55

3.2.5. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas veiklos pelnu .....	60
3.2.6. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra kombinuotas .....	65
<b>Išvados .....</b>	<b>71</b>
<b>Literatūros sąrašas .....</b>	<b>72</b>
<b>Priedai.....</b>	<b>78</b>
1 priedas.....	78

## Lentelių sąrašas

<b>1 lentelė.</b> ESG įtakos įmonių finansiniams rezultatams tyrimų palyginimas.....	16
<b>2 lentelė.</b> Sumaišymo matrica .....	34
<b>3 lentelė.</b> Reikšmių pasiskirstymas pagal tikslo kintamuosius .....	40
<b>4 lentelė.</b> Šalių pasiskirstymas .....	40
<b>5 lentelė.</b> Įmonių veiklos sektoriai .....	41
<b>6 lentelė.</b> XGBoost hiperparametrai.....	42
<b>7 lentelė.</b> LightGBM hiperparametrai .....	43
<b>8 lentelė.</b> Atsitiktinių miškų hiperparametrai.....	43
<b>9 lentelė.</b> EasyEnsamble hiperparametrai .....	44
<b>10 lentelė.</b> Logistinės regresijos hiperparametrai.....	44
<b>11 lentelė.</b> Atraminių vektorių hiperparametrai .....	44
<b>12 lentelė.</b> Modelių rezultatai (be ESG), kai tikslo kintamasis yra pagal Altman'o rodiklį .....	45
<b>13 lentelė.</b> Modelių rezultatai (su ESG), kai tikslo kintamasis yra pagal Altman'o rodiklį .....	45
<b>14 lentelė.</b> Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas neigiamu grynuoju pelnu .....	50
<b>15 lentelė.</b> Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas neigiamu grynuoju pelnu .....	51
<b>16 lentelė.</b> Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą.....	55
<b>17 lentelė.</b> Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą.....	56
<b>18 lentelė.</b> Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas neigiamu veiklos pelnu .....	60
<b>19 lentelė.</b> Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas neigiamu veiklos pelnu .....	61
<b>20 lentelė.</b> Modelių rezultatai (be ESG), kai tikslo kintamasis yra kombinuotas .....	65
<b>21 lentelė.</b> Modelių rezultatai (su ESG), kai tikslo kintamasis yra kombinuotas .....	65

## Paveikslų sąrašas

<b>1 pav.</b> Atraminiai vektoriai .....	29
<b>2 pav.</b> ROC kreivė.....	35
<b>3 pav.</b> „SHAP Summary Plot“ .....	37
<b>4 pav.</b> „SHAP Waterfall Plot“ .....	37
<b>5 pav.</b> Tyrimo eigos schema.....	39
<b>6 pav.</b> Agreguoto aplinkosaugos balo histograma .....	41
<b>7 pav.</b> Kintamųjų svarba LightGBM su SMOTE (be ESG).....	46
<b>8 pav.</b> Kintamųjų svarba XGBoost (su ESG).....	47
<b>9 pav.</b> „SHAP summary plot“ XGBoost metodui.....	48
<b>10 pav.</b> „SHAP waterfall“ XGBoost metodui.....	49
<b>11 pav.</b> DiCE metodo alternatyvūs scenarijai pasirinktai įmonei .....	49
<b>12 pav.</b> Kintamųjų svarba XGBoost (be ESG) .....	52
<b>13 pav.</b> Kintamųjų svarba XGBoost (su ESG).....	52
<b>14 pav.</b> „SHAP summary plot“ XGBoost modeliui.....	53
<b>15 pav.</b> „SHAP waterfall“ XGBoost modeliui.....	54
<b>16 pav.</b> DiCE metodo alternatyvūs scenarijai pasirinktai įmonei .....	54
<b>17 pav.</b> Kintamųjų svarba RandomForest modelyje su svoriais (be ESG) .....	57
<b>18 pav.</b> Kintamųjų svarba RandomForest modelyje (su ESG) .....	57
<b>19 pav.</b> „SHAP summary plot“ RandomForest modeliui .....	58
<b>20 pav.</b> „SHAP waterfall“ RandomForest modeliui .....	59
<b>21 pav.</b> DiCE metodo alternatyvūs scenarijai pasirinktai įmonei .....	59
<b>22 pav.</b> Kintamųjų svarba XGBoost su SMOTE modelyje (be ESG) .....	62
<b>23 pav.</b> Kintamųjų svarba LightGBM su SMOTE modelyje (su ESG).....	62
<b>24 pav.</b> „SHAP summary plot“ LightGBM su SMOTE modeliui .....	63
<b>25 pav.</b> „SHAP waterfall“ LightGBM su SMOTE modeliui.....	64
<b>26 pav.</b> Kintamųjų svarba LightGBM modelyje (be ESG).....	66
<b>27 pav.</b> Kintamųjų svarba RandomForest su SMOTE modelyje (su ESG) .....	67
<b>28 pav.</b> „SHAP summary plot“ LightGBM su SMOTE modeliui .....	68
<b>29 pav.</b> „SHAP waterfall“ LightGBM su SMOTE modeliui.....	69
<b>30 pav.</b> DiCE metodo alternatyvūs scenarijai pasirinktai įmonei .....	69

## **Santrumpų ir terminų sąrašas**

ESG – aplinkosaugos, socialinės atsakomybės ir valdysenos veiksniai (angl. *Enviromental, Social and Governance*) [1].

## Įvadas

Pastaraisiais metais ESG (aplinkosaugos, socialinės atsakomybės ir valdysenos) rodikliai tampa vis labiau svarbesni ir įtraukiami į įmonių veiklos ir finansinio stabilumo vertinimą. Šie veiksniai leidžia įvertinti ne tik finansinius rezultatus, bet ir su įmonės veikla susijusius nefinansinius aspektus. Dėl augančio ESG duomenų prieinamumo, šie rodikliai vis dažniau analizuojami finansinės rizikos prognozavimo kontekste, nes prastesni tvarumo rezultatai dažnai siejami su didesnėmis kapitalo pritraukimo sąnaudomis, reputacinėmis krizėmis ir veiklos nestabilumu. Praktiniu lygmeniu šio darbo rezultatai gali būti naudingi įvairioms suinteresuotoms šalims: įmonėms – formuojant veiklos vykdymo strategijas, finansų institucijoms – vertinant kredito riziką, investuotojams – priimant sprendimus dėl portfelio sudėties. Visa tai rodo, kad nagrinėjama tema yra ne tik **aktuali**, bet ir turinti realų poveikį šiuolaikiniam verslo sprendimų priėmimui. Darbo **naujumas** susiejęs su tuo, kad finansinė rizika prognozuojama panaudojant ne tik finansinius įmonių duomenis, bet visas ESG dedamąsias ir taikant pažangius mašininio mokymosi metodus.

Nepaisant to, kad moksliniai tyrimai patvirtina, kad ESG (aplinkosaugos, socialiniai ir valdysenos) veiksniai reikšmingai veikia įmonių finansinius rezultatus ir rizikos lygį, vis dar trūksta mokslinių tyrimų, nagrinėjančių efektyvų ESG rodiklių integravimą į pažangius finansinių sunkumų prognozavimo modelius, paremtus mašininio mokymosi technologijomis. Neaišku, kaip ir kokiu mastu ESG duomenų įtraukimas gali pagerinti finansinės rizikos vertinimo tikslumą bei finansinių sunkumų prognozavimo patikimumą.

Šio darbo tyrimo metu realizuotų interpretavimo metodų rezultatai gali būti pritaikomi finansinės rizikos vertinime, padedant priimti labiau pagrįstus sprendimus, kadangi gaunamas ne tik prognozės rezultatas – patirs finansinių sunkumų ar ne, bet ir tą rezultatą nulėmę kintamieji, jų reikšmės. Toks pritaikomumas gali būti naudingas investuotojams, finansų analitikams, įmonių vadovams.

**Baigiamojo projekto tikslas** – įvertinti ESG rodiklių įtaką įmonių finansinės rizikos prognozavime, taikant mašininio mokymosi modelius.

### Uždaviniai:

- Atlikti literatūros analizę, siekiant įvertinti ryšį tarp nefinansinių (aplinkosaugos, socialinės atsakomybės ir valdysenos) veiksnių ir įmonių finansinės rizikos bei identifikuoti tinkamiausius finansinės rizikos matavimo rodiklius, kurie galėtų būti naudojami kaip priklausomi (tikslų) kintamieji mašininio mokymosi modeliuose.
- Parinkti tinkamus mašininio mokymosi metodus finansinės rizikos prognozavimui bei sudaryti prognozavimo modelius dviem tyrimo duomenų imtims – su ESG rodikliais ir be ESG rodiklių.
- Palyginti sukurtų mašininio mokymosi modelių prognozavimo tikslumą, įvertinti ar tarp geriausio modelio be ESG rodiklių ir geriausio modelio su ESG rodikliais yra statistiškai reikšmingas prognozavimo tikslumo skirtumas.
- Pasirinktam, geriausiai finansinę riziką prognozuojančiam modeliui, pritaikyti mašininio mokymosi modelių interpretavimo metodus, siekiant identifikuoti svarbiausius, prognozę lemiančius, nepriklausomus kintamuosius.

## **1. Literatūros analizė**

### **1.1. ESG rodiklių samprata ir reikšmė vertinant finansinę riziką**

Trumpinys ESG yra plačiai vartojamas finansų srityje, įmonių veikloje, akademinėje veikloje, viešajame sektoriuje ar politikoje. Verčiant iš anglų kalbos, šis trumpinys savyje saugo tris aspektus: aplinkosauginį, susijusį su įmonės poveikiu gamtinei aplinkai ir klimato kaitai, socialinį, apimančią santykius su darbuotojais, klientais, visuomene ir valdysenos, kuris atspindi įmonės valdymo struktūrą, valdymo praktiką bei skaidrumą [1]. Pastaraisiais metais ESG kriterijai tapo svarbia verslo vertinimo dalimi. Įmonių vertinimas remiantis ESG principais reiškia ne tik tradicinių finansinių rodiklių analizę, bet ir gebėjimą įvertinti nematerialius veiksniai, darančius ilgalaikį poveikį įmonės vertei.

#### **1.1.1. Aplinkosauginiai veiksniai**

Aplinkosaugos aspektas ESG koncepte akcentuoja įstaigų poveikio gamtinei aplinkai vertinimą ir valdymą, apimdamas tiek kiekybinius rodiklius, tokius kaip šiltnamio efektą sukeliančių dujų emisijos, sunaudojamą energiją, vandens kiekį ir šių resursų panaudojimo efektyvumą, tiek sunkiau apibrėžiamus kokybinius aspektus, pavyzdžiui, strateginį įmonės požiūrį klimato kaitos perspektyvoje [2]. Bendras aplinkosauginis ESG rodiklis išreiškiamas agreguotu balu, kuris apibendrina įmonės poveikį aplinkai ir gebėjimą valdyti su tuo susijusias rizikas. Iš pradžių aplinkosaugos kriterijaus integravimas į finansinę analizę buvo pasiūlytas kaip viena iš rizikos vertinimo formų, vėliau, po Jungtinių Tautų ataskaitos „Who Cares Wins“ [3], buvo raginama įtraukti aplinkosaugos veiksniai ir į investavimo strategiją, taip sukuriant aplinkybes investuotojams pasirinkti tvarias investicijas į korporacijas, kurios atsižvelgė į aplinkosaugą [1]. Mackey'aus straipsnyje, kuris fokusuojasi į Jungtines Amerikos Valstijas, teigiama, kad ESG koncepcija kilo kaip reakcija į vyriausybės nesugebėjimą tvarkytis su plataus masto problemomis – aplinkos ir gamtos išsaugojimas yra viena iš jų [4]. Šiame straipsnyje pabrėžiama, kad atsakomybė spręsti aplinkosaugos ir socialines problemas palaipsniui buvo perkeliama iš viešojo į privatų sektorių. Nors tai yra tik Jungtinių Amerikos Valstijų pavyzdys, po tarptautinių iniciatyvių, tokių kaip Paryžiaus klimato susitarimas ir Jungtinių tautų darnaus vystymosi tikslai, buvo dar labiau sustiprintas spaudimas, viso pasaulio mastu, verslui atsižvelgti ir taikyti aplinkosauginius principus savo veikloje [2]. Visa tai atsispindi ir įmonių finansinėse ataskaitose, reitingų sistemose, kuriose įtraukiami aplinkosauginiai rodikliai kaip sudedamoji dalis vertinant bendrą įmonės padėtį rinkoje bei tvarumą. Taigi, ESG aplinkosauginio rodiklio E balas gali būti interpretuojamas kaip indikatorius, rodantis įmonės jautrumą aplinkosauginės rizikos veiksniams ir galimą jų finansinį poveikį įmonės veiklai.

„Task Force on Climate-related Financial Disclosures“ (TCFD) yra tarptautinė iniciatyva, inicijuota Finansinio stabilumo tarnybos, kurios tikslas – padėti įmonėms identifikuoti ir atskleisti su klimato kaita susijusias finansines rizikas. Remiantis TCFD išleista ataskaita [5], aplinkosauginiai veiksniai tiesiogiai veikia įmonių finansinės veiklos rodiklius ir nesuvaldytos rizikos gali pasireikšti per didesnes veiklos sąnaudas, turto nuvertėjimą ar pajamų sumažėjimą. Empiriniai tyrimai [6] taip pat rodo, kad prastesni aplinkosauginiai rodikliai yra siejami su didesne kapitalo kaina ir didesniu finansiniu neapibrėžtumu. Taigi, aplinkosauginė rizika vis labiau pripažįstama kaip reikšmingas sisteminės rizikos faktorius.

### 1.1.2. Socialiniai veiksniai

Socialinis aspektas ESG koncepcijoje akcentuoja organizacijų santykį su darbuotojais, klientais, bendruomenėmis bei visuomene plačiąja prasme, vertinant tiek įmonės poveikį socialinei aplinkai, tiek gebėjimą kurti ilgalaikę vertę per socialinę atsakomybę ir įsitraukimą. Šis komponentas taip pat gali būti vertinamas tiek per kiekybinių, tiek per kokybinių rodiklių prizmę: darbuotojų sauga ir sveikata, lyčių pasiskirstymas, lygybė, įvairovė ir įtrauktis, darbo teisės, klientų duomenų apsauga, tiekimo grandinės, darbo sąlygų skaidrumas ar bendruomenės gerovės kūrimas [7]. Kaip pažymi Macey'us, socialinė dalis ESG sistemoje atspindi ne tik privataus sektoriaus atsakomybę ten, kur valstybė nebesugeba užtikrinti socialinio teisingumo, bet ir platesnę visuomenės lūkesčių transformaciją – verslas turi būti nebe tik pelno siekiantis subjektas, bet ir aktyvus socialinės pažangos dalyvis [4]. Socialinis ESG rodiklis S taip pat išreiškiamas agreguotu balu, kuris gali būti interpretuojamas kaip indikatorius, rodantis įmonės gebėjimą valdyti socialines (žmogiškąsias) rizikas. Žema šio rodiklio reikšmė gali atspindėti vykstančius streikus, didelę darbuotojų kaitą, klientų praradimą ar reputacijos kritimą, o tai indikuoja nestabilius pinigų srautus ir didesnę finansinę riziką [7].

Atlikti tyrimai ESG ir finansinės rizikos srityje rodo, kad socialinis aspektas yra reikšmingai susijęs su įmonių finansine rizika, ypač per jo poveikį įmonės veiklos stabilumui, reputacijai ir pinigų srautų nepastovumui. Khan'as ir kt. [7] savo tyrime parodė, kad įmonės, kurios skiria dėmesį ir valdo socialines rizikas, pasižymi stabilesniais ir geresniais finansiniais rezultatais ir mažesniu finansinės rizikos lygiu, tačiau autoriai akcentuoja, kad ne visi tiek aplinkosauginiai, tiek socialiniai, tiek valdysenos aspektai yra svarbūs. Finansinės rizikos įvertinimui reikšmingi yra tik tie, kurie susiję su įmonės veikla plačiąja prasme. Panašias išvadas pateikė ir Servaes'as ir Tamayo'as [8], kurie nustatė, kad stipresnė socialinė atsakomybė yra siejama su didesne įmonės verte, ypač tais atvejais, kai įmonės reputacija rinkoje yra labai svarbi. Kadangi yra išlaikomas didesnis investuotojų pasitikėjimas, dėl aukštos įmonės reputacijos, atitinkamai mažėja finansinis neapibrėžtumas – rizika. Dyck'as ir kt. analizavo, kaip instituciniai investuotojai (pensijų fondai, bankai, investiciniai fondai) daro įtaką įmonių aplinkos ir socialiniams ESG rodikliams, ir nustatė, kad didesnė institucinė nuosavybė yra susijusi su aukštesniu šių rodiklių lygiu. Autoriai taip pat parodo, kad investuotojai yra motyvuoti ne tik finansine grąža, bet jiems svarbūs ir socialiniai įmonės veiksniai – darbuotojai, etika, aplinka, todėl įmonės yra skatinamos gerinti socialinį ESG aspektą [9]. Straipsnyje finansinė rizika nėra tiesiogiai analizuojama, bet galima daryti išvadą, kad geresnis socialinių veiksmų valdymas, skatinamas investuotojų, gali prisidėti prie įmonės veiklos stabilumo ir mažesnės finansinės rizikos.

### 1.1.3. Valdymo veiksniai

Valdymo komponentas ESG struktūroje apima organizacijos valdymo sistemų, struktūrų, procesų ir etinių principų vertinimą, kurie lemia tiek įmonės skaidrumą, tiek jos ilgalaikį veiklos stabilumą. Ši dimensija apima tokius elementus kaip valdybos sudėtis ir nepriklausomumas, akcininkų teisės, atlygio politika, antikorpucinės priemonės, vidaus kontrolė bei aiškus atsakomybių pasiskirstymas [2]. Tinkamas valdymas laikomas pagrindiniu veiksmu, užtikrinančiu, kad aplinkosauginiai ir socialiniai įsipareigojimai nebus vien tik simboliniai – būtent nuo valdymo priklauso, ar ESG principai bus realiai įgyvendinti. Šis valdymo komponentas, kaip aplinkosauginis ir socialinis, išreiškiamas agreguotu balu, atspindinčiu kaip gerai įmonė yra valdoma ir kontroliuojama. Žemas šio komponento balas siejamas su vadovo veikimu savo naudai, netinkamais investiciniais sprendimais ar apskaitos klaidomis, o finansinės rizikos kontekste tai siejasi su didesne bankroto tikimybe ir

finansiniu nestabilumu. Viena dažniausių problemų įmonės valdymo srityje yra interesų nesutapimas tarp įmonės savininkų (akcininkų) ir vadovų, nes vadovai, turintys sprendimų priėmimo galią, gali siekti jiems asmeniškai naudingų tikslų, kurie nebūtinai sutampa su įmonės tikslais ir įmonės vertės auginimu [10]. Tokia situacija gali lemti neefektyvų investicijų pasirinkimą ir perteklines išlaidas, atitinkamai didinant įmonės finansinę riziką. Gomper'io ir kt. atliktas tyrimas [11] rodo, kad įmonėse, kuriose akcininkų „balsas“ yra silpnesnis ir vadovų kontrolė mažesnė, pasižymi žemesne rinkos verte ir akcijų grąža bei mažesniu pelningumu. Taip pat, Tettamanzi'io straipsnyje [2] minima, kad skaidrus ir atsakingas valdymas, ypač kai aiškiai apibrėžti tikslai ir kontrolės mechanizmai, siejasi su aukštesne ESG ataskaitų kokybe bei didesniu investuotojų pasitikėjimu. Tvarumo ataskaitų rengimo standarte ISSB [12] aiškiai įtvirtinama vadovų atsakomybė už ESG veiksmų integravimą į įmonės strategiją ir rizikos valdymą. Taigi, valdymo aspektas ESG sistemoje užtikrina institucinį pagrindą, kuriuo pagrįstas tiek aplinkosaugos, tiek socialinių tikslų įgyvendinimas.

Edmans'as savo kritinėje analizėje [13] teigia, kad ESG neturėtų būti laikomas nei išskirtiniu, nei atskiru investicijų vertinimo elementu. Jo vertinimu, metrikos, ateinančios iš ESG ideologijos, yra tik vienos iš daugelio nematerialių įmonės vertės sudedamųjų dalių, pavyzdžiui, tokių kaip darbuotojų lojalumas. ESG analizė naudinga tik tada, kai ji padeda identifikuoti veiksmus, darančius įtaką ilgalaikiai įmonės verte, ar kitaip tariant, kai ESG rodikliai tampa pagrindiniais veiklos rodikliais (angl. *Key performance indicator*). Šio autoriaus pozicija kritikuoja perteklinę ESG metrikų naudojimą bei rodo pavojų, kad įmonės gali koncentruotis į lengvai pamatuojamus, bet ne esminius rodiklius.

Apibendrinant galima daryti išvadą, kad ESG yra sistema, integruojanti aplinkosauginius, socialinius ir valdymo aspektus į verslo ir investicijų praktiką, siekiant kurti ilgalaikę vertę ne tik akcininkams, bet ir visai visuomenei. Nors trumpinys ESG iš pradžių buvo neaiškus ir vartotas tik tam tikrose, siaurose nišose, šiandien jis tapo kertiniu principu globaliame tvarumo, atsakomybės ir rizikų valdymo kontekste. ESG integracija ir vertinimas neapsiriboja vien tik rizikos valdymu, o padeda išvelgti galimybes, pavyzdžiui, inovatyvius verslo modelius, kurie yra labiau pritaikyti prie būsimo tvarumo politikos [14].

## **1.2. ESG įtaka įmonių finansiniams rezultatams**

ESG ir finansinių rezultatų ryšys pastaraisiais metais tapo svarbia empirinių tyrimų kryptimi. Vis didesnis dėmesys tvarumui skatina analizuoti ESG reikšmę ne tik bendram įmonės vertinimui, bet ir finansiniams rezultatams. Tokiuose tyrimuose bandoma kiekybiškai įvertinti, ar ESG veiksniai yra susiję su įmonių pelningumu, verte bei finansine rizika. Empiriniai tyrimai skiriasi ne tik naudojamais duomenimis, bet ir pasirinktais priklausomais kintamaisiais, kurie dažniausiai apima įmonės pelningumo rodiklius, rinkos vertės matavimus ar kapitalo kaštus. Labai svarbiu aspektu šiuose tyrimuose tampa ESG rodiklių skaidymas į atskirus komponentus ir jų reikšmingumo vertinimas, siekiant tiksliau nustatyti jų poveikį finansiniams rezultatams.

Khan'as ir kt. atliko tyrimą [7], kuriame siekė nustatyti ar investicijos į tvarumą didina įmonės vertę. Autoriai atskyrė reikšmingus (svarbius konkrečiai pramonės šakai ir turinčius realų poveikį) ir nereikšmingus (mažiau su įmonės veikla susijusius) ESG veiksmus ir sukūrė reikšmingumo ir ne reikšmingumo indeksus, kuriuos naudojo, kaip nepriklausomus kintamuosius analizuojant jų poveikį būsimai akcijų grąžai. Tyrimo metu, pritaikius regresijos modelį, buvo nustatyta, kad tik materialūs ESG turi statistiškai reikšmingą teigiamą poveikį būsimai akcijų grąžai, o ne materialūs ESG

veiksniai statistiškai reikšmingo poveikio būsimai akcijų grąžai neturi. Remiantis šiuo tyrimu, galima daryti išvadą, kad įmonei svarbus ne tik ESG rodiklių lygis, bet ir jų reikšmė, tiesioginė įtaka konkrečiai įmonei ar sektoriui.

ESG poveikį įmonės nuosavo kapitalo kainai analizavo Ng'as ir Rezaee'as [15]. Tyrimo metu, kaip priklausomas kintamasis, buvo pasirinkta nuosavo kapitalo kaina, o nepriklausomi kintamieji – ESG komponentai ir ECON (ekonominiai įmonės rodikliai). Autoriai nustatė, kad aukštesni ESG rodiklių įverčiai yra susiję su mažesne kapitalo kaina, tačiau skirtingi ESG komponentai daro nevienodą poveikį. Atliktame tyrime didžiausią reikšmę turėjo aplinkosaugos ir valdymo komponentai, būtent jie stipriau prisideda prie kapitalo kainos mažinimo. Šie rezultatai leidžia daryti prielaidą, kad ESG gali būti vertinamas ir kaip rizikos mažino veiksnys, turintis įtakos būtent investuotojų reikalaujamai grąžai.

Dalies tyrimų dėmesys sutelkiamas į ESG poveikį įmonių pelningumui ir rinkos vertei. Puri atliko tyrimą [16], kuriame analizavo ar ESG atskleidimo (padengiamumo) balas turi reikšmingą poveikį finansiniams pelningumo rodikliams (ROA, ROE) ir rinkos vertei (Tobin's Q). Rezultatai rodo, kad įmonės norėdamos įsitraukti į ESG rodiklių išpildymą, ar ESG rodiklių gerinimą gali didinti veiklos kaštus, dėl to trumpuoju laikotarpiu buvo pastebėtas neigiamas poveikis pelningumui, tačiau nustatytas teigiamas ESG poveikis įmonės rinkos vertei ir tai rodo, kad investuotojai ESG veiklą vertina palankiai, nepaisant trumpalaikių sąnaudų padidėjimo. Autorė tyrimą papildė nepriklausomu kintamuoju „finansiniai ištekliai“, siekdama nustatyti ar laisvi finansiniai įmonės resursai keičia ESG poveikį pelningumo rodikliams. Finansiniai resursai tyrime naudojami kaip moderuojantis kintamasis, kuris parodo, kad ESG poveikis finansiniams rezultatams priklauso nuo įmonės finansinių galimybių. Įmonėse su didesniais finansiniais resursais ESG turi teigiamą poveikį pelningumui, o ribotų išteklių atveju – neigiamą. Xu ir Zhu [17] taip pat analizavo ESG poveikį įmonių finansiniams rezultatams, panaudojant moderuojantį kintamąjį „finansiniai resursai“. Autoriai nustatė, kad ESG teigiamai veikia įmonių pelningumą, tačiau šis poveikis nėra tiesioginis. Pelningumas yra veikiamas per inovacijas – ESG skatina inovacijas, o inovacijos didina pelningumą. Šių autorių tyrimas, kaip ir Puri [16], atskleidžia, kad finansiniai apribojimai silpnina ESG poveikį pelningumui, kadangi riboja įmonių galimybes investuoti į tvarumo iniciatyvas.

Nazarova'os ir kt. tyrime [18] analizuojamas ESG informacijos atskleidimo poveikis įmonių finansiniams rezultatams, ypatingą dėmesį skiriant skirtumams tarp privalomo ir savanoriško ESG atskleidimo. Tyrimo aktualumas grindžiamas sparčiai didėjančiu ESG informacijos poreikiu iš investuotojų, reguliuotojų ir kitų suinteresuotųjų šalių, taip pat tuo, kad iki šiol mokslinėje literatūroje nėra vieningos nuomonės, ar ESG atskleidimas kuria finansinę vertę įmonėms. Pagrindinis tyrimo tikslas – empiriškai įvertinti, ar privalomas ESG atskleidimas daro kitokį poveikį įmonių finansiniams rezultatams nei savanoriškas ESG atskleidimas. Šiuo tikslu Europos įmonės laikomos tiriamųjų įmonių grupe, o JAV įmonės – kontrolinė grupe, nes analizuojamu laikotarpiu JAV nebuvo taikomi privalomi ESG atskleidimo reikalavimai. Empirinio tyrimo rezultatai rodo, kad privalomas ESG atskleidimas turi statistiškai reikšmingą neigiamą poveikį visiems trimis finansinių rezultatų rodikliams – ROA, ROE ir Tobin's Q. Neigiamas sąveikos kintamojo koeficientas leidžia daryti išvadą, kad po direktyvos įsigaliojimo Europos įmonių finansiniai rezultatai pablogėjo, palyginti su JAV įmonėmis. Autoriai interpretuoja gautus rezultatus teigdami, kad privalomas ESG atskleidimas gali paskatinti įmones perinvestuoti į tam tikras ESG sritis, ypač aplinkosaugą, siekiant formaliai atitikti reguliacinius reikalavimus, tačiau tokios investicijos nebūtinai kuria finansinę vertę.

Naujesni tyrimai analizuoja ESG poveikio stabilumą skirtingais ekonominiais laikotarpiais. Azizah‘o ir Haron‘o tyrime [19] nagrinėjamas ESG poveikis Indonezijos ir Malaizijos įmonių finansiniams rezultatams prieš ir po COVID 19 pandemijos. Autoriai pažymi, kad nors ESG integracija dažnai laikoma strateginiu pranašumu, jos finansinis poveikis gali būti nevienodas skirtingais laikotarpiais ir gali priklausyti nuo išorinių ekonominių sąlygų. Tyrimo rezultatai rodo, kad ESG ir finansinių rezultatų ryšys yra dinamiškas ir priklauso nuo ekonominio konteksto. Prieš pandemiją socialiniai ir valdysenos veiksniai turėjo teigiamą ir statistiškai reikšmingą poveikį ROA bei Tobin‘o Q, o tai leidžia daryti išvadą, kad stipri valdysena ir socialinė atsakomybė prisidėjo prie didesnio pelningumo ir rinkos vertės. Priešingai, aplinkosaugos veiksniai nei prieš, nei po pandemijos neturėjo reikšmingo poveikio finansiniams rodikliams, kas gali būti siejama su silpnesniu aplinkosaugos reguliavimu tirtose šalyse.

Galiausiai, Gillan‘as ir kt. pateikia platesnę ESG ir finansinių rezultatų sąsajos analizę [20]. Autoriai atliko literatūros analizę, apimančią ESG įtaką pelningumui, įmonės vertei, kapitalo kainai ir rizikai. Gillan‘as ir kt. pabrėžia, kad ESG poveikis yra nevienareikšmis ir labai priklauso nuo tyrimo metodologijos, duomenų šaltinių bei reguliacinės aplinkos. Apibendrinant šį tyrimą galima daryti išvadą, kad nors ir ESG poveikis nėra vienareikšmis, tačiau dažniausiai nustatomas reikšmingas ryšys su įmonių verte, pelningumu ar rizika. Siekiant palyginti ir susisteminti skirtingų tyrimų tikslą, metodologiją ir rezultatus, pateikiama apibendrinanti aprašytų tyrimų lentelė (1 lentelė).

**1 lentelė.** ESG įtakos įmonių finansiniams rezultatams tyrimų palyginimas

<b>Autoriai, metai</b>	<b>Tikslas</b>	<b>Tikslo kintamasis</b>	<b>Rezultatai</b>
Khan ir kt. , 2016 [7]	Nustatyti ar investicijos į tvarumą didina įmonės vertę	Būsima akcijų grąža ( $\alpha$ )	Įmonės su aukštu materialu ESG turi didesnę būsimą akcijų grąžą, o nematerialus ESG neturi statistiškai reikšmingo poveikio būsimai akcijų grąžai.
Ng ir Rezaee, 2015 [15]	Nustatyti, kaip ESG ir ECON veikia įmonės nuosavo kapitalo kainą	Nuosavo kapitalo kaina	Aukštas ESG mažina nuosavo kapitalo kainą, ne visi ESG komponentai vienodai svarbūs, E ir G turi stiprius efektus.
Puri, 2022 [16]	Ištirti, kaip ESG veikia ROA, ROE ir įmonės rinkos vertę. Nustatyti ar laisvi finansiniai resursai (angl. financial slack) keičia šiuos ryšius.	ROE ROA Tobin‘o Q	ESG didina įmonės veiklos kaštus, todėl pastebėtas neigiamas poveikis pelningumui (mažesnis ROA ir ROE). ESG turi teigiamą poveikį įmonės rinkos vertei.
Xu ir Zhu, 2024 [17]	Ištirti, kaip ESG veikia įmonės finansinius rezultatus	ROA	ESG turi teigiamą poveikį ROA, taip pat skatina inovaciją, o jos atitinkamai didina pelningumą. Finansiniai apribojimai (pinigų trūkumas) silpnina ESG efektą.
Nazarova ir kt., 2023 [18]	Nustatyti, ar privalomas ESG atskleidimas skiriasi nuo savanoriško pagal poveikį įmonės finansiniams rezultatams	ROA ROE Tobin‘o Q	Po įvesto privalomo ESG teikimo įmonių ROA sumažėjo, tačiau teikiant savanoriškai neigiamo efekto nėra. Autoriai pateikia išvadą, kad ne pats ESG, o jo įgyvendinimo būdas (privalomas ar savanoriškas) yra svarbus
Azizah‘as ir Haron‘as, 2025 [19]	Ištirti, kaip ESG veikia įmonių finansinius rezultatus prieš ir po COVID-19 pandemijos	ROA ROE Tobin‘o Q	Prieš pandemiją yra teigiamas S ir G poveikis finansiniams rezultatams, o E nereikšmingas. Po pandemijos S poveikis tampa neigiamas, G susilpnėja, tačiau bendras ESG išlieka teigiamas. ESG poveikis nėra pastovus, jis

			priklauso nuo ekonominių sąlygų, krizės laikotarpio.
Gillan'as ir kt., 2021 [20]	Nustatyti, kaip ESG susijęs su įmonės finansiniais rezultatais, verte ir rizika	įmonės vertė ROA Akcijų grąža Kapitalo kaina rizika	Tyrimo metodas – literatūros apžvalga. Autoriai palyginamosios analizės būdų nustatė, kad ESG poveikis įmonių finansiniams rezultatams yra įvairus, labai priklausantis, nuo šalies, industrijos reguliavimo. ESG dažnai mažina įmonių finansinę riziką, gerina stabilumą

Apibendrinant nagrinėtus tyrimus, galima teigti, kad ESG poveikis įmonių finansiniams rezultatams yra kompleksinis ir priklauso nuo įvairių veiksnių. Nors atskirų empirinių tyrimų rezultatai nėra vienareikšmiai, bendrą išvadą padeda formuluoti Friede ir kt. atlikta meta-analizė, apimanti daugiau nei 2000 empirinių tyrimų [21]. Joje nustatyta, kad apie 90% tyrimų nustato teigiamą arba neutralų ESG ir finansinių rezultatų ryšį, o neigiamas poveikis pasitaiko retai. Vis dėlto ši išvada neturėtų būti interpretuojama, kaip universali taisyklė, kadangi analizuoti tyrimai rodo, jog ESG poveikis reikšmingai priklauso nuo įmonės finansinių išteklių, reguliacinės aplinkos ir taikomų analizės metodų. Kai kuriais atvejais ESG atsakomybių išpildymas gali pareikalauti papildomų išlaidų, taip mažinant pelningumą, tačiau ilgalaikėje perspektyvoje tai prisideda prie rizikos mažinimo, kapitalo kainos mažėjimo ir vertės kūrimo. Remiantis išanalizuotais tyrimais, galima daryti išvadą, kad ESG nėra savaime įmonės vertę kuriantis veiksnys, tačiau jis gali tapti svarbia strategine priemone, stiprinančia įmonės finansinį tvarumą ir konkurencingumą.

### 1.2.1. ESG įtaka prognozuojant įmonių finansinius rodiklius

ESG veiksnių įtraukimas į finansinių rodiklių prognozavimą pastaraisiais metais tapo svarbia tyrimų kryptimi, tačiau empirinių tyrimų rezultatai rodo, kad ESG įtraukimas skirtingai veikia prognozavimo tikslumą. Dinca ir kt. tyrimas [22] analizavo 2548 įmones iš 98 šalių ir vertino, ar ESG rodikliai pagerina finansinių prognozių tikslumą taikant įvairius mašininio mokymosi modelius, tokius kaip ARIMA, XGBoost. Tyrimo rezultatai parodė, kad aukštesni ESG rodikliai ne visada prisideda prie geresnio prognozavimo tikslumo. Teigiamas jų poveikis nustatytas tik tam tikruose sektoriuose, pavyzdžiui, verslo paslaugų srityje. Kita vertus, šis tyrimas taip pat pabrėžia, kad ESG gali turėti netiesioginį poveikį prognozavimui per įmonės finansinės rizikos mažinimą ir veiklos skaidrumą. Įmonės, turinčios aukštesnius ESG rodiklius, dažnai pasižymi stabilesniais pinigų srautais, skaidresne ir aiškiai apibrėžta įmonės veikla ir strategija, o tai, autorių nuomone, turėtų savaime indikuoti geresnes finansines prognozes.

Terzioglu ir kt. tyrimas [23] analizavo 427 Stambulo vertybinių popierių biržoje kotiruojamas įmones ir siekė nustatyti, ar tvarumo ir valdymo rodikliai gali padėti prognozuoti nuosavo kapitalo grąžą (angl. *return on equity* - ROE). Tyrimo rezultatai parodė, kad ESG rodikliai labai mažai pagerina klasifikavimo tikslumą. Autorių sukurti klasifikavimo modeliai su aplinkosauginiais, socialiniais ir valdymens kintamaisiais tik labai nežymiai lenkė atsitiktinį spėjimą. Tyrimo metu nustatyta, kad tokie kintamieji kaip tvarumo ataskaitų buvimas ar valdymo struktūros elementai buvo identifikuoti kaip svarbūs, tačiau jų poveikis prognozavimo tikslumui – nedidelis. Tyrimuose [22, 23] išryškėjo ESG duomenų kokybės ir standartizacijos problema. ESG rodiklių reikšmės dažnai priklauso nuo šaltinio, kuris pateikė tyrimuose naudojamus ESG duomenis. Skirtingi šaltiniai (agentūros) remiasi skirtingomis metodologijomis apdorojant ir agreguojant ESG rodiklius, todėl tyrimų rezultatų palyginamumas yra ribotas. Ši ESG duomenų problema apsunkina jų integraciją į prognozavimo modelius ir gali sumažinti jų praktinę vertę, pritaikomumą.

ESG svarbą, prognozuojant įmonių pelningumą, savo tyrime [24] vertino D'Amato ir kt. Autoriai analizavo apie 400 Europos įmonių iš EuroStoxx 600 indekso ir vertino ESG rodiklių poveikį pelningumui (EBIT). Tyrime naudoti mašininio mokymosi modeliai, tokie kaip sprendimų medžiai, atsitiktiniai miškai ir gradientinis stiprinimas (angl. *gradient boosting*). Rezultatai rodo, kad ESG rodiklis yra reikšmingas pelningumo prognozavimo kintamasis (pagal kintamųjų svarbos diagramą „ESG.Score“ yra trečioje vietoje), o modeliai pasiekė aukštą tikslumo lygį (vidutinė  $R^2$  reikšmė taip modelių yra 0,83). Šio tyrimo rezultatai rodo, kad mašininio mokymosi metodai gali efektyviai išnaudoti ESG duomenis įmonių finansinių rodiklių prognozavime. Talas'as ir kt. tyrime [25] siekė nustatyti ar įmonių ESG atkleidimas pagerina ROE prognozavimą. Autoriai nenaudojo įprastų ESG srities kintamųjų, ar agreguotų E, S, G balų, jie orientavosi į tai ar įmonė teikia ESG atskaitas ir kokia informacija yra jose. Prognozavimui buvo pasirinkti mašininio mokymosi modeliai: atsitiktiniai miškai, XGBoost, LightGBM, CatBoost. Tyrimo rezultatai autoriams leido daryti išvadą, kad ESG pagerina ROE prognozę, tačiau pažymi, kad ESG nepakeičia finansinių rodiklių, bet prideda papildomos informacijos.

Svarbus finansinių rodiklių prognozavimo tyrimų aspektas yra paaiškinamumo (interpretavimo) metodų taikymas. Vienas tokių metodų yra SHAP, kuris leidžia nustatyti, kaip konkretūs ESG veiksniai prisideda prie prognozės. Terzioglu ir kt. [23] būtent pasinaudoję SHAP metodu nustatė, kad tokie kintamieji kaip tvarumo atskaitos ar valdymo struktūros elementai yra tarp svarbiausių pagal SHAP reikšmes, tačiau jų poveikis prognozei vis vien nedidelis.

Apibendrinant tyrimus galima teigti, kad mašininio mokymosi metodai suteikia pažangias galimybes analizuoti ESG poveikį finansiniams rodikliams, tačiau jų efektyvumas priklauso nuo duomenų kokybės ir konteksto. Kai kurie tyrimai rodo teigiamą ESG poveikį prognozavimo rezultatui, tačiau yra tokių, kurie pabrėžia ribotą šių duomenų naudą, todėl ESG turėtų būti naudojamas kaip papildomas, o ne pagrindinis prognozavimo veiksnys.

### **1.3. Finansinės rizikos prognozavimas**

Finansinės rizikos samprata literatūroje nėra vienareikšmė, nes skirtingi autoriai ją interpretuoja skirtingais aspektais ir taiko įvairiuose kontekstuose. Dažniausiai finansinė rizika siejama su galimybe patirti nuostolių dėl neapibrėžtų finansinių rezultatų, rinkos svyravimų, kredito įsipareigojimų nevykdymo ar likvidumo problemų. Finansinė rizika taip pat skirstoma į sisteminę ir nesisteminę: sisteminė rizika kyla dėl bendrų rinkos pokyčių ir negali būti eliminuojama, o nesisteminė rizika susijusi su konkrečios įmonės veikla ir gali būti sumažinama [26, 27]. Šis skirstymas yra itin svarbus finansinės rizikos vertinime, kadangi leidžia identifikuoti, kokia rizikos dalis yra neišvengiama.

Modigliani ir Miller'io kapitalo struktūros teorija [28] yra svarbus teorinis pagrindas finansinės rizikos vertinime, kadangi paaiškina ryšį tarp įmonės finansavimo struktūros ir rizikos lygio. Pradinė teorijos versija teigia, kad tobuloje (be mokesčių, išlaidų ir nuostolių) rinkoje kapitalo struktūra neturi įtakos įmonės vertei [28]. Šį teiginį paneigė patys autoriai, išleidę pradinės teorijos papildymus [29]. Nuo tobulos rinkos buvo pereita prie realesnės finansinės aplinkos – įtraukti mokesčiai. Autoriai parodė, kad skolinto kapitalo naudojimas gali didinti įmonės vertę, nes palūkanos už skolą yra pripažįstamos sąnaudomis ir mažina apmokestinamąjį pelną. Nors tobuloje rinkoje kapitalo struktūra nekeičia įmonės vertės, bet finansinė rizika tiesiogiai kyla dėl didesnio įsiskolinimo [30]. Dėl šios priežasties susiformavo kompromiso (angl. *Trade-off*) teorija, teigianti, kad įmonė turi siekti

optimalios kapitalo struktūros, kurioje būtų subalansuota skolinto kapitalo teikiama mokestinė nauda ir didėjantys bankroto kaštai – papildomi kaštai susiję su finansiniais sunkumais, pavyzdžiui, patiriami nuostoliai dėl įmonės reputacijos mažėjimo, teisinės išlaidos. Ši teorija tiesiogiai paaiškina, kaip skola susijusi su finansine rizika [30, 31]:

- kuo didesnė skola, tuo didesnė finansinių sunkumų rizika,
- rizika didėja ne tiesiškai, o eksponentiškai, kai įmonė artėja prie per didelio įsiskolinimo,
- finansinė rizika yra ne tik skolos dydis, bet ir įmonės gebėjimas generuoti stabilius pinigų srautus.

Damodaran'as [32] finansinę riziką sieja ne tik su nemokumu, bankroto tikimybe, bet ir su įmonės gebėjimu palaikyti stabilius pinigų srautus, veiklos nepastovumu bei kapitalo struktūra. Toks požiūris leidžia suvokti riziką kaip daugiamatę sąvoką, apimančią tiek veiklos, tiek finansinę bei rinkos riziką ir išplečia tradicinį finansinės rizikos supratimą, sudarydamas prielaidas įtraukti papildomus veiksnius, tokius kaip ESG rodikliai, kurie galėtų turėti reikšmingą įtaką vertinant ir prognozuojant įmonių finansinę riziką

Finansinės rizikos vertinimas ir prognozavimas yra vienas iš įmonių finansų valdymo uždavinių, kadangi leidžia anksčiau identifikuoti galimus mokumo sutrikimus, finansinį nestabilumą bei veiklos tęstinumo grėsmes. Finansinė rizika apima bankrotą, finansinius sunkumus, kredito riziką, įsipareigojimų nevykdymo tikimybę, kapitalo struktūros silpnėjimą bei likvidumo problemas. Finansinės rizikos nustatymas leidžia įmonėms taikyti prevencines priemones, mažinti galimus nuostolius ir stiprinti finansinį atsparumą [33]. Finansinė rizika dažnai vystosi palaipsniui, todėl pirmieji signalai gali būti pastebimi gerokai anksčiau nei formaliai paskelbiamas bankrotas, todėl vis daugiau dėmesio skiriama ne tik bankroto prognozavimui, bet ir ankstyvajam finansinių sunkumų (angl. *financial distress*) identifikavimui.

Finansinė rizika ir finansiniai sunkumai yra glaudžiai susiję priežasties–pasekmės ryšiu. Finansinė rizika yra potenciali galimybė patirti nuostolių, o finansiniai sunkumai – materializuota rizikos pasekmė, kai įmonė realiai susiduria su sunkumais vykdydama finansinius įsipareigojimus ar valdydama pinigų srautus. Didelė finansinė rizika, pavyzdžiui, didelis finansinis svertas ar nepastovūs pinigų srautai, didina finansinių sunkumų tikimybę. Jei finansiniai sunkumai išskyla, jie gali dar labiau paaštrinti riziką, nes įmonės dažnai priverstos imtis brangaus ar rizikingo finansavimo, kad išlaikytų veiklą. Todėl efektyvus finansinės rizikos valdymas tiesiogiai mažina finansinių sunkumų tikimybę ir yra esminė priemonė užtikrinant įmonės finansinį stabilumą.

Finansiniai sunkumai dažnai pasireiškia mažėjančiu pelningumu, augančiu finansiniu svertu, didėjančiais įsipareigojimais bei prastėjančiais pinigų srautais. Laiku identifikavus šiuos požymius galima išvengti finansinių nesėkmių ir sumažinti bankroto riziką [34]. Namaki's ir kt. [35] pastebi, kad finansinės rizikos vertinimas pereina nuo apskaitinių ir finansinių rodiklių analizės prie sudėtingesnių ankstyvo perspėjimo sistemų (angl. *Early Warning Systems*), integruojančių, tiek finansinius, tiek nefinansinius veiksnius, įskaitant ir ESG rodiklius. Taigi, finansinės rizikos prognozavimas yra kompleksinis procesas, kuriame vertinami ne tik įstoriniai finansiniai rezultatai, bet ir platesni tvarumo ir ilgalaikio veiklos stabilumo aspektai.

Atsižvelgiant į aprašytas teorijas, finansinės rizikos apibrėžimus ir jos vertinimo būdus, šiame darbe finansinė rizika bus siejama su įmonės tikimybe patirti finansinius sunkumus. Remiantis Modigliani ir Miller'io kapitalo struktūros teorija, kompromiso teorija ir Damodaran'o požiūriu, finansiniai

sunkumai apibrėžiami kaip įmonės negebėjimas generuoti stabilų pinigų srautų, negebėjimas padengti skolų ir neigiamas veiklos pelnas. Tikslios finansinių sunkumų vertinimo formulės yra plačiau pristatomos tyrimo metodų skyriuje.

### 1.3.1. Mašininis mokymasis prognozuojant finansinę riziką

Įmonių bankroto prognozavimas per pastaruosius penkis dešimtmečius reikšmingai išsiplėtė ir įgijo didelę svarbą finansinės analizės srityje, ypač po 2008 metų pasaulinės finansų krizės. Mokslininkės Shi ir Li atliko literatūros analizę [36], kurioje apžvelgė 496 mokslinius straipsnius šioje srityje. Autorės pažymi, kad nėra vieningos verslo nesėkmės sampratos, tačiau dažniausiai ji apima finansinius sunkumus, nemokumą, įsipareigojimų nevykdymą ir bankrotą. Straipsnyje išskiriamos dvi prognozavimo metodų grupės: klasikiniai statistiniai modeliai ir dirbtinio intelekto bei mašininio mokymosi metodai. Analizės rezultatai parodė, kad logistinės regresijos modelis išlieka plačiausiai taikomas dėl savo metodologinio paprastumo ir interpretavimo galimybių, o neuroniniai tinklai bei kiti mašininio mokymosi metodai pasižymi didesniu prognozavimo tikslumu, nes gali modeliuoti sudėtingus netiesinius ryšius tarp kintamųjų. Autorės taip pat pastebi, kad įmonių bankroto prognozavimo srityje vis dažniau yra taikomi inovatyvūs mašininio mokymosi metodai, o tai leidžia daryti prielaidą, kad ateityje šie metodai gali tapti dominuojančiais šioje srityje.

Rahman'as ir kt. atliko tyrimą [37], kuriame analizavo Kinijos biržoje kotiruojamų statybos sektoriaus įmonių finansinius sunkumus. Tyrime buvo naudoti trys mašininio mokymosi modeliai: AdaBoost, CUSBoost ir CART (klasifikavimo ir regresijos medžių algoritmas), taip pat keturi tradiciniai Z-score modeliai: Altman, Sprigate, Sorins/Voronova ir Ng. Tikslu kintamasis „finansiniai sunkumai“ buvo klasifikuojamas į dvi grupes – „patiria sunkumus“ ir „nepatiria sunkumų“. Tikslu kintamojo reikšmė buvo nustatoma remiantis finansiniais rodikliais, tokiais kaip, veiklos nuostoliai, mažėjantis mokumas, neigiami pinigų srautai bei ribinės Z-score modelių reikšmės. Modelių tikslumas vertintas naudojant AUC ir AUPR rodiklius. Tyrimo rezultatai parodė, kad visi mašininio mokymosi modeliai pranoko tradicinius metodus, o didžiausią tikslumą pasiekė CUSBoost algoritmas, kuris geriausiai identifikavo finansinių sunkumų turinčias įmones. Autoriai pateikė išvadą, kad mašininio mokymosi modeliai yra pranašesni nei jų tirti klasikiniai metodai.

Lokanan'as ir Ramzan'as finansinės rizikos prognozavimo tyrime [38] naudojo sprendimų medžio (angl. *Decision Tree*), atsitiktinių miškų (angl. *Random Forest*), atraminių vektorių (angl. *Support Vector Machine*) ir dirbtinių neuroninių tinklų (angl. *Artificial Neural Networks*) modelius. Reikšmingų kintamųjų atrankai autoriai pasitelkė Recursive Feature Elimination with Cross-Validation metodą. Rezultatai parodė, kad svarbiausi kintamieji prognozuojant finansinius sunkumus buvo pajamos, dividendai, pinigų santykis su trumpalaikiais įsipareigojimais ir pelno marža. Geriausią prognozavimo tikslumą pasiekė dirbtinių neuroninių tinklų modelis, kurio klasifikavimo tikslumas siekė 98%. Dirbtinius neuroninius tinklus taip pats panaudojo ir Wu'as ir kt. finansinės rizikos prognozavimo tyrime [39]. Autoriai prognozavo bankroto riziką ir rezultatus lygino su tradiciniu Altman'o Z-score modeliu. Rezultatai atkleidė, kad dirbtinių neuroninių tinklų modelis pasiekė geresnę prognozavimo tikslumą nei tradicinis Altman'o modelis. Autoriai pabrėžė, kad mašininio mokymosi metodai geriau identifikavo sudėtingus netiesinius ryšius tarp finansinių rodiklių.

Yang ir kt. prognozavo finansinius sunkumus ir pasiūlė metodologinę sistemą tokioms prognozavimo užduotims spręsti [40]. Autorių tikslas buvo ne tik sukurti tikslų prognozavimo modelį, bet ir

identifikuoti svarbiausius finansinės rizikos veiksnius. Tyrimą sudarė dvi dalys: prognozavimas XGBoost mašininio mokymosi metodu ir rezultatų interpretavimas pasitelkiant SHAP (angl. *Shapley Additive Explanations*) metodą. Autoriai nustatė, kad XGBoost lenkia tradicinę logistinę regresijos modelį savo prognozavimo tikslumu, o tai dar kartą patvirtina mašininio mokymosi metodu pranašumą. Taip pat, didžiausią įtaką finansinių sunkumų tikimybei turi tokie veiksniai kaip finansinis svertas, įmonės vertės bei pelningumo rodikliai. Remiantis šiuo ir anksčiau aptartais straipsniais, galima teigti, kad mašininio mokymosi metodai, derinami su SHAP interpretavimo sistema, suteikia ne tik aukštą prognozavimo tikslumą, bet ir užtikrina modelio praktinį pritaikomumą, leidžiantį interpretuoti rezultatus.

Viena pagrindinių finansinės rizikos prognozavimo problemų yra nesubalansuoti duomenys, kai finansinių sunkumų patiriančių įmonių yra žymiai mažiau nei finansiškai stabilių. Duomenų disbalansas neigiamai veikia modelių tikslumą ir jų pritaikomumą. Straipsnyje [41] pažymima, kad modeliai, apmokyti naudojant nesubalansuotus duomenis, gali pasiekti aukštą bendrą tikslumą, tačiau nesugebėti tinkamai identifikuoti finansinių sunkumų turinčių įmonių. Duomenų disbalansas daro tiesioginę įtaką skirtingų tipų klaidų skaičiui. Didesnė bankrutuojančių įmonių dalis imtyje mažina antros rūšies klaidą (neatpažintus bankrotus), bet didina pirmos rūšies klaidą, ir atvirkščiai. Šiai problemai spręsti, literatūroje siūlomi įvairūs sprendimai, įskaitant duomenų lygmens metodus, tokius kaip perrinkimas (angl. *oversampling*) ir sumažinimas (angl. *undersampling*) bei modelio lygmens korekcijas. Vienas dažniausiai naudojamų metodų yra sprendimų slenksčio (angl. *cut-off threshold*) koregavimas, kai vietoje standartinės 0,5 tikimybės ribos parenkama kita reikšmė, leidžianti geriau subalansuoti pirmos ir antros rūšies klaidas. Taip pat dažnai taikomas kaštams jautrus mokymasis (angl. *cut-sensitive learning*), kai klasifikavimo modelio mokymo metu skirtingiems klaidų tipams priskiriami nevienodi svoriai, pavyzdžiui, didesnė bauda už nepastebėtą finansinio sunkumo atvejį nei už klaidingą įmonės priskyrimą prie finansinių sunkumų turinčių įmonių. Liu ir kt. straipsnyje [42] nagrinėjamas finansinių sunkumų prognozavimas, pagrindinį dėmesį skiriant ansamblių metodų (angl. *ensemble learning*) taikymui. Tyrime buvo naudojami įvairūs mašininio mokymosi metodai, įskaitant XGBoost, AdaBoost, atsitiktinius miškus, atraminius vektorius, logistinę regresiją ir EasyEnsemble. Kintamųjų atrankai autoriai naudojo požymių svarbos vertinimo permutacijos būdu (angl. *permutation importance*) ir rekursinės požymių atrankos (angl. *recursive feature elimination*) metodus. Šio tyrimo duomenyse yra ryškus priklausomo kintamojo klasių disbalansas – finansinių sunkumų patiriančių įmonių skaičius yra daugiau nei 30 kartų mažesnis nei finansiškai stabilių. Pradiniai prognozavimo rezultatai atskleidė, kad tradiciniai klasifikavimo modeliai pasižymi aukštu tikslumu, virš 90%, tačiau visiškai nesugeba identifikuoti finansinių sunkumų turinčių įmonių (labai žemas jautrumo rodiklis). Geriausias prognozavimo rezultatus pavyko pasiekti naudojant EasyEnsemble ir kitus ansamblio mokymosi metodus, pavyzdžiui, XGBoost. Įvykdžius kintamųjų atranką, jų skaičių pavyko sumažinti perpus. Sun ir kt. atliko tyrimą [43], kuriame taip pat didžiausią dėmesį skyrė duomenų disbalanso problemai spręsti. Šios problemos sprendimui, autoriai pasiūlė hibridinį metodą, integruojantį sintetinį mažumos klasės pavyzdžių generavimą (SMOTE) su Adaboost-SVM ansambliu, papildytu laiko svorių komponentu, kuomet naujesni metai yra didesnės svarbos. Šio tyrimo rezultatai parodė, kad autorių taikomas metodas reikšmingai pagerino mažumos klasės atpažinimą ir pranoko bazinius modelius, kurie prognozavimą atlieka su nesubalansuotais duomenimis.

Duomenų disbalanso problemos sprendimą, finansinių sunkumų prognozavime, SMOTE metodu sprendė ir Nguyen'as ir kt. Tyrime [44] mašininio mokymosi metodai taikomi kartu su

interpretuojamo dirbtinio intelekto priemonėmis (angl. *explainable artificial intelligence*). Pagrindinis dėmesys skiriamas ne tik prognozavimo tikslumui, bet ir modelių sprendimo paaiškinimui, siekiant identifikuoti svarbiausius veiksnius bei suteikti praktinę vertę sprendimų priėmėjams. Tyrimo imtį sudaro 1097 įmonės iš kurių tik 76 patiria finansinius sunkumus, todėl duomenų disbalanso problemai spręsti buvo taikomas SMOTE metodas, kuris leidžia subalansuoti klases ir pagerinti finansinių sunkumų turinčių įmonių aptikimą. Prognozavimui taikomi įvairūs klasifikavimo metodai, tiek klasikiniai, tiek mašininio mokymosi. Tyrimo rezultatai rodo, kad neuroniniai tinklai ir XGBoost pasižymi aukščiausiu prognozavimo tikslumu. Modelių rezultatų interpretacijai buvo taikomi du metodai: SHAP ir DiCE. SHAP panaudotas kintamųjų svarbai nustatyti, įvertinant tiek bendrą, tiek individualų kiekvieno veiksnio poveikį prognozei. DiCE metodas buvo panaudotas rekomendacijoms, jis nustatė minimalius įvesties kintamųjų pokyčius, kurie pakeistų modelio prognozuojamą rezultatą. Šio metodo integravimas yra reali vertė sprendimo priėmėjams, kadangi konstatuojamas ne tik finansinių sunkumų faktas, bet ir pateikiamos rekomendacijos, ką vertėtų pakeisti. Nguyen'o ir kt. atliktas tyrimas parodo, kad mašininio mokymosi modeliai gali būti sėkmingai derinami su interpretavimo metodais, užtikrinant tiek aukštą prognozavimo tikslumą, tiek sprendimų paaiškinamumą. Toks tyrimo modelis suteikia galimybę ne tik prognozuoti finansinius sunkumus, bet ir geriau suprasti jų priežastis bei galimus jų prevencijos būdus.

Gao'o ir kt. tyrime [45] siūlomas naujas finansinių sunkumų prognozavimo modelis, skirtas spręsti pagrindines problemas: didelės apimties duomenys, klasių disbalansas ir hiperparametrų optimizavimas. Autorių pasiūlyto modelio struktūra susideda iš trijų pagrindinių etapų. Pirmiausia atliekama požymių atranka, taikant porinį skirtingų laikotarpių duomenų palyginimą mRMR metodą. Antrajame etape naudojamas daugialypis savimokos ansamblinis mokymasis (MHSPE), kuriame derinami keli baziniai klasifikatoriai, tokie kaip, LightGBM, XGBoost ir GBDT. Toks metodas sprendžia klasių disbalanso problemą. Ansamblinio mokymosi metodų pranašumą akcentavo ir Liu ir kt. [42]. Trečiajame etape modelio veikimas optimizuojamas naudojant dalelių spiečiaus optimizavimo (PSO) algoritimą, kuris leidžia parinkti optimalias hiperparametrų reikšmes ir pagerinti prognozavimo tikslumą. Šio tyrimo rezultatai rodo, kad pasiūlytas modelis savo prognozavimo tikslumu lenkia visus kitus.

Apibendrinant, finansinės rizikos prognozavimas analizuojamas ne tik per bankroto tikimybės įvertinimą, bet per platesnę finansinių sunkumų sampratą, apimančią likvidumo problemas, išpareigojimų nevykdymo riziką, kapitalo struktūros silpnėjimą. Tradiciniai modeliai finansinės rizikos vertinimui, tokie kaip Altman'o Z-score, Ohlson'o O-score ir Zmijewski'o, ilgą laiką buvo pagrindiniai finansinės rizikos vertinimo modeliai, tačiau pastarojo dešimtmečio tyrimai rodo, kad jų prognozavimo tikslumas dažnu atveju yra mažesnis nei mašininio mokymosi modelių. Analizuoti tyrimai rodo, kad tradiciniai metodai nėra pakankami, kuomet yra duomenų disbalanso problema. Geriausi rezultatai pasiekiami derinant SMOTE metodą su ansamblinio mokymosi algoritmais. Taip pat svarbi tyrimo dalis turėtų būti ir interpretavimas, pavyzdžiui SHAP metodo panaudojimas.

### **1.3.2. ESG įtraukimas į mašininio mokymosi modelius**

ESG rodiklių integravimas į mašininio mokymosi modelius yra sudėtingas procesas, susijęs ne tik su papildomų kintamųjų įtraukimu, bet ir su reikšmingais duomenų struktūros, kokybės bei interpretacijos iššūkiais. Priešingai nei įprasti finansiniai rodikliai, ESG duomenys pasižymi didele subjektyvumu ir nevienodumu problema, nes skirtingos, ESG duomenis apdorojančios, agentūros

naudoja nevienodas metodikas ir vertinimo kriterijus, pavyzdžiui, ta pati įmonė gali gauti skirtingus ESG įvertinimus priklausomai nuo to, kuri agentūra ją vertina [46].

Vienas pagrindinių iššūkių integruojant ESG duomenis į mašininio mokymosi modelius yra šių duomenų nevienalytiškumas. ESG informacija gali būti pateikiama tiek kiekybine (emisijų kiekis, darbuotojų kaita), tiek kokybine (reputacijos aspektai, valdymo praktika) forma, dėl to integruojant ESG rodiklius į mašininio mokymosi modelius būtina atlikti duomenų transformacijas [46, 47]. Kitas svarbus aspektas yra duomenų trūkumas ir jų pasiskirstymas. ESG rodikliai yra sąlyginai nauji, ne visame pasaulyje jų rinkimas ir teikimas yra privalomas ir reglamentuotas, todėl susiduriama su trūkstamų reikšmių problema. Šiai problemai spręsti gali būti taikomi įvairūs duomenų užpildymo metodai, tačiau netinkamas ar perteklinis duomenų apdorojimas gali iškreipti mašininio mokymosi modelių rezultatus ir sumažinti jų patikimumą [7]. Dar viena ESG rodiklių problema – multikolinearumas. ESG rodikliai dažnai yra stipriai tarpusavyje koreliuoti, todėl labai svarbu prieš pradėdant kurti prognozavimo modelį atlikti kintamųjų atranką. Pagrindinių komponentų analizė gali būti vienas iš kintamųjų atrankos metodų [46].

ESG rodikliai dažnai pateikiami ne tik kaip atskiri rodikliai iš aplinkosaugos, socialinės atsakomybės ir valdysenos sričių, bet kaip agreguoti, atskiri E, S ir G įverčiai, kuriuos skaičiuoja agentūros, tokios kaip MSCI, Sustainalytics ar Refinitiv. Šie agreguoti rodikliai apjungia kelias dešimtis individualių kintamųjų į vieną bendrą indeksą, kuris turėtų atspindėti bendrą kiekvienos srities lygį. Šių agreguotų rodiklių naudojimas palengvina duomenų paruošimą ir sumažina kintamųjų skaičių, bet mokslinėje literatūroje pabrėžiama, kad vis dėlto agreguoti ESG įverčiai skiriasi tarp agentūrų, dėl nevienodų skaičiavimo metodikų, svorių ir vertinimo kriterijų [46]. Tyrime [48] nustatyta, kad ESG rodiklių koreliacija tarp skirtingų agentūrų yra nedidelė ir sutampa tik apie 15% reikšmių. Skirtingos ESG apibrėžtys lemia ir skirtingus požymių rinkinius kiekvienai ESG sričiai, kurie daro tiesioginę įtaką modelių mokymuisi ir rezultatams. Toks skirtingas agregavimas gali lemti informacijos praradimą, nes skirtingi rodikliai iš tos pačios srities, pavyzdžiui, aplinkosaugos, gali turėti nevienodą ar net priešingą poveikį prognozuojamam rezultatui. Taigi, jei mašininio mokymosi modeliuose pasirenkama naudoti agreguotus ESG rodiklius, itin svarbu užtikrinti jų nuoseklumą, naudoti duomenis iš to paties šaltinio ar agentūros.

### **1.3.3. Finansinės rizikos prognozavimas įtraukiant ESG rodiklius**

Tradiciškai finansinė rizika buvo vertinama remiantis įmonių finansiniais rodikliais, tačiau naujausi tyrimai rodo, kad ESG gali suteikti papildomos vertės prognozuojant įmonių finansinius sunkumus, kredito riziką ar įsipareigojimų nevykdymo tikimybę.

Vienas iš empirinių tyrimų yra Song ir kt. [49], kuriame buvo analizuojamas ESG informacijos panaudojimas įmonių finansinių sunkumų prognozei energetikos sektoriuje. Tyrime buvo naudojami dviejų rūšių modeliai: pavieniai, pavyzdžiui, logistinė regresija ir sprendimų medis, ir ansamblinio mokymosi modeliai, pavyzdžiui, atsitiktiniai miškai ir XGBoost, o ESG socialinio aspekto informacija integruojama tiek per kiekybinius rodiklius, tiek per tekstinę analizę (metinių ataskaitų turinį). Autoriai taip pat sprendė ir duomenų disbalanso problemą, jos sprendimui panaudojo SMOTE metodą. Tyrimo rezultatai parodė, kad geriausiai veikia ansamblinio mokymosi modeliai, tokie kaip GBDT, atsitiktiniai miškai ir XGBoost, tačiau autoriai ypač išskiria GBDT modelio tikslumą. Taip pat nustatyta, kad tekstiniai rodikliai iš metinių ataskaitų yra itin svarbūs, o socialinės atsakomybės rodikliai pagerina prognozavimo tikslumą. Nors šis tyrimas parodė, kad socialinės atsakomybės

rodikliai gali pagerinti finansinės rizikos prognozavimą naudojant mašininio mokymosi metodus, tačiau jis apsiriboja tik socialinio aspekto integravimu ir neapima pilno aplinkosaugos, socialinių ir valdysenos veiksnių integravimo.

Šią tyrimų kryptį papildė ir Citterio darbas [50], kuriame jau nagrinėjamas pilnas ESG veiksnių vaidmuo prognozuojant finansinius sunkumus bankų sektoriuje. Autorius analizuoja 362 JAV ir Europos bankų imtį ir vertina, kaip bendras ESG įvertis bei atskiros jo dedamosios E, S ir G veikia finansinės rizikos lygį. Tyrimui pasirinkti tiek tradiciniai statistiniai metodai, tiek mašininio mokymosi ir ansamblinio mokymosi metodai. Rezultatai rodo, kad ESG rodikliai turi statistiškai reikšmingą neigiamą ryšį su bankų sunkumų tikimybe – aukštesni ESG įverčiai siejami su mažesne rizika. Autorius ypač pabrėžia socialinio komponento svarbą, kadangi jis turi stipriausią riziką mažinantį poveikį, o tuo tarpu aplinkosauginiai ir valdysenos komponentai pasižymi silpnesniu, tačiau vis tiek teigiamu efektu. Nors Citterio tyrimas analizuoja ESG veiksnių vaidmenį prognozuojant finansinius sunkumus, jo empirinis tyrimas apsiriboja tik bankų sektoriumi, kuriam būdinga specifinė veiklos logika, aukštas reguliavimo lygis ir veiklos struktūra. Dėl šios priežasties gauti rezultatai ne visada gali būti tiesiogiai pritaikomi ne finansinių įmonių kontekste, kur ESG veiksniai dažniau susiję su operacine veikla, tiekimo grandinėmis bei aplinkos poveikiu.

Bonacorsi ir kt. tyrimas [51] analizuoja ESG veiksnių ir įmonių kredito rizikos sąsajas, taikant mašininio mokymosi metodus Europoje listinguojamų įmonių imčiai. Kredito rizika šiame darbe vertinama naudojant Altman'o Z-score rodiklį, o ESG rodikliai šiame darbe yra ne agreguoti, naudojamas platus ESG rodiklių rinkinys. Poveikis kredito rizikai vertinamas naudojant regresinę analizę. Gauti rezultatai rodo, kad ESG veiksniai reikšmingai prisideda prie įmonių nemokumo tikimybės paaiškinimo ir pagerina prognozavimo modelio tikslumą, lyginant su modeliu į kurį įtraukiami tik finansiniai rodikliai. Autoriai taip pat išskyrė ir ESG rodiklius, turinčius didžiausią poveikį kredito rizikai. Aplinkosauginiai veiksniai, susiję su emisijomis, bei socialiniai aspektai, tokie kaip darbuotojų kvalifikacija ir duomenų apsauga yra išskiriami, kaip turintys didžiausią poveikį tarp ESG rodiklių rinkinio. Šiame darbe kredito rizika vertinama naudojant tradicinę regresijos modelį, o ne mašininio mokymosi metodus, todėl išlieka erdvė tolimesniems tyrimams, kurie analizuoja ESG veiksnių įtaką naudojant būtent mašininio mokymosi ir jų interpretavimo metodus.

Apibendrinant nagrinėtą literatūrą, galima daryti išvadą, kad ESG integravimas į finansinės rizikos prognozavimą yra besivystanti, tačiau dar nepakankamai išplėta tyrimu sritis. Esami tyrimai patvirtina, kad ESG informacija gali reikšmingai pagerinti finansinės rizikos vertinimą ir prognozavimo tikslumą. Tyrimuose ESG veiksniai dažnai naudojami kaip papildomi kintamieji, o jų gilesnė sąveika su finansine rizika ir interpretavimas išlieka ribotas.

#### **1.4. Darbo pagrindimas**

Finansinės rizikos prognozavimas yra viena svarbiausių finansinės analizės sričių. Tradiciškai finansinė rizika yra vertinama remiantis finansiniais rodikliais, tačiau augantis dėmesys ESG leidžia išplėsti finansinės rizikos vertinimą į jį integruojant nefinansinius rodiklius. Mokslinėje literatūroje ESG rodiklių vaidmuo finansinės rizikos kontekste yra vis labiau nagrinėjamas, tačiau trūksta tyrimų, orientuotų būtent į finansinių sunkumų prognozavimą naudojant pažangius finansinių sunkumų prognozavimo modelius, paremtus mašininio mokymosi technologijomis. Dauguma esamų tyrimų daugiausia koncentruojasi į kredito riziką, įsipareigojimų nevykdymo tikimybę ar finansinių rodiklių prognozavimą arba įtraukia tik pavienius ESG komponentus. Be to, tik dalis šių tyrimų taiko

pažangius mašininio mokymosi metodus, todėl ESG ir mašininio mokymosi derinimas finansinės rizikos prognozavime išlieka besiformuojanti tyrimų kryptis.

Atsižvelgiant į tai, galima teigti, kad egzistuoja tyrimų spraga – trūksta darbų, kurie vertintų ESG rodiklių, ypač agreguotų, panaudojimą finansinės rizikos prognozavime taikant mašininio mokymosi metodus. Šiuo darbu siekiama prisidėti prie šios srities nagrinėjimo, analizuojant, ar ESG rodikliai gali pagerinti finansinių sunkumų prognozavimą ir kokią įtaką jie turi modelių tikslumui, vertinant nefinansinių, skirtingų pasaulio šalių įmonių imtį. Darbo aktualumas grindžiamas tuo, kad finansinės rizikos vertinimas yra ypač svarbus investuotojams, kreditoriams ir pačiai įmonei, jos sprendimų priėmėjams.

## 2. Tyrimo metodai

### 2.1. Duomenys

Duomenys tyrimui surinkti iš LSEG duomenų platformos<sup>1</sup>. Pradinis duomenų rinkinys apima apie 60 tūkstančių viso pasaulio įmonių ir jų finansinių bei ESG rodiklių. Dėl išsiskiriančių ESG reglamentų skirtingose pasaulio šalyse, didelė dalis šių įmonių neteikia ESG duomenų, todėl nėra tinkamos šio darbo tyrimo imčiai. Siekiant išsaugoti kuo daugiau įmonių, neturinčių trūkstatų reikšmių duomenyse, papildomai buvo surinkti finansiniai duomenys iš Yahoo Finance puslapio<sup>2</sup>. Šiais duomenimis buvo papildytas pradinis duomenų rinkinys. Duomenų struktūra ir kintamųjų aprašymai yra pateikti 1 priede.

### 2.2. Finansinės rizikos (priklausomo kintamojo) apibrėžimas

Finansinė rizika šio darbo tyrime apibrėžiama, kaip įmonės patiriami finansiniai sunkumai (angl. financial distress), nes jie yra ankstyvesnė, jautresnė ir platesnė finansinės rizikos išraiška nei bankrotas, leidžianti identifikuoti įmonės finansinio nestabilumo požymius gerokai anksčiau, nei rizika materializuojasi į kraštutinį įvykį. Remiantis atlikta literatūros apžvalga, galima pastebėti, kad finansiniai sunkumai skirtinguose empiriniuose tyrimuose yra apibrėžiami nevienodai, pavyzdžiui, vienuose jų finansinė rizika suprantama kaip bankrotas, kituose kaip mažėjantys pinigų srautai. Atliekamo tyrimo pradiniam duomenų rinkinyje nėra jokio priklausomojo kintamojo, todėl šiame poskyryje jis bus apibrėžiamas. Šiame darbe taikomi 5 skirtingi finansinių sunkumų apibrėžimai. Visos priklausomojo kintamojo reikšmės priskiriamos remiantis tik 2025 metų duomenimis.

1. Atsižvelgiant į tai, kad tyrimo imtį sudaro įvairių sektorių (ne tik gamybos) ir įvairių šalių įmonės, priklausomojo kintamojo sukūrimui naudojamas modifikuotas Altman'o Z-score finansinių sunkumų vertinimo metodas – Altman'o Z'' [52, 53].

Altman'o Z'' apskaičiuojamas:

$$Z'' = 6,56X_1 + 3,26X_2 + 6,72X_3 + 1,05X_4, \quad (1)$$

$$\text{čia } X_1 = \frac{\text{apivartinis kapitalas}}{\text{visas turtas}}, X_2 = \frac{\text{nepaskirstytas pelnas}}{\text{visas turtas}}, X_3 = \frac{EBIT}{\text{visas turtas}}, X_4 = \frac{\text{nuosavo kapitalo vertė}}{\text{visų įsipareigojimų vertė}}$$

Įmonė laikoma patiriančia finansinius sunkumus („finansiniai sunkumai“ = 1), jei Z'' rodiklis 2025 metais yra mažesnis nei 1,1.

2. Finansiniai sunkumai apibrėžiami kaip neigiamas grynas pelnas [41, 54]. Jei grynas pelnas yra mažesnis už 0, tai įmonė laikoma patiriančia finansinius sunkumus („finansiniai sunkumai“ = 1).
3. Finansiniai sunkumai apibrėžiami, kai įmonės nesugebėjimas padengti skolą. Jei  $\frac{EBIT}{\text{palūkanų sąnaudos}}$  yra mažiau už 1 tai įmonė laikoma patiriančia finansinius sunkumus („finansiniai sunkumai“ = 1) [55, 56]. Žemas palūkanų padengimo rodiklis rodo, kad įmonė negeneruoja pakankamai pelno skolų įsipareigojimams vykdyti ir tai yra laikoma svarbiu finansinių sunkumų indikatoriumi.

<sup>1</sup> <https://www.lseg.com/en/data-analytics/products/workspace>

<sup>2</sup> <https://finance.yahoo.com>

4. Finansiniai sunkumai apibrėžiami kaip neigiamas veiklos pelnas (uždirbti pinigai prieš palūkanas ir mokesčius) [57, 58]. Įmonė laikoma patiriančia finansinius sunkumus („finansiniai sunkumai“ = 1) jei  $EBIT < 0$ .
5. Siekiant atsižvelgti į tai, kad finansiniai sunkumai yra daugialypė sąvoka finansiniai sunkumai apibrėžiami, kaip kombinuotas anksčiau paminėtų finansinių sunkumų rinkinys. Įmonė laikoma patiriančia finansinius sunkumus („finansiniai sunkumai“ = 1) jei tenkinama bent viena sąlyga:
  - $Z'' < 1,1$
  - Grynasis pelnas mažesnis už 0
  - $\frac{EBIT}{\text{palūkanų sąnaudos}} < 1$

Toks metodas leidžia apimti skirtingus finansinės rizikos aspektus ir sumažinti priklausomybę nuo vieno rodiklio, kas ir yra rekomenduojama literatūroje [58].

### 2.3. Duomenų disbalanso problema ir jos sprendimo būdai

Prognozavimo uždaviniuose, kurių tikslas yra prognozuoti dvireikšmio priklausomojo kintamojo reikšmę (0 arba 1), dažnai susiduriama su duomenų disbalanso problema, kai vienos iš reikšmių yra reikšmingai mažiau nei kitos. Ne visi mašininio mokymosi metodai yra pajėgūs susitvarkyti su taip pasiskirsčiusiais duomenimis. Modeliai tampa šališki daugumos klasei ir linkę ją visada prognozuoti, tačiau modelio gerumas gali būti apgaulingas, pavyzdžiui, prognozavimo tikslumas 98% indikuotų labai gerą ir aukštą prognostinį rezultatą, tačiau jei modelyje 98% stebinių priklauso vienai klasei ir modelis viską jai ir priskiria taip pasiekdamas 98% tikslumą, jis yra bevertis.

Duomenų disbalanso problema sprendžiama taikant duomenų transformavimo metodus, tokius kaip daugumos klasės mažinimas (angl. *undersampling*), mažumos klasės didinimas (angl. *oversampling*) arba šių metodų deriniai. Tokių metodų tikslas yra subalansuoti duomenų pasiskirstymą ir pagerinti modelių gebėjimą tiksliau prognozuoti abi klases.

Šio darbo tyrime duomenų disbalansui mažinti taikomas SMOTE (angl. *Synthetic Minority Oversampling Technique*) metodas, kuris generuoja naujus, sintetinius mažumos klasės stebėjimus remdamasis esamais mažumos klasės stebimais [59].

Trakime, kad  $\mathbf{x}$  yra vienas iš mažumos klasės stebėjimų. SMOTE metodas kiekvienam tokiam stebėjimui  $\mathbf{x}$  suranda  $k$  artimiausių kaimynų mažumos klasėje. Artimiausius kaimynus renka pagal pasirinktą artumo metriką, dažniausiai tai yra Euklido atstumas:

$$d(i, j) = \sqrt{\sum_{s=1}^p (x_{is} - x_{js})^2}, \quad (2)$$

čia  $p$  – požymių (nepriklausomų kintamųjų) skaičius,  $x_i$  –  $i$ -asis stebėjimas,  $x_j$  –  $j$ -asis stebėjimas.

Toliau metodas parenką vieną kaimyną  $\mathbf{x}^R$  ir generuoja naują stebėjimą  $\mathbf{s}$  pagal formulę [59]:

$$\mathbf{s} = \mathbf{x} + u \cdot (\mathbf{x}^R - \mathbf{x}), \quad (3)$$

čia  $u \in [0; 1]$ .

Šio metodo veikimui būtina parinkti artimiausių kaimynų  $k$  skaičių bei nurodyti kiek naujų stebėjimų sugeneruoti. Nauji stebėjimai pridedami prie duomenų rinkinio, taip padidinant mažumos klasės dalį.

Apibendrinant, SMOTE metodas leidžia išplėsti mažumos klasės stebinių skaičių, sumažinti mašininio mokymosi modelio šališkumą daugumos klasės atžvilgiu ir pagerinti mažumos klasės prognozavimo jautrumą.

## 2.4. Mašininio mokymosi metodai

Šiame skyriuje pateikiama tyrime naudotų mašininio mokymosi metodai. Modeliai finansinės rizikos prognozavimui pasirinkti remiantis atlikta literatūros apžvalga.

### 2.4.1. Logistinė regresija

Logistinė regresija yra statistinis ir kartu mašininio mokymosi metodas, skirtas modeliuoti dvireikšmio priklausomojo kintamojo tikimybę. Logistinės regresijos modelis sudaromas ne pačiam tikslo kintamajam, o jo tikimybių santykio logaritmui [60].

Logistinės regresijos modelis užrašomas taip:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n, \quad (4)$$

čia  $p$  yra tikimybė, kad priklausomas kintamasis įgyja reikšmę 1,  $x_n$  – nepriklausomi kintamieji, o  $\beta_n$  – modelio parametrai.

Iš šios logistinės regresijos išraiškos gaunama logistinės regresijos tikimybinė forma, kuri naudojama praktiniam prognozavimui:

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}. \quad (5)$$

Pažymėkime  $e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n} = z$  ir tuomet tikimybę galime užrašyti:

$$p = \frac{e^z}{1 + e^z}. \quad (6)$$

Toks logistinės regresijos modelis gali būti taikomas tik tokiu atveju, jei duomenis tenkina reikalavimus:

1. Priklausomas kintamasis yra dvireikšmis – įgyja tik 0 arba 1 reikšmes.
2. Nėra duomenų disbalanso – tarp priklausomojo kintamojo reikšmių (0 ir 1) negali vyrėti viena iš jų.
3. Nedvireikšmiai kategoriniai nepriklausomi kintamieji, turi būti perkoduoti į dvireikšmius kintamuosius.
4. Nepriklausomi kintamieji negali būti stipriai koreliuoti.

Šio darbo tyrime yra ryški klasių disbalanso problema, todėl prieš sudarant logistinės regresijos modelį buvo panaudotas SMOTE metodas.

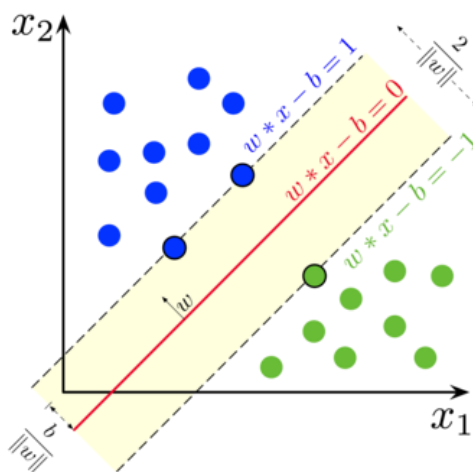
### 2.4.2. Atraminų vektorių metodas

Atraminiai vektoriai yra vienas iš dažniausiai taikomų mašininio mokymosi metodų klasifikavimo uždaviniam spręsti. Pagrindinė šio metodo idėja yra rasti tokią sprendimo ribą, kuri kuo geriau atskirtų skirtingų klasių stebėjimus. Ši riba yra vadinama hiperplokštuma.

Paprasciausiai atveju, jei duomenys yra tiesiškai atskiriami, atraminių vektorių modelis gali būti aprašomas kaip [61]:

$$f(x) = \mathbf{w}^T \cdot \mathbf{x} + b, \quad (7)$$

čia  $\mathbf{w}$  - svorių vektorius, o  $b$  – poslinkis. Sprendimas priimamas pagal tai, kurioje šios hiperplokštumos pusėje atsiduria stebėjimas. Atraminių vektorių metodo iliustracija pateikiama 1 paveiksle.



1 pav. Atraminiai vektoriai<sup>3</sup>

Didelės apimties duomenyse dažnai pasitaiko situacijos, kai tiesinė riba nėra pakankama, todėl tokiais atvejais taikomas branduolio metodas, leidžiantis duomenis perkelti į aukštesnės dimensijos erdvę, kurioje duomenys galbūt tampa geriau atskiriami. Bendruoju atveju atraminių vektorių sprendimo funkcija užrašoma kaip [62]:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b, \quad (8)$$

čia  $x_i$  – stebiny,  $y_i$  – stebinio tikroji klasė,  $x$  – stebiny, kuriam prognozuojama tikslo kintamojo reikšmė,  $K(x_i, x)$  – branduolio funkcija,  $\alpha_i$  – modelio parametrai.

Šiame tyrime buvo išbandytos kelios skirtingos branduolio funkcijos, siekiant įvertinti kuri iš jų geriausiai tinka nagrinėjamiems duomenims.

- Tiesinė branduolio funkcija[62]:

$$K(x_i, x_j) = x_i^T x_j. \quad (9)$$

Šis branduolys tinkamas kai duomenys yra tiesiškai atskiriami.

- Polinominė ( $d$ -tojo laipsnio) branduolio funkcija [62]:

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad (10)$$

čia  $d$  nusako polinomo laipsnį, todėl naudojant šią branduolio funkciją modelis gali įvertinti sudėtingesnius, netiesinius ryšius tarp kintamųjų.

- RBF (radialinės bazės) branduolio funkcija [62]:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right). \quad (11)$$

<sup>3</sup> WIKIMEDIA COMMONS. SVM margin [interaktyvus], 2025. Prieiga per internetą: [https://commons.wikimedia.org/w/index.php?title=File:SVM\\_margin.png&oldid=978793380](https://commons.wikimedia.org/w/index.php?title=File:SVM_margin.png&oldid=978793380) [žiūrėta gegužės 17, 2026].

Ji leidžia modeliuoti sudėtingas, netiesines priklausomybes ir praktikoje dažnai pasižymi gerais rezultatais.

– Sigmoidinė branduolio funkcija [62]:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r). \quad (12)$$

Ši funkcija taip pat gali modeliuoti sudėtingas, netiesines priklausomybes, tačiau ji daugiau susijusi su neuroniniais tinklais, nes šios branduolio funkcijos forma atitinka aktyvacijos funkciją neuroniniuose tinkluose.

Atraminų vektorių metodas ieško tokios sprendimo ribos, kuri būtų kiek įmanoma toliau nuo skirtingų klasių ir kartu modelis nedarytų per daug klaidų. Šio metodo tikslas yra minimizuoti:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i, \quad (13)$$

čia  $\|\omega\|$  – svorių vektoriaus ilgis,  $\xi_i$  – klaida,  $C$  – regularizacijos parametras („bauda“ už klaidas).

Svarbus atraminų vektorių modelio parametras yra  $C$ , kuris leidžia kontroliuoti modelio sudėtingumą ir padaromas klaidas. Didesnė  $C$  reikšmė skatina modelį sumažinti klaidų skaičių, tačiau tuomet atitinkamai didėja persimokymo rizika.

### 2.4.3. XGBoost metodas

XGBoost yra ansamblinio mokymosi metodas, pagrįstas gradientiniu stiprinimu (angl. *gradient boosting*). Gradientinis stiprinimas yra iteracinis metodas, kuriame kiekvienas naujai pridamas modelis, dažniausiai sprendimų medis, mokosi iš anksčiau pridėtų modelių klaidų [63]. XGBoost yra gradientinio stiprinimo metodo realizacija, papildyta regularizacija, kuri padeda sumažinti persimokymą, leidžia kontroliuoti modelio sudėtingumą ir efektyviau minimizuoti nuostolių funkciją.

Modelis sudaromas kaip kelių regresinių medžių suma [64]:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (14)$$

čia  $\hat{y}_i$  – prognozuojama reikšmė,  $f_k$  – k-asis sprendimų medis,  $\mathcal{F}$  - visų galimų medžių erdvė,  $K$  – medžių skaičius.

XGBoost modelio optimizuojama funkcija sudaryta iš dviejų dalių: nuostolių funkcijos ir regularizacijos nario:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (15)$$

čia  $l(y_i, \hat{y}_i)$  – skirtumas tarp tikrosios reikšmės  $y_i$  ir prognozės  $\hat{y}_i$ ,  $K$  – medžių skaičius,  $\Omega(f)$  – regularizacijos funkcija.

XGBoost modelyje gali būti taikoma tiek L1, tiek L2 regularizacija. Jos įtraukiamos į regularizacijos funkciją  $\Omega(f)$ . Regularizacijos funkcija padeda sumažinti persimokymo riziką, nes sudėtingesni medžiai yra papildomai „baudžiami“.

XGBoost modelis kuriamas iteracijomis, kiekvienoje jų prie jau esamos prognozės pridamas naujas medis. Todėl t-ojoje iteracijoje prognozė gali būti išreikšta taip:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (16)$$

Įstačius šią išraišką į minimizuojamą funkciją, gaunama vienos iteracijos tikslo funkcija:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t). \quad (17)$$

Ši formulė parodo, kad kiekvienoje iteracijoje optimizuojamas tik naujai pridamas medis ( $f_t$ ), o ankstesnių iteracijų prognozė ( $\hat{y}_i^{(t-1)}$ ) laikoma fiksuota, o tai reiškia, kad naujas medis mokomas taip, kad sumažintų ankstesnio modelio klaidas.

Svarbu pažymėti, kad XGBoost metodas taip pat gali būti pritaikytas darbui su nesubalansuotais duomenimis. Vienas iš pagrindinių būdų yra svorių priskyrimas modeliui padarytoms klaidoms, kai skirtingoms klasėms priskiriami skirtingi svoriai. Tai įgyvendinama naudojant parametą „*scale\_pos\_weight*“. Modelis skiria didesnes baudas klaidoms, padaromoms neteisingai klasifikuojant mažumos klasės stebėjimus [65].

#### 2.4.4. LightGBM metodas

LightGBM yra gradientinio stiprinimo sprendimų medžių (angl. *gradient boosting decision trees* – GBDT) metodas, skirtas užduotims su dideliais duomenų rinkiniais. Šis metodas išplečia tradicinį GBDT algoritmą pagerindamas skaičiavimų greitį, tačiau išlaikydamas ar net pagerindamas prognozavimo tikslumą [66].

Tradicinis GBDT modelis konstruojamas kaip daugelio sprendimų medžių ansamblis, kuriame kiekvienas medis apmokomas nuosekliai. Kaip ir XGBoost algoritmo metu, tai ir LightGBM kiekvienas naujas medis mokos taip, kad sumažintų ankstesnio modelio klaidas. Kiekvienos iteracijos metu modelis mokosi pagal nuostolių funkcijos gradientus. Gradientai šiuo atveju yra nuostolių funkcijos išvestinės pagal modelio prognozes, jie parodo, kaip turi būti koreguojama prognozė siekiant sumažinti prognozavimo klaidą:

$$g_i = -\frac{\partial \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad (18)$$

čia  $\ell(y_i, \hat{y}_i)$  žymi nuostolių funkciją, įvertinančią skirtumą tarp tikrosios reikšmės  $y_i$  ir modelio prognozės  $\hat{y}_i$ .

GBDT modelio galutinė prognozė yra išreiškiama kaip medžių suma:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (19)$$

čia  $\hat{y}_i$  – prognozuojama reikšmė,  $f_k$  – k-asis sprendimų medis,  $\mathcal{F}$  - visų galimų medžių erdvė,  $K$  – medžių skaičius.

Didžiausios skaičiavimų sąnaudos šiame algoritme tenka optimalių skaidymo taškų (angl. *split points*) paieškai, nes reikia analizuoti visus stebėjimus ir visas jų požymių reikšmes. LightGBM sumažina skaičiavimų apimtį, kadangi naudoja histogramų pagrindu veikiančią algoritmą, kuris kiekvieno požymio reikšmes susikirsto į intervalus ir priskiria intervalo numerį. Tokiu būdu vietoj tikslų kiekvieno stebėjimo požymių reikšmių modelis vertina šiuos intervalus, todėl sumažėja galimų skaidymo taškų skaičius ir pagreitėja modelio apmokymas [66]. Svarbi LightGBM dalis yra GROSS ir EFB metodai. GROSS metodas remiasi tuo, kad stebėjimai su didesniais gradientais turi didesnę įtaką modelio apmokymui. EFB apjungia tarpusavyje išskirtinius, nesuderinamus požymius (kintamuosius) į vieną bendrą požymį, o tai sumažina nepriklausomų kintamųjų skaičių duomenyse ir atitinkamai didėja apmokymo greitis.

Lyginant LightGBM su XGBoost, abu metodai priklauso tai pačiai gradientinio stiprinimo algoritmu klasei ir paremti ansambliniu medžių mokymusi. XGBoost orientuotas į optimizavimą ir

regularizaciją, o LightGBM į skaičiavimo efektyvumą. LightGBM naudoja lapų auginimo strategiją, leidžiančią greičiau sumažinti modelio klaidą ir pasiekti gerą tikslumą, tačiau atitinkamai yra didesnė persimokymo rizika.

#### 2.4.5. Atsitiktiniai miškai

Atsitiktiniai miškai (angl. *Random Forest*) yra ansamblinio mokymosi metodas. Pagrindinė šio metodo idėja yra ta, kad vietoje vieno sprendimų medžio modelio sudaroma daug skirtingų (mažesnių) sprendimų medžių, o jų rezultatai sujungiami [67].

Kiekvienas sprendimų medis atsitiktinių miškų modelyje yra apmokomas naudojant skirtingą duomenų imtį. Šios duomenų imtys sudaromos atsitiktinai parenkant stebinius iš pradinio, modelio mokymui skirto, duomenų rinkinio. Panaudoti stebiniai gražinami atgal ir vėl gali būti panaudojami kitam sprendimų medžiui apmokinti. Kadangi stebiniai kiekvienam medžiui parenkami naujai, kiekvienas medis gauna skirtingus duomenis ir išmoksta skirtingus dėsningumus. Sprendimų medžio mokymo metu kiekviename šio medžio skaidymo žingsnyje yra naudojamas tik atsitiktinai parinktas požymių (nepriklausomų kintamųjų) rinkinys, o tai reiškia, kad medis nevertina visų duomenų rinkinyje esančių kintamųjų vienu metu.

Kadangi šio darbo tyrime nagrinėjamas tikslo kintamasis yra dvireikšmis, tai atsitiktinių miškų modelyje galutinė prognozė nustatoma balsavimo principu, pasirenkant dažniausiai pasikartojančia klasę tarp visų sprendimų medžių prognozių. Taigi, galutinė prognozė gali būti išreiškiama per tikimybę:

$$P(y = 1|x) = \frac{1}{T} \sum_{t=1}^T f_t(x), \quad (20)$$

čia  $f_t(x)$  – atskiro sprendimų medžio prognozė, o  $T$  – bendras medžių skaičius.

Jei  $P > 0,5$ , tai priskiriama priklausomojo kintamojo reikšmė yra 1, kitais atvejais 0.

Šiame darbe atsitiktinių miškų metodas taikomas prognozavimo uždaviniui spręsti. Jis pasirinktas dėl savo stabilumo, gebėjimo dirbti su didelės apimties duomenimis ir gerų prognozavimo rezultatų pastebėtų ir aprašytų literatūros apžvalgos skyriuje.

#### 2.4.6. EasyEnsemble metodas

EasyEnsemble yra mašininis ansamblio mokymosi metodas, skirtas prognozavimo ir klasifikavimo uždaviniams spręsti tai atvejais, kai duomenų rinkinys yra nesubalansuotas [68]. Prieš tai jau aprašytas SMOTE metodas duomenų disbalanso problemai spręsti remiasi mažumos klasės didinimu (angl. *oversampling*), o šis EasyEnsemble metodas – daugumos klasės mažinimu (angl. *undersampling*).

EasyEnsemble algoritmas [68]:

1. Duomenys suskirstomi į dvi aibes: mažumos klasės stebėjimų aibė ( $\mathcal{P}$ ) ir daugumos klasės stebėjimų aibė ( $\mathcal{N}$ ), kai  $|\mathcal{P}| < |\mathcal{N}|$ .
2. Parenkamas ansamblio dydis ( $\mathcal{T}$ ), kuris nusako, kiek subalansuotų duomenų poabių bus sudaryta.
3. Nustatomas  $s_i$  – iteracijų skaičius nurodantis kiek kartų apmokyti AdaBoost ansamblį  $H_i$ .
4. Kiekvienai iteracijai nuo  $i = 1$  iki  $\mathcal{T}$ :

- a) atsitiktinai parenkamas daugumos klasės poaibis  $\mathcal{N}_i$ , kurio dydis lygus mažumos klasės  $\mathcal{P}$  dydžiui ( $|\mathcal{N}_i| = |\mathcal{P}|$ ).
- b) sudaromas subalansuotas mokymo rinkinys  $\mathcal{P} \cup \mathcal{N}_1$ .
- c) Kiekvienam subalansuotam rinkiniui apmokomas atskiras AdaBoost ansamblis ( $H_i$ ), sudarytas iš  $s_i$  silpnų klasifikatorių  $h_{i,j}$  su atitinkamais svoriais  $\alpha_{i,j}$ .
- d) Kiekvieno ansamblio prognozė paskaičiuojama pagal formulę:

$$H_i(x) = \text{sgn} \left( \sum_{j=1}^{s_i} \alpha_{i,j} h_{i,j}(x) - \theta_i \right), \quad (21)$$

čia  $\theta$  – sprendimo slenkstis,  $\text{sgn}(\cdot)$  - ženklų funkcija.

5. Galutinė prognozė gaunama agreguojant visų ( $\mathcal{T}$ ) ansamblių rezultatus:

$$H(x) = \text{sgn} \left( \sum_{i=1}^{\mathcal{T}} \sum_{j=1}^{s_i} \alpha_{i,j} h_{i,j}(x) - \sum_{i=1}^{\mathcal{T}} \theta_i \right). \quad (22)$$

Šis metodas suteikia galimybę spręsti prognozavimo ir klasifikavimo uždavinius naudojant nesubalansuotus duomenis. Priešingai nei jau aptartas SMOTE metodas, EasyEnsemble visiškai nepakeičia esamo duomenų rinkinio, modelio mokymas vyksta naudojant tik tikrus, o ne sintetinius stebinius.

## 2.5. Mašininio mokymosi modelių gerumo vertinimas

Mašininio mokymosi modelių gerumą ir kokybę galima įvertinti tiek grafinėmis priemonėmis, tiek statistiniais rodikliais. Šiame tyrime modeliai prognozuoja tikimybę, kad įmonė patirs finansinius sunkumus 2025 metais. Įprastu atveju esant tikimybei didesnei už arba lygiai 0,5 įmonė laikoma patiriančia finansinius sunkumus, tačiau tyrime sprendimo riba (angl. *threshold*) yra optimizuojama pagal F1 rodiklį ir kiekvienam mašininio mokymosi modeliui ji yra skirtinga.

### 2.5.1. Statistiniai gerumo matai

Mašininio mokymosi modelių kokybė tyrime vertinama naudojant pagrindinius statistinius gerumo rodiklius, leidžiančius įvertinti jų prognozavimo tikslumą.

Sumaišymo matrica tiesiogiai įvertinti modelio tikslumo nepadedą, tačiau ji reikalinga tam tikrų tikslumo rodiklių skaičiavimams aprašyti. Ji sudaryta iš 4 reikšmių:

- Teisingas teigiamas, parodo kiek tikrų teigiamų stebinių suprognozuota teisingai;
- Teisingas neigiamas, parodo kiek tikrų neigiamų stebinių suprognozuota teisingai;
- Klaidingas teigiamas, parodo kiek tikrų teigiamų stebinių suprognozuota neteisingai;
- Klaidingas neigiamas, parodo kiek tikrų neigiamų stebinių suprognozuota neteisingai;

Visas šias 4 reikšmes galima pavaizduoti lentele (2 lentelė).

**2 lentelė.** Sumaišymo matrica

		Prognozuojama reikšmė	
		Patirs finansinių sunkumų	Nepatirs finansinių sunkumų
Tikroji reikšmė	Patirs finansinių sunkumų	<i>TT</i>	<i>KN</i>
	Nepatirs finansinių sunkumų	<i>KT</i>	<i>TN</i>

Pagrindinis modelių vertinimo matas yra tikslumas:

$$Tikslumas = \frac{TT+TN}{TT+TN+KT+KN}. \quad (23)$$

Ši metrika vertina teisingai suprognuotų stebinių dalį visame duomenų rinkinyje. Šio darbo kontekste, kuomet duomenys yra nesubalansuoti, tikslumas nėra pagrindinis ir vienintelis modelio gerumo vertinimo matas. Kadangi duomenyse įmonių, patiriančių sunkumus, yra mažuma, didžiausias dėmesys vertinant modelių kokybę yra jautrumo matui:

$$Jautrumas = \frac{TT}{TT+KN}. \quad (24)$$

Jautrumas parodo, kaip gerai modelis atskiria ir prognozuoja tikrus teigiamus stebinius – įmones patiriančias finansinius sunkumus.

Tyrimė taip pat naudojamas ir specifiškumas:

$$Specifiškumas = \frac{TN}{TN+KT}. \quad (25)$$

Specifiškumas parodo, kaip gerai modelis atskiria ir prognozuoja tikrus neigiamus stebinius – įmones nepatiriančias finansinių sunkumų.

Preciziškumas parodo, kiek iš visų įmonių, kurioms yra prognozuojami finansiniai sunkumai, iš tiesų yra tos, kurios finansinius sunkumus patirs.

$$Preciziškumas = \frac{TT}{TT+KT}. \quad (26)$$

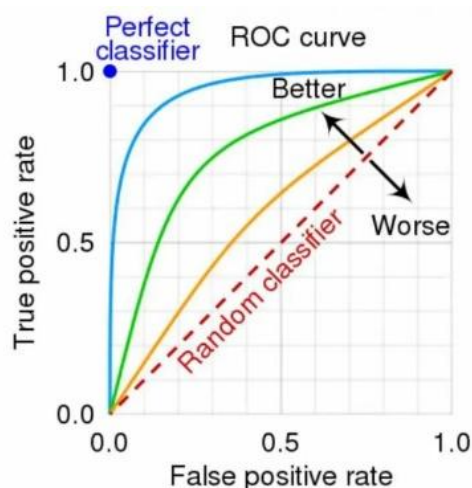
F1 matas yra harmoninis preciziškumo ir jautrumo harmoninis vidurkis, leidžiantis įvertinti modelio gebėjimą tiek teisingai identifikuoti teigiamus stebinius, tiek išvengti klaidingų prognozių.

$$F1 = \frac{2}{\frac{1}{Preciziškumas} + \frac{1}{Jautrumas}}. \quad (27)$$

Šis rodiklis ypač svarbus atliekant tyrimą su nesubalansuotais duomenimis.

## 2.5.2. Modelių rezultatų kreivės

Šiame tyrime mašininio mokymosi modelių kokybė vertinama naudojant ROC (angl. *Receiver Operating Characteristic*) kreivę, kuri leidžia įvertinti modelio gebėjimą atskirti klases. Ši kreivė, pateikta 2 paveiksle, vaizduoja jautrumo ir specifškumo priklausomybę.



2 pav. ROC kreivė<sup>4</sup>

Pagal ROC kreivę yra apskaičiuojamas ir AUC rodiklis (angl. *Area under the curve*), kuris parodo kaip gerai modelis atskiria skirtingas klases. Kuo didesnė AUC reikšmė, tuo modelis kokybiškesnis, o jei reikšmė artima 0,5, tai modelis prognozės prilygsta atsitiktiniam spėjimui.

## 2.6. Kintamųjų atranka

Tyrime atliekama kintamųjų atranka naudojant požymių svarbos vertinimo permutacijos būdu (angl. *permutation importance*) metodą. Pritaikius šį metodą, gaunami kintamųjų svarbos įverčiai, parodantys, kiek pablogėja modelio prognozavimo tikslumas sumaišius konkretaus kintamojo reikšmes, o tai leidžia identifikuoti kintamuosius, turinčius didžiausią įtaką modelio prognozėms.

Kintamųjų svarbos įverčiai priklauso nuo pasirinkto modelio vertinimo rodiklio, pavyzdžiui, AUC, bendrojo tikslumo, F1 mato. Gautas įvertis interpretuojamas, kaip modelio kokybės sumažėjimas. Jei sumaišius kintamąjį „pajamos“ modelio AUC sumažėja per 0,12, o sumaišius kintamąjį „skolos“ sumažėja per 0,02, tai laikoma, kad kintamasis „pajamos“ turi didesnę įtaką modelio rezultatų kokybei. Jei sumaišius kintamąjį modelio kokybės vertinimo matas padidėja – gali būti gaunamos ir neigiamos svarbos reikšmės, o tai rodo, kad modelio veikimas pagerėja ir toks kintamasis gali būti laikomas nereikšmingu arba turinčiu neigiamą įtaką modelio veikimui.

## 2.7. Mašininio mokymosi modelių interpretavimas

Mašininio mokymosi modeliai pasižymi aukštu prognozavimo tikslumu, tačiau jų rezultatus sunku interpretuoti. Logistinė regresija yra gana lengvai interpretuojama, nes modelio koeficientai tiesiogiai nusako nepriklausomų kintamųjų įtaką prognozei. Sudėtingesni metodai, pavyzdžiui, XGBoost, veikia „juodosios dėžės“ principu, todėl jų sprendimų logika nėra lengvai paaiškinama. Dėl šios priežasties, atliekant tyrimą svarbu ne tik sukurti tiksliai prognozuojantį modelį, bet ir suprasti,

<sup>4</sup> WIKIMEDIA COMMONS. *ROC curve* [interaktyvus], 2018. Prieiga per internetą: [https://commons.wikimedia.org/wiki/File:Roc\\_curve.svg](https://commons.wikimedia.org/wiki/File:Roc_curve.svg) [žiūrėta gegužės 17, 2026].

paaiškinti kokie veiksniai šias prognozes lemia. Toks supratimas leidžia neapsiriboti vien tik modelio prognozuojamais rezultatais (patirs finansinių sunkumų ar ne), bet ir suteikia galimybę juos interpretuoti bei pritaikyti praktikoje. Verslo vadyboje, sprendimų priėmimo kontekste, kur svarbus ne tik prognozės tikslumas, bet ir jos pagrįstumas, aiškumas ypač svarbu gebėti paaiškinti modelio rezultatus ir identifikuoti pagrindinius juos lemiančius veiksnius. Šio darbo tyrime, geriausiai prognozuojančiam modeliui taikomi SHAP ir DiCE interpretavimo metodai.

### 2.7.1. SHAP metodas

SHAP (angl. *Shapley additive explanations*) yra vienas iš plačiausiai naudojamų aiškinamojo dirbtinio intelekto metodų, skirtų interpretuoti mašininio mokymosi modelių prognozes. Šis metodas remiasi Shapley'aus reikšmėmis iš žaidimų teorijos. SHAP metodo kontekste „žaidėjais“ laikomi modelio kintamieji (požymiai), o „laimėjimas“ – modelio prognozė. Kiekvienam požymiui priskiriama reikšmė, parodanti jo indėlį į galutinę prognozę.

SHAP metodas vertina kiekvieno požymio svarbą analizuodamas, kaip keičiasi modelio prognozė, kai konkretus įtraukiamas į kitų požymių rinkinius. Požymio svarba nustatoma vertinant jo vidutinį poveikį visose galimose požymių aibėse. SHAP reikšmės apskaičiuojamos remiantis formule [69]:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_{S \cup \{i\}}(x) - f_S(x)], \quad (28)$$

čia  $\phi_i$  – konkretaus požymio indėlis,  $S$  – pasirinktas požymių poaibis,  $F$  – visas požymių rinkinys, o  $f_S(x)$  – modelio prognozė, naudojant tik tam tikrą požymių rinkinį  $S$ .

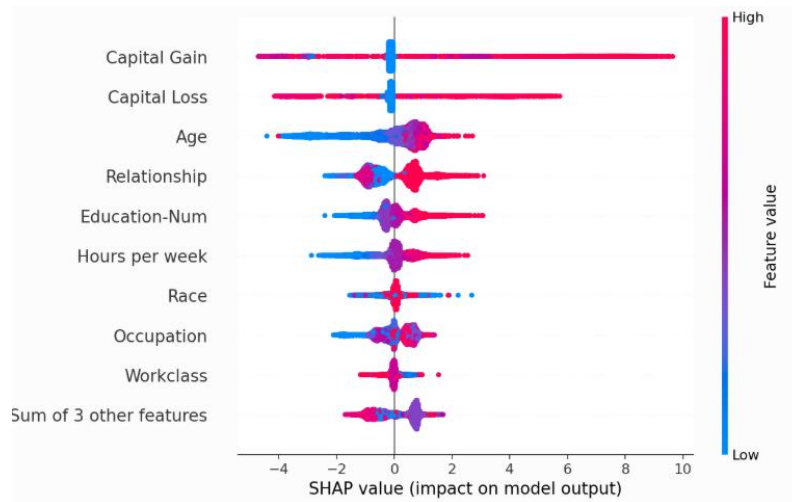
Straipsnyje [69] išskiriamos trys SHAP metodo savybės (privalumai): lokalus tikslumas, nebuvimo savybė ir nuoseklumas. Lokaliojo tikslumo savybė reiškia, kad visų SHAP reikšmių suma turi būti lygi modelio prognozei:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i. \quad (29)$$

Nebuvo savybė užtikrina, kad požymio SHAP reikšmė bus lygi nuliui, jei požymis neįtraukimas į konkrečią požymių kombinaciją, o nuoseklumo savybė garantuoja, kad didėjant požymio svarbai SHAP reikšmė nemažės.

Šiame darbe modelio interpretavimui naudojamas Tree SHAP metodas, skirtas būtent medžių pagrindu veikiantiems modeliams. Palyginti su bendroju Kernel SHAP metodu, Tree SHAP pasižymi didesniu skaičiavimų greičiu vertinant medžių tipo modelius.

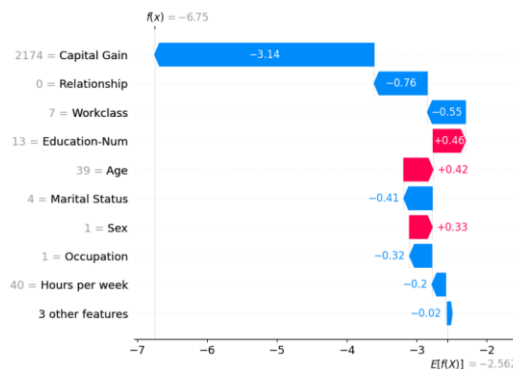
SHAP metodas taip pat suteikia galimybę vizualizuoti modelio sprendimus naudojant įvairius grafikus. Vienas dažniausiai naudojamų grafikų yra „SHAP Summary Plot“, pateiktas 3 paveiksle.



3 pav. „SHAP Summary Plot“<sup>5</sup>

Šis grafikas pateikia bendrą informaciją apie visų požymių svarbą modelyje. Jame galima vienu metu matyti kurie požymiai yra svarbiausi, kokią įtaką jie daro prognozei ir kaip požymių reikšmės susijusios su modelio atsaku. 3 paveiksle galima matyti, kad kintamasis „Age“ yra vienas iš svarbiausių modelio sprendimų priėmimo, taip pat didėjant šio kintamojo reikšmėms didėja ir tikimybė, kad modelis pasirinktų klasę 1. Kintamasis „Workclass“ tikimybę veikia atvirkščiai – didesnės šio kintamojo reikšmės mažina tikimybę.

Individualių prognozių interpretavimui naudojamas „SHAP Waterfall Plot“ grafikas, pateiktas 4 paveiksle.



4 pav. „SHAP Waterfall Plot“<sup>6</sup>

Šis grafikas paaiškina kodėl modelis konkrečiam objektui pateikia tokią prognozę. Paveiksle 4 pažymėta, kad bazinė  $f(x)$  reikšmė (vidutinė modelio SHAP reikšmių suma visame duomenų rinkinyje) yra apie -2,562. Raudoni požymiai didina modelio prognozę, o mėlyni požymiai atitinkamai mažina.

<sup>5</sup> SHAP Documentation. *SHAP beeswarm plot* [interaktyvus]. Prieiga per internetą: [https://shap.readthedocs.io/en/latest/example\\_notebooks/api\\_examples/plots/beeswarm.html](https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html) [žiūrėta gegužės 17, 2026].

<sup>6</sup> SHAP Documentation. *Waterfall plot example* [interaktyvus]. Prieiga per internetą: [https://shap.readthedocs.io/en/latest/example\\_notebooks/api\\_examples/plots/waterfall.html](https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/waterfall.html) [žiūrėta gegužės 17, 2026].

## 2.7.2. DiCE metodas

DiCE (angl. *Diverse Counterfactual Explanations*) taip pat yra mašininio mokymosi modelių interpretavimo metodas. Priešingai nei SHAP, kurio tikslas paaiškinti kaip požymiai prisideda prie prognozės, DiCE siekia atsakyti į klausimą: kokius minimalius pokyčius reikėtų atlikti, kad prognozė pasikeistų [70]?

Pagrindinė šio metodo idėja yra generuoti alternatyvius scenarijus (kontrafaktus), kurie yra kuo artimesni pradinei situacijai (įvesties duomenims), tačiau lemia kitokį modelio sprendimą, o tai leidžia interpretuoti modelį, praktiškai įvertinant galimus sprendimų pokyčius. DiCE metodo suteikiamos interpretavimo galimybės ypač svarbios verslo vadybos kontekste, kur sprendimų priėmimas reikalauja ne tik suprasti esamą situaciją, bet ir įvertinti galimas alternatyvas, reikalingus pokyčius.

Alternatyvus scenarijus (kontrafaktas) apibrėžiamas kaip naujas stebėjimas  $x'$ , kuris yra panašus pradiniam stebėjimui  $x$ , tačiau lemia kitokį prognozės rezultatą:

$$f(x) \neq f(x'), \quad (30)$$

čia  $f(x)$  – modelio prognozė naudojant pradinį stebinį, o  $f(x')$  – modelio prognozė naudojant naujai sugeneruotą stebinį.

Taigi, DiCE metodas siekia rasti minimalų skirtumą tarp  $x$  ir  $x'$  stebinių, išlaikant realistiškas ir interpretuojamas reikšmes. Metodo tikslas yra minimizuoti atstumą tarp pradinio ir alternatyvaus stebinio, užtikrinti prognozės rezultato pasikeitimą ir išlaikyti sugeneruotų pavyzdžių įvairovę [70]. Jis taip pat leidžia apriboti, kurie kintamieji gali būti keičiami, todėl generuojami paaiškinimai gali būti pritaikomi praktiniams sprendimams. Šiame tyrime DiCE metodas naudojamas papildyti SHAP analizę ir pateikti alternatyvius scenarijus, kurie parodo, kokie kintamųjų pokyčiai galėtų nulemti įmonės prognozės pasikeitimą iš klasės „patirs finansinių sunkumų 2025 metais“ į klasę „nepatirs finansinių sunkumų 2025 metais“.

## 2.8. Statistinis modelių palyginimas

Siekiant įvertinti, ar skirtumas tarp modelių (su ESG ir be ESG) prognozavimo tikslumo yra statistiškai reikšmingas, šiame tyrime taikomas Vilkoksono ženklinių rangų kriterijus (angl. *Wilcoxon signed-rank test*) [71]. Šis testas pasirinktas dėl to, nes nėra būtina normalumo prielaida ir galima lyginti dvi tarpusavyje susijusias imtis (porines imtis).

Mašininio mokymosi modelių prognozavimo kokybė vertinama naudojant AUC metriką. Taikant  $k$ -kartų kryžminę validaciją, kiekvienam modeliui gaunami poriniai AUC įverčiai:

$$(AUC_i^{ESG}, \quad AUC_i^{beESG}), \quad i = 1, \dots, k.$$

Toliau apskaičiuojami poriniai skirtumai:

$$d_i = AUC_i^{ESG} - AUC_i^{beESG}. \quad (31)$$

Atvejais, kai  $d_i = 0$ , yra atmetami, o likusiems skirtumams apskaičiuojamos absoliučios reikšmės  $|d_i|$ , kurios išrikiuojamos didėjimo tvarka ir priskiriamas rangas  $R_i$  bei ženklas:

$$\text{sgn}(d_i) = \begin{cases} +1, & d_i > 0 \\ -1, & d_i < 0 \end{cases} \quad (32)$$

Tuomet apskaičiuojamos teigiamų ir neigiamų rangų sumos:

$$W^+ = \sum_{d_i > 0} R_i, \quad (33)$$

$$W^- = \sum_{d_i < 0} R_i. \quad (34)$$

Iš šių dviejų sumų parenkama mažesnė ir priskiriama, kaip Vilkoksono statistikos reikšmė:

$$W = \min(W^+, W^-). \quad (35)$$

Tyrimo tikrintos hipotezės:

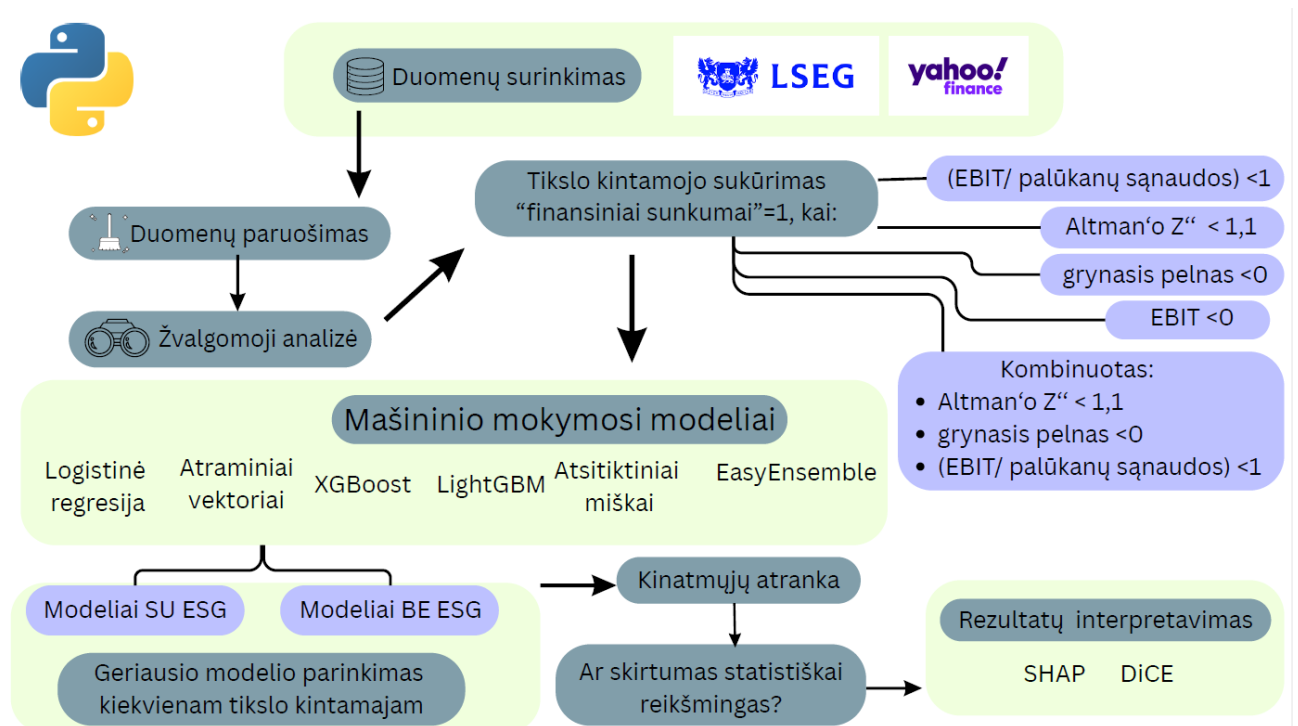
$$H_0: \text{skirtumų } (d_i) \text{ mediana} = 0,$$

$$H_1: \text{skirtumų } (d_i) \text{ mediana} \neq 0.$$

Kadangi tyrime porinių stebėjimų skaičius buvo mažas ( $k = 5$ ), naudota tiksli Vilkoksonos testo p-reikšmė. Statistinis reikšmingumas vertinamas pagal 0,05 reikšmingumo lygmenį.

## 2.9. Tyrimo eiga

Paveiksle 5 pateikiama šio darbo tyrimo eigos schema.



5 pav. Tyrimo eigos schema

### 3. Tyrimo rezultatai

#### 3.1. Žvalgomoji analizė

Iš pradinių duomenų buvo palikti tik tie įrašai, kurie neturėjo trūkstančių reikšmių. 5 tikslo kintamieji yra išskaičiuoti remiantis 2025 metų duomenimis, kurie po to buvo pašalinti, paliekant tik 5 tikslo kintamuosius 2025 metams. Taip pat buvo suskaičiuoti 16 kintamųjų, kurie atspindi skirtumus tarp 2023 metų ir 2024 metų duomenų. Galiausiai duomenų rinkinį sudarė 1315 įrašai, turintys po 64 kintamuosius. Visų kintamųjų sąrašas pateikiamas 1 priede. Tikslo kintamųjų reikšmės buvo taip arba ne. Iš šių kintamųjų reikšmių pasiskirstymų matoma, jog kiekvieno kintamojo atveju yra klasių disbalansas. Reikšmių pasiskirstymas yra pateiktas 3 lentelėje.

**3 lentelė.** Reikšmių pasiskirstymas pagal tikslo kintamuosius

Tikslo kintamasis	Ne	Taip
financial_d_altzzz_2025	1060	255
financial_d_net_income_2025	1206	109
financial_d_interest_coverage_2025	1189	126
financial_d_ebit_2025	1234	81
financial_d_combined_2025	977	338

Duomenų rinkinyje taip pat yra ir kategorinių kintamųjų: šalis, įmonės veiklos sektorius. Daugiausia įmonių yra iš Jungtinių Amerikos Valstijų - 597. Daugiau šalių ir jų pasiskirstymai pateikiami 4 lentelėje.

**4 lentelė.** Šalių pasiskirstymas

Reikšmė	Kiekis	Dažnis (%)
Jungtinės Amerikos Valstijos	597	45,4
Jungtinė Karalystė	142	10,8
Australija	120	9,1
Švedija	53	4
Kanada	50	3,8
Vokietija	40	3
Prancūzija	40	3
Kinija	37	2,8
Korėjos Respublika (Pietų Korėja)	28	2,1
Šveicarija	25	1,9
Kitos reikšmės (23)	183	13,9

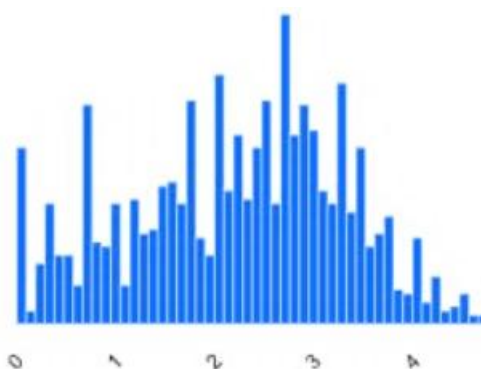
Įmonės veiklos sektoriaus reikšmių pasiskirstymas pateikiamas 5 lentelėje.

**5 lentelė.** Įmonių veiklos sektoriai

Reikšmė	Kiekis	Dažnis (%)
Programinė įranga ir kompiuterinės paslaugos	102	7,8
Kelionės ir laisvalaikis	76	5,8
Mažmeninės prekybos įmonės	73	5,6
Nafta, dujos ir anglis	71	5,4
Statyba ir medžiagos	56	4,3
Pramoninis transportas	53	4
Technologinė aparatinė įranga ir įrenginiai	52	4
Pramoninės palaikymo paslaugos	51	3,9
Medicinos įranga ir paslaugos	49	3,7
Pramoniniai metalai ir kasyba	45	3,4
Kitos reikšmės (30)	684	52

Rinkos kapitalizacijos mažiausia reikšmė 2023 metais buvo 4 343 167,8, o 2024 metais 4 467 968,4. Didžiausia reikšmė 2023 metais buvo 2 794 827 900 000, o 2024 metais buvo 3 288 762 100 000. Vidurkis 2023 metais buvo 33 117 081 000, 2024 metais 38 565 191 000, o mediana 2023 metais buvo 7 631 019 700 ir 2024 metais buvo 7 748 083 400. Rinkos kapitalizacijos skirstinys yra stipriai asimetriškas – dauguma įmonių pasižymi santykinai mažesne rinkos kapitalizacija, tačiau imtyje esantis nedidelis skaičius itin didelės kapitalizacijos įmonių reikšmingai padidina vidutinę reikšmę.

Duomenų imtyje yra ir ESG rodiklių. Vienas iš jų yra aplinkosaugos. Šio agreguoto balo reikšmės 2023 metais yra intervale nuo 0 iki 5, kur 0 reiškia žemiausią, o 5 – aukščiausią aplinkosaugos įvertinimą, vidurkis yra 2,214, o histograma pateikiama 6 paveiksle.



**6 pav.** Agreguoto aplinkosaugos balo histograma

### 3.2. Finansinių sunkumų prognozavimas

Tyrimo metodų skyriuje tikslo kintamasis buvo apibrėžtas penkiais skirtingai būdai, todėl pagal kiekvieną tikslo kintamąjį buvo atlikta finansinių sunkumų prognozavimo analizė. Kiekvienam kintamajam ji apima:

- Modelių apmokymą, neįtraukiant ESG rodiklių;

- Modelių apmokymą, įtraukiant ESG rodiklius;
- Geriausio modelio be ESG rodiklių parinkimą;
- Geriausio modelio su ESG rodikliais parinkimą;
- Kintamųjų atranką abiem geriausiems modeliams;
- Įvertinimą ar statistiškai reikšmingai pasikeičia modelio kokybę įtraukus ESG;
- Geriausiai prognozuojančio modelio su ESG rodikliais interpretaciją su SHAP ir DiCE metodais.

Šio skyriaus lentelėse yra pateikiami metodų pavadinimų sutrumpinimai ir nustatymai:

- „LG“ – logistinė regresija;
- „SVM“ – atraminiai vektoriai;
- „RF“ – atsitiktiniai miškai;
- „XGB“ – XGBoost metodas;
- „LGBM“ – LightGBM metodas;
- „EE“ – EasyEnsemble metodas;
- „+S“ – prieš atliekant modelio apmokymą buvo panaudotas SMOTE metodas duomenų disbalanso problemai išspręsti.
- „+W“ klasių disbalanso problemai spręsti uždedami klasių svoriai;

### 3.2.1. Hiperparametrų optimizavimas

Mašininio mokymosi modelių mokymo metu, kiekvienas metodas turėjo rinkinį hiperparametrų kurie buvo optimizuojami.

XGBoost modelyje klasių disbalansas buvo valdomas naudojant arba SMOTE metodą, arba pritaikant klasių svorius. Hiperparametrai optimizuoti naudojant RandomizedSearchCV su 5 dalių kryžmine validacija, optimizavimo kriterijumi pasirenkant AUC rodiklį. Papildomai optimizuotas klasifikavimo slenkstis, maksimizuojantis F1 rodiklį. XGBoost optimizuoti hiperparametrai pateikiami 6 lentelėje.

6 lentelė. XGBoost hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
n_estimators	100–800	Medžių skaičius ansamblyje
max_depth	2-6	Maksimalus medžio gylis
learning_rate	0,005–0,1	Mokymosi greitis
subsample	0.6–1.0	Naudojama stebėjimų dalis kiekvienam medžiui
colsample_bytree	0.6–1.0	Naudojama požymių dalis kiekvienam medžiui
min_child_weight	1–20	Minimalus stebėjimų svoris lape
gamma	0–5	Minimalus pagerėjimas skaidymui atlikti
reg_alpha	0–5	L1 reguliarizacija
reg_lambda	0.1–50	L2 reguliarizacija

LightGBM modelyje klasių disbalansas buvo valdomas naudojant arba SMOTE metodą, arba pritaikant klasių svorius. Hiperparametrai optimizuoti su 5 dalių kryžmine validacija. LightGBM optimizuoti hiperparametrai pateikiami 7 lentelėje.

**7 lentelė.** LightGBM hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
n_estimators	100–1200	Medžių skaičius ansamblyje
max_depth	1-8	Maksimalus medžio gylis
learning_rate	0,005–0,1	Mokymosi greitis
subsample	0.6–1.0	Naudojama stebėjimų dalis kiekvienam medžiui
colsample_bytree	0.6–1.0	Naudojama požymių dalis kiekvienam medžiui
min_child_samples	5-120	Minimalus stebėjimų skaičius lape
num_leaves	7-127	Maksimalus lapų skaičius
reg_alpha	0–5	L1 reguliarizacija
reg_lambda	0.1–50	L2 reguliarizacija

Atsitiktinių miškų modelyje klasių disbalansas buvo valdomas naudojant arba SMOTE metodą, arba pritaikant klasių svorius. Hiperparametrai optimizuoti naudojant 5 dalių kryžminę validaciją. Atsitiktinių miškų optimizuoti hiperparametrai pateikiami 8 lentelėje.

**8 lentelė.** Atsitiktinių miškų hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
n_estimators	100–1200	Medžių skaičius ansamblyje
max_depth	1-20	Maksimalus medžio gylis
min_samples_split	2-20	Minimalus stebėjimų skaičius mazgo skaidymui
min_samples_leaf	1-12	Minimalus stebėjimų skaičius lape kiekvienam medžiui
max_features	0,3-0,7	Naudojama požymių dalis
bootstrap	True, False	Bootstrap imčių naudojimas
criterion	Gini, entropy, log_loss	Skaidymo kriterijus

EasyEnsemble modelis pasirinktas nesubalansuotų klasių problemai spręsti, nes jis taiko daugumos klasės sumažinimą ir sudaro kelis ansamblio modelius. Bazinis klasifikatorius buvo AdaBoost su sprendimų medžiais. EasyEnsemble optimizuoti hiperparametrai pateikiami 9 lentelėje.

### 9 lentelė. EasyEnsamble hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
n_estimators	5-30	EasyEnsamble iteracijų skaičius
replacement	True, False	Atranka su grąžinimu
AdaBoost n_estimators	25-200	AdaBoost iteracijų skaičius
AdaBoost learning_rate	0.01–1.0	AdaBoost mokymosi greitis
DecisionTree max_depth	1-3	Naudojama požymių dalis
DecisionTree min_samples_leaf	1-10	Bootstrap imčių naudojimas

Logistinėje regresijoje klasių disbalansas buvo valdomas naudojant SMOTE metodą. Logistinės regresijos optimizuoti hiperparametrai pateikiami 10 lentelėje.

### 10 lentelė. Logistinės regresijos hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
penalty	l1, l2, elasticnet	Reguliarizacijos tipas
C	$10^{-4} - 10^3$	Reguliarizacijos stiprumas
L1_ratio	0,05-0,95	L1 ir L2 santykis Elastic Net reguliacijoje

Atraminų vektorių modeliuose buvo tikrintos keturios branduolio funkcijos: tiesinė, RBF, polinominė ir sigmoidinė. Klasių disbalansas buvo valdomas naudojant SMOTE metodą. Atraminų vektorių modeliuose optimizuoti hiperparametrai pateikiami 11 lentelėje.

### 11 lentelė. Atraminų vektorių hiperparametrai

Hiperparametras	Tikrintos reikšmės	Paskirtis
kernel	linear, rbf, poly, sigmoid	Branduolio funkcija
C	$10^{-3} - 10^3$	Reguliarizacijos stiprumas
gamma	scale, auto, 0,001–1	Branduolio koeficientas
degree	2-4	Polinominio branduolio laipsnis
coef0	0-1	Nepriklausomas branduolio narys

### 3.2.2. Finansinių sunkumų prognozavimas, kai tikslo kintamas paremtas Altman'o reikšmėmis

Klasių pasiskirstymas:

- 255 įmonės patiriančios finansinius sunkumus 2025 metais;
- 1060 įmonių nepatiriančių finansinių sunkumų 2025 metais.

12 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį nebuvo įtraukti ESG rodikliai.

**12 lentelė.** Modelių rezultatai (be ESG), kai tikslo kintamasis yra pagal Altman'o rodiklį

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
LGBM+S	0,46	0,89	0,67	0,95	0,91	0,76	0,71	0,76
XGB+S	0,58	0,90	0,65	0,96	0,91	0,79	0,71	0,74
XGB	0,64	0,89	0,65	0,94	0,90	0,73	0,69	0,69
XGB+W	0,69	0,88	0,63	0,94	0,90	0,73	0,67	0,68
RF+S	0,54	0,89	0,57	0,96	0,90	0,78	0,66	0,73
EE	0,51	0,84	0,75	0,86	0,89	0,56	0,64	0,64
RF+W	0,42	0,86	0,67	0,91	0,89	0,63	0,65	0,65
RF	0,37	0,87	0,67	0,92	0,89	0,67	0,67	0,63
LG	0,73	0,89	0,57	0,97	0,88	0,83	0,67	0,75
LGBM+W	0,53	0,87	0,65	0,92	0,88	0,66	0,65	0,65
LG+S	0,79	0,89	0,53	0,98	0,88	0,84	0,65	0,74
LGBM	0,26	0,83	0,73	0,85	0,87	0,54	0,62	0,61
SVM+S (POLY)	0,69	0,85	0,43	0,95	0,79	0,69	0,53	0,53
SVM (POLY)	0,23	0,84	0,45	0,93	0,79	0,61	0,52	0,55

13 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį buvo įtraukti ESG rodikliai.

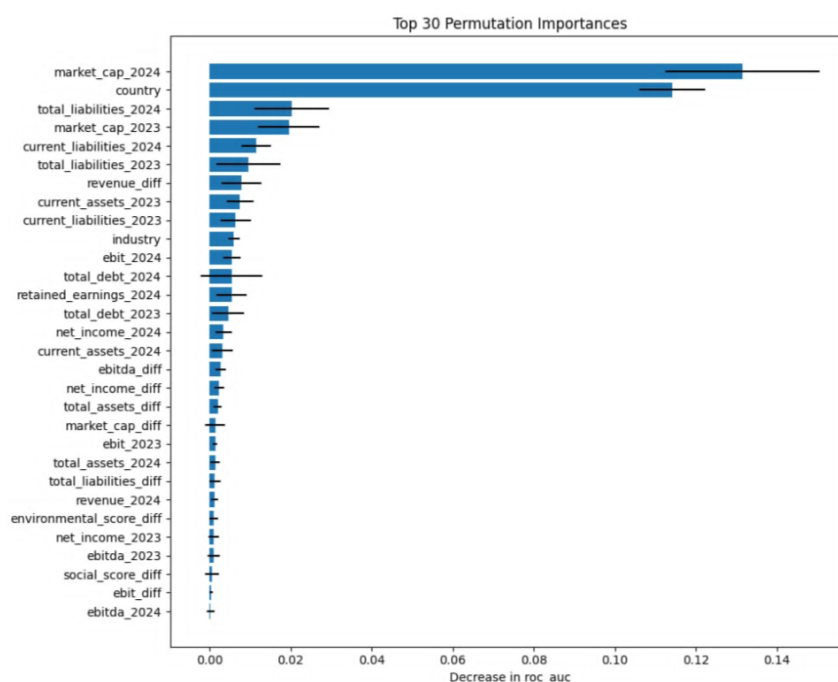
**13 lentelė.** Modelių rezultatai (su ESG), kai tikslo kintamasis yra pagal Altman'o rodiklį

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
EE	0,53	0,89	0,61	0,95	0,91	0,76	0,67	0,67
XGB+S	0,27	0,88	0,76	0,91	0,90	0,66	0,71	0,72
RF+W	0,38	0,86	0,67	0,91	0,89	0,64	0,65	0,66
LGBM+S	0,43	0,88	0,65	0,93	0,89	0,70	0,67	0,71
XGB+W	0,56	0,87	0,67	0,92	0,89	0,65	0,66	0,66
RF	0,57	0,88	0,57	0,95	0,89	0,74	0,64	0,66
XGB	0,42	0,86	0,75	0,88	0,89	0,60	0,67	0,67
RF+S	0,53	0,88	0,55	0,96	0,89	0,78	0,64	0,72
LG+S	0,77	0,90	0,57	0,98	0,88	0,85	0,68	0,74
LGBM+W	0,07	0,84	0,73	0,87	0,88	0,58	0,64	0,63
LGBM	0,57	0,87	0,61	0,93	0,88	0,67	0,64	0,64
LG	0,74	0,89	0,57	0,97	0,88	0,83	0,67	0,74
SVM+S (POLY)	0,25	0,78	0,75	0,79	0,83	0,46	0,57	0,55

SVM (POLY)	0,24	0,79	0,67	0,82	0,81	0,47	0,55	0,52
------------	------	------	------	------	------	------	------	------

Kai neįtraukiami ESG duomenys geriausias modelis pagal AUC reikšmę yra LightGBM su SMOTE metodu disbalanso problemos spręsti. Įtraukiant ESG duomenis geriausias modelis pagal AUC gaunasi EasyEnsamble, tačiau vertinti pagal tik AUC reikšmę yra kiek netikslu šio tyrimo kontekste. Jautrumas šiuo atveju taip pat turi didelę reikšmę, kadangi tikslas yra padaryti kuo mažiau antros rūšies klaidų (nepraleisti įmonių, kurios patirs finansinius sunkumus 2025 metais). Svarbu atsižvelgti į jautrumą ne tik dėl neteisingai prognozuotų, finansinius sunkumus patiriančių įmonių, bet ir dėl to, kad tad šios klasės duomenyse yra mažuma ir geras jos atskyrimas yra pranašumas vertinant modelio kokybę. Taigi, geriausias modelis, kai prognozuojama be ESG rodiklių pasirinktas LightGBM su SMOTE (*jautrumas* = 0,67, *AUC* = 0,91, *F1* = 0,71, *tikslumas* = 0,89). Geriausias modelis duomenų imčiai su ESG rodikliais – XGBoost su SMOTE (*jautrumas* = 0,76, *AUC* = 0,9, *F1* = 0,71, *tikslumas* = 0,88).

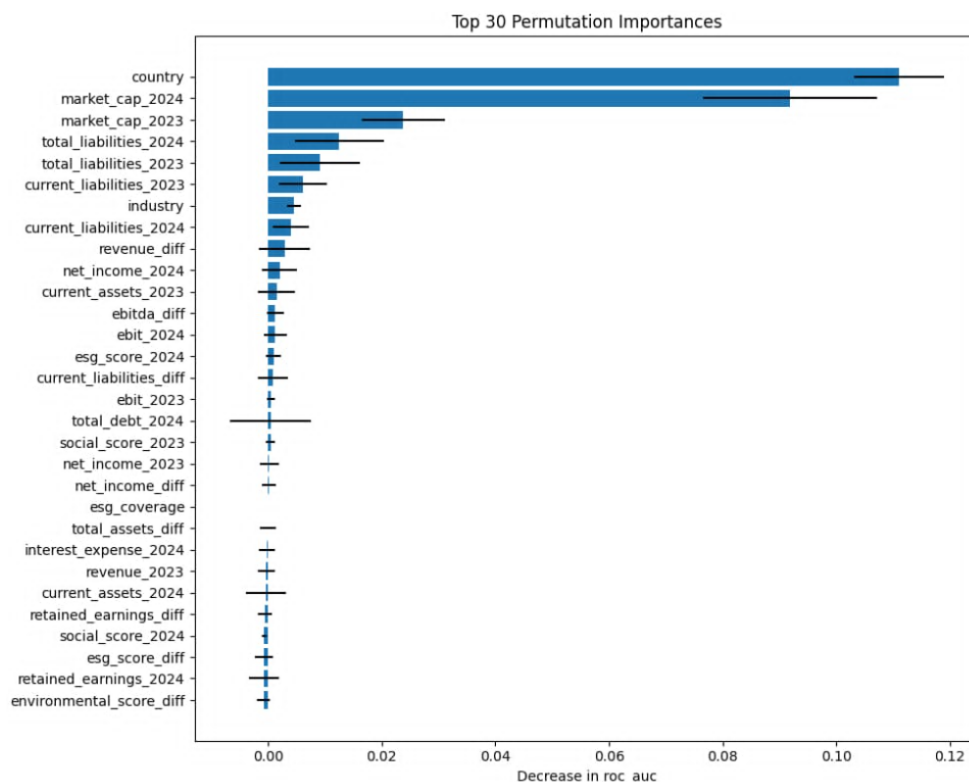
Kintamųjų svarba LightGBM su SMOTE modelyje, kai duomenų imtyje nėra ESG rodiklių, pateikiama 7 paveiksle.



7 pav. Kintamųjų svarba LightGBM su SMOTE (be ESG)

Iš 7 paveiksle pateikiamos diagramos galima daryti išvadą, kad įmonės rinkos kapitalizacijos vertė 2024 metais daro didžiausią įtaką modelio prognozavimo tikslumui (sumaišius šio kintamojo reikšmes AUC reikšmė sumažėja per 0,13). Taip pat, svarbus kintamasis yra šalis, kurio sumaišymas modelio prognozavimo AUC įvertį sumažina per 0,115.

Kintamųjų svarba XGBoost modelyje, kai duomenų imtyje įtraukti ESG rodikliai, pateikiama 8 paveiksle.

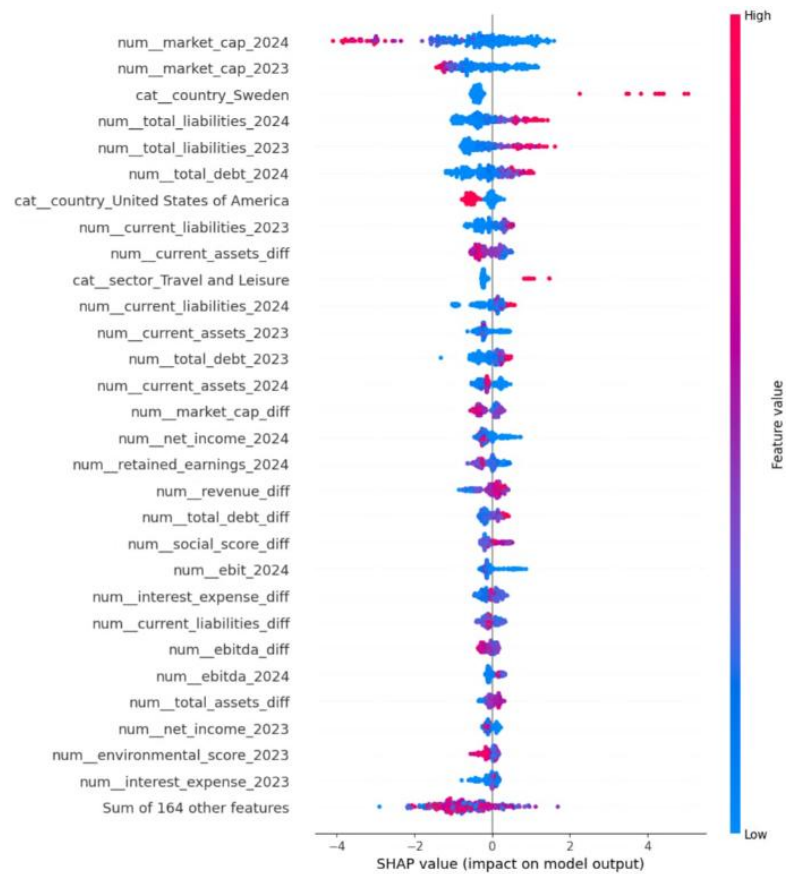


8 pav. Kintamųjų svarba XGBoost (su ESG)

Remiantis 8 paveiksle pateikta diagrama galima pastebėti, kad tie patys kintamieji: rinkos kapitalizacijos vertė 2024 metais ir šalis daro didžiausią įtaką modelio prognozavimo tikslumui, atitinkamai AUC vertę sumažindami per 0,09 ir 0,11.

Toliau labai svarbu įvertinti ar duomenų papildymas ESG rodikliais statistiškai reikšmingai pagerina prognozavimo tikslumą, remiantis AUC įverčiu. Šiam klausimui atsakyti buvo panaudotas Vilkoksono testas. Šio testo nulinė hipotezė: skirtumų mediana tarp lyginamų modelių AUC reikšmių kryžminės validacijos metu yra lygi nuliui. Gauta p-reikšmė yra 0,3125, dėl to nulinė hipotezė yra priimama ir laikoma, kad tarp šių dviejų modelių (LightGBM ir XGBoost) prognozavimo tikslumo nėra statistiškai reikšmingo skirtumo. Galima daryti išvadą, kad ESG rodikliai neprideda papildomos vertės prognozuojant finansinius sunkumus, apibrėžtus remiantis Altman‘o Z‘‘ reikšme.

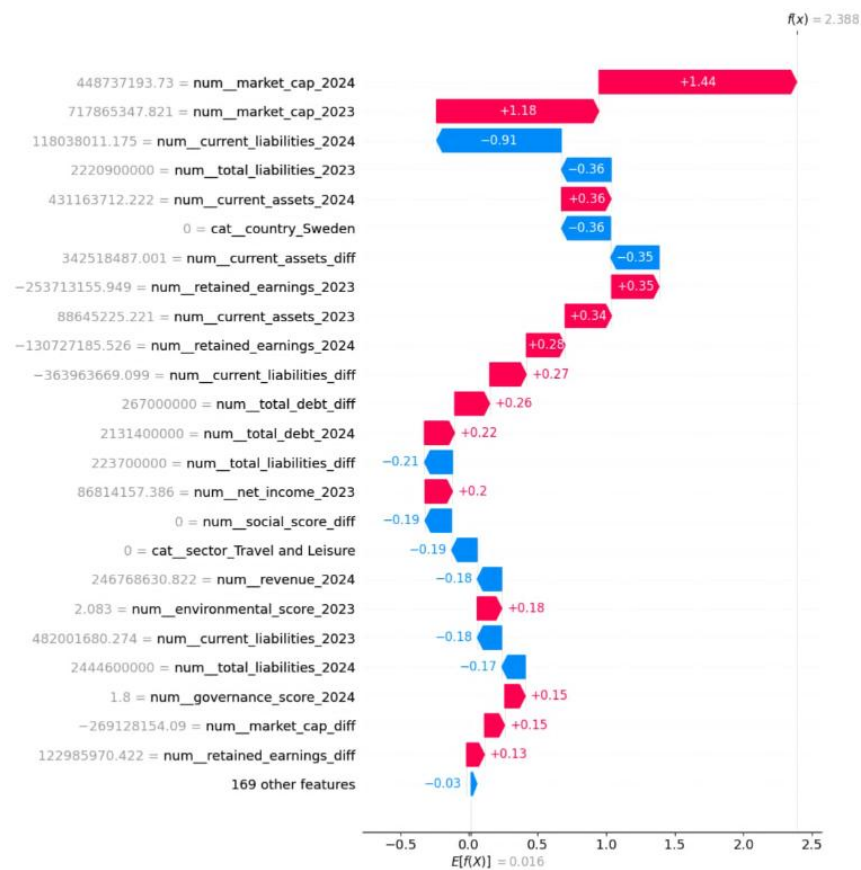
Tolimesnė tyrimo dalis apima šio tyrimo realią praktinę vertę – rezultatų interpretavimą. Interpretavimo metodai pritaikyti tik geriausiam modeliui su ESG rodikliais – XGBoost. 9 paveiksle pateikiama SHAP metodo metu gauta „SHAP summary plot“ diagrama, parodanti kiekvieno kintamojo Shapley‘aus reikšmes.



**9 pav.** „SHAP summary plot“ XGBoost metodui

Remiantis 9 paveikslu, galima daryti išvadą, kad mažesnė rinkos kapitalizacijos vertė 2024 metais lemia didesnes Shapley'aus reikšmes, kurios artina tikimybę link 1 (didėja tikimybė, kad įmonė patirs finansinius sunkumus).

Šis metodas taip pat paaiškina kodėl modelis konkrečiam stebėjimui pateikė vienokią ar kitokią prognozę. Toliau, 10 paveiksle, pateikimas pavyzdys įmonės, kuriai yra prognozuojami finansiniai sunkumai 2025 metais.



10 pav. „SHAP waterfall“ XGBoost metodu

Raudonai pažymėti kintamieji artino modelio prognozę link verdikto, kad įmonė patirs finansinius sunkumus, tuo tarpu mėlyni – tolino. 10 paveiksle pateiktame pavyzdyje galima matyti, kad didžiausią postūmį į teigiamą pusę padarė rinkos kapitalizacijos vertė tiek 2023 metais, tiek 2024 metais. Svarbu paminėti ir tai, kad įmonės išsipareigojimais 2024 metais mažino finansinių sunkumų tikimybę.

Modelio rezultatų interpretavimui buvo panaudotas ir DiCE metodas. Šio metodo pritaikymas iš verslo vadybos pusės leidžia sprendimų priėmėjui numatyti kokius finansinius rodiklius ir kaip juos koreguoti, kad įmonės finansinių sunkumų prognozė 2025 metams pasikeistų. Metodas pateikia 5 alternatyvius scenarijus, pagal nurodytus kintamuosius, kuriuos reiktų keisti konkrečios įmonės kontekste. Pasirinktas pavyzdys, pateikiamas 11 paveiksle.

	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024
0	431163712.0	118038008.0	448737184.0	93822400.0	-130727184.0	246768624.0	4.303100e+09	2.131400e+09	2.444600e+09

Counterfactuals – changed columns + target:

	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	target
0	4773664.5	1816712.0	4467968.5	84380632.0	-2.035500e+10	2.395904e+08	8.580881e+06	0.0	1860174.0	0
0	4773664.5	1816712.0	4467968.5	84380632.0	-2.035500e+10	9.870409e+08	8.580881e+06	0.0	1860174.0	0
0	4773664.5	1816712.0	4467968.5	84380632.0	-2.035500e+10	2.012919e+09	8.580881e+06	0.0	1860174.0	0
0	4773664.5	1816712.0	4467968.5	84380632.0	-2.035500e+10	0.000000e+00	8.580881e+06	0.0	1860174.0	0
0	4773664.5	1816712.0	413316096.0	84380632.0	-2.035500e+10	2.356265e+08	1.117593e+09	531017792.0	1860174.0	0

11 pav. DiCE metodo alternatyvūs scenarijai pasirinktai įmonei

Šiame pavyzdyje nurodyta, kad keisti galima tik kintamuosius „current\_assets\_2024“, „current\_liabilities\_2024“, „market\_cap\_2024“, „net\_income\_2024“, „retained\_earnings\_2024“, „revenue\_2024“, „total\_assets\_2024“, „total\_debt\_2024“ ir „total\_liabilities\_2024“.

DiCE metodo keliamos sąlygos, kad įmonė nepatirtų finansinių sunkumų, sunkiai įgyvendinamos ir nerealistiškos, ne visada atitinkančios ekonomines sąlygas ar finansinius apribojimus. Pasiūlyti scenarijai turėtų būti interpretuojami ne tiesiogiai remiantis pateiktomis reikšmėmis, bet jų kryptimis.

Remiantis 11 paveiksle pateiktais alternatyviais scenarijais, finansinės rizikos mažėjamas siejamas su mažesniu įsipareigojimų lygiu, didėjantis grynasis pelnas ir pajamos taip pat svarbūs veiksniai mažinant finansinę riziką.

Apibendrinat rezultatus tikslo kintamajam, apskaičiuotam pagal Altman'o Z<sup>o</sup> reikšmę, galima daryti išvadą, kad ESG rodiklių įtraukimas statistiškai reikšmingai **nepagerino** mašininio mokymosi modelių prognozavimo tikslumo. Svarbiausi kintamieji modelių prognozėms yra rinkos kapitalizacijos vertė 2024 ir 2023 metais bei šalis.

### 3.2.3. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas grynuoju pelnu

Klasių pasiskirstymas:

- 109 įmonės patiriančios finansinius sunkumus 2025 metais;
- 1206 įmonės nepatiriančios finansinių sunkumų 2025 metais.

14 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį nebuvo įtraukti ESG rodikliai.

**14 lentelė.** Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas neigiamu grynuoju pelnu

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
XGB+S	0,41	0,91	0,86	0,91	0,93	0,48	0,61	0,48
EE	0,56	0,92	0,77	0,93	0,93	0,52	0,62	0,58
RF+W	0,62	0,92	0,73	0,94	0,93	0,52	0,60	0,53
LGBM	0,85	0,94	0,68	0,96	0,92	0,60	0,64	0,57
RF	0,48	0,94	0,73	0,96	0,92	0,62	0,67	0,52
LGBM+W	0,69	0,93	0,73	0,95	0,92	0,55	0,63	0,52
XGB	0,79	0,94	0,73	0,96	0,92	0,62	0,67	0,59
RF+S	0,42	0,92	0,82	0,93	0,92	0,50	0,62	0,47
LGBM+S	0,75	0,92	0,73	0,94	0,91	0,53	0,62	0,47
XGB+W	0,52	0,92	0,82	0,93	0,91	0,53	0,64	0,51
LG	0,55	0,86	0,68	0,87	0,84	0,33	0,44	0,38
LG+S	0,54	0,82	0,68	0,83	0,81	0,27	0,39	0,35
SVM (POLY)	0,09	0,74	0,68	0,75	0,80	0,20	0,31	0,27
SVM+S (SIGMOID)	0,57	0,87	0,27	0,92	0,68	0,24	0,26	0,15

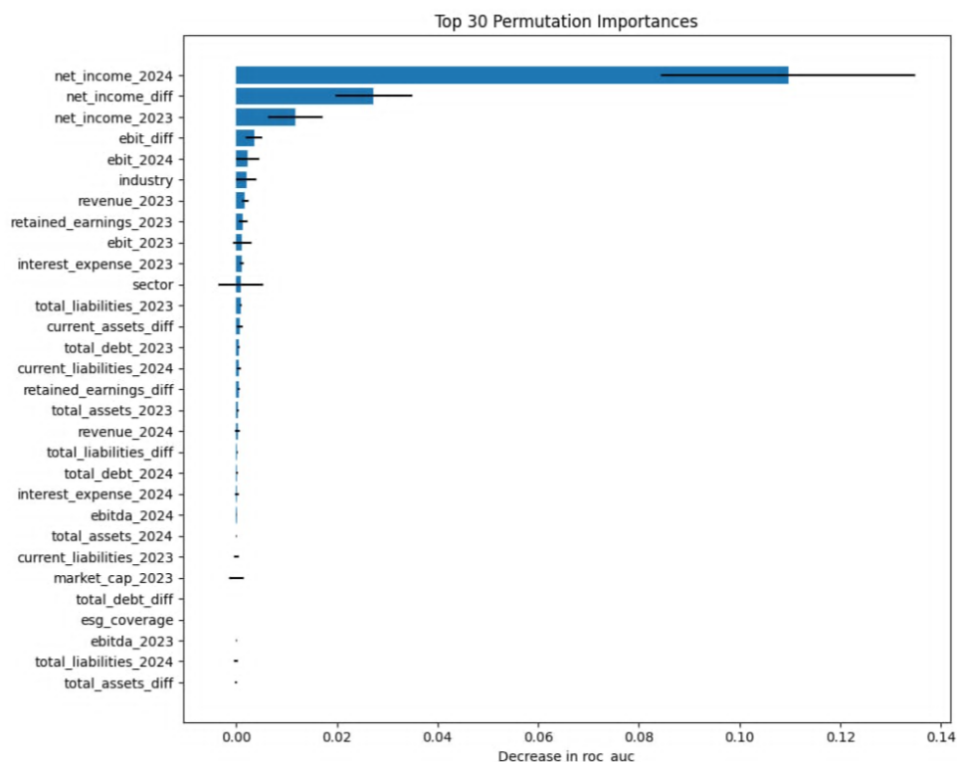
15 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį buvo įtraukti ESG rodikliai.

**15 lentelė.** Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas neigiamu grynuoju pelnu

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
RF+W	0,66	0,94	0,68	0,96	0,93	0,60	0,64	0,57
XGB+S	0,42	0,92	0,86	0,93	0,93	0,53	0,66	0,47
XGB+W	0,27	0,91	0,86	0,91	0,93	0,48	0,61	0,54
RF	0,44	0,93	0,73	0,95	0,92	0,57	0,64	0,51
XGB	0,40	0,91	0,82	0,92	0,92	0,49	0,61	0,56
RF+S	0,64	0,94	0,68	0,97	0,92	0,65	0,67	0,49
EE	0,61	0,92	0,73	0,94	0,92	0,53	0,62	0,51
LGBM+W	0,73	0,93	0,68	0,95	0,91	0,56	0,61	0,53
LGBM+S	0,62	0,93	0,73	0,95	0,91	0,55	0,63	0,46
LGBM	0,62	0,92	0,73	0,94	0,91	0,52	0,60	0,53
LG+S	0,61	0,87	0,55	0,90	0,82	0,34	0,42	0,37
LG	0,49	0,80	0,68	0,81	0,81	0,25	0,36	0,35
SVM + S (POLY)	0,33	0,77	0,68	0,78	0,78	0,22	0,33	0,26
SVM (POLY)	0,08	0,12	1,00	0,04	0,75	0,09	0,16	0,21

Kai neįtraukiami ESG duomenys, pagal AUC išskiriami trys geriausi modeliai: XGBoost su SMOTE ( $AUC = 0,93$ ), EasyEnsamble ( $AUC = 0,93$ ) ir RandomForest su svoriais ( $AUC = 0,93$ ). XGBoost turi geriausią jautrumo reikšmę ( $jautrumas = 0,86$ ) iš visų lyginamų modelių šio tikslo kintamojo atveju, kuomet nėra įtraukiami ESG duomenys, todėl geriausias pasirenkamas būtent XGBoost su SMOTE. Kai įtraukiami ESG duomenys, pagal AUC išskiriami trys geriausi modeliai: RandomForest su svoriais ( $AUC = 0,93$ ), XGBoost su SMOTE ( $AUC = 0,93$ ) ir XGBoost su svoriais ( $AUC = 0,93$ ). Iš šių trijų pirmiausia atmetamas RandomForest, nes jo jautrumas ( $jautrumas = 0,68$ ) kitų dviejų modelių atžvilgiu yra gerokai žemesnis. Iš dviejų likusių modelių, visų rezultatų, išskyrus PR-AUC, reikšmės yra geresnės XGBoost su SMOTE modelio.

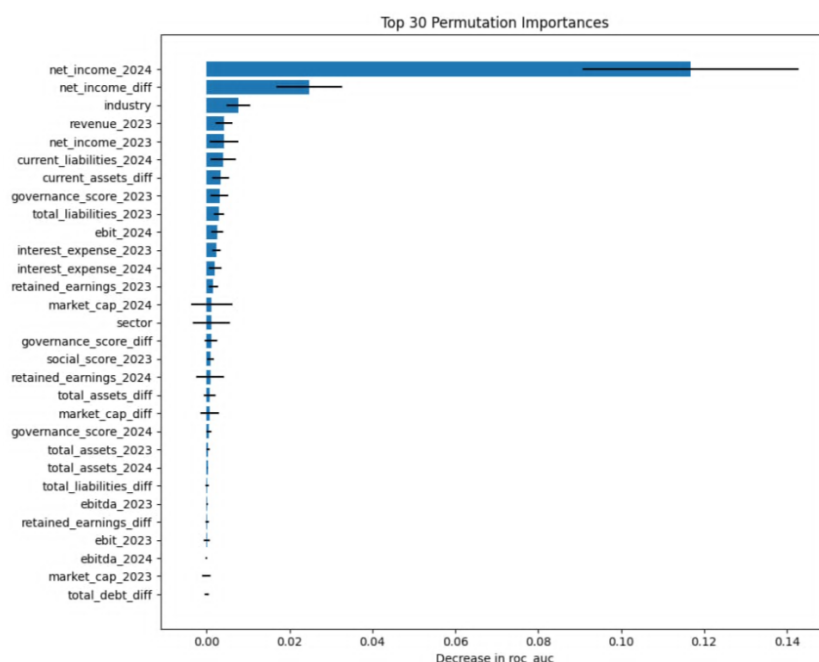
Kintamųjų svarba XGBoost modelyje, kai duomenų imtyje nėra ESG rodiklių, pateikiama 12 paveiksle.



12 pav. Kintamųjų svarba XGBoost (be ESG)

Iš 12 paveiksle pateikiamos diagramos galima daryti išvadą, kad įmonės gryniais pelnas 2024 metais daro didžiausią įtaką modelio prognozavimo tikslumui (sumaišius šio kintamojo reikšmės AUC reikšmė sumažėja per 0,11). Šiek tiek mažiau, tačiau vis tiek svarbus kintamasis yra grynojo pelno pokytis per 2023-2024 metus, kurio sumaišymas modelio prognozavimo AUC įvertį sumažina per 0,03.

Kintamųjų svarba XGBoost modelyje, kai duomenų imtyje įtraukti ESG rodikliai, pateikiama 13 paveiksle.

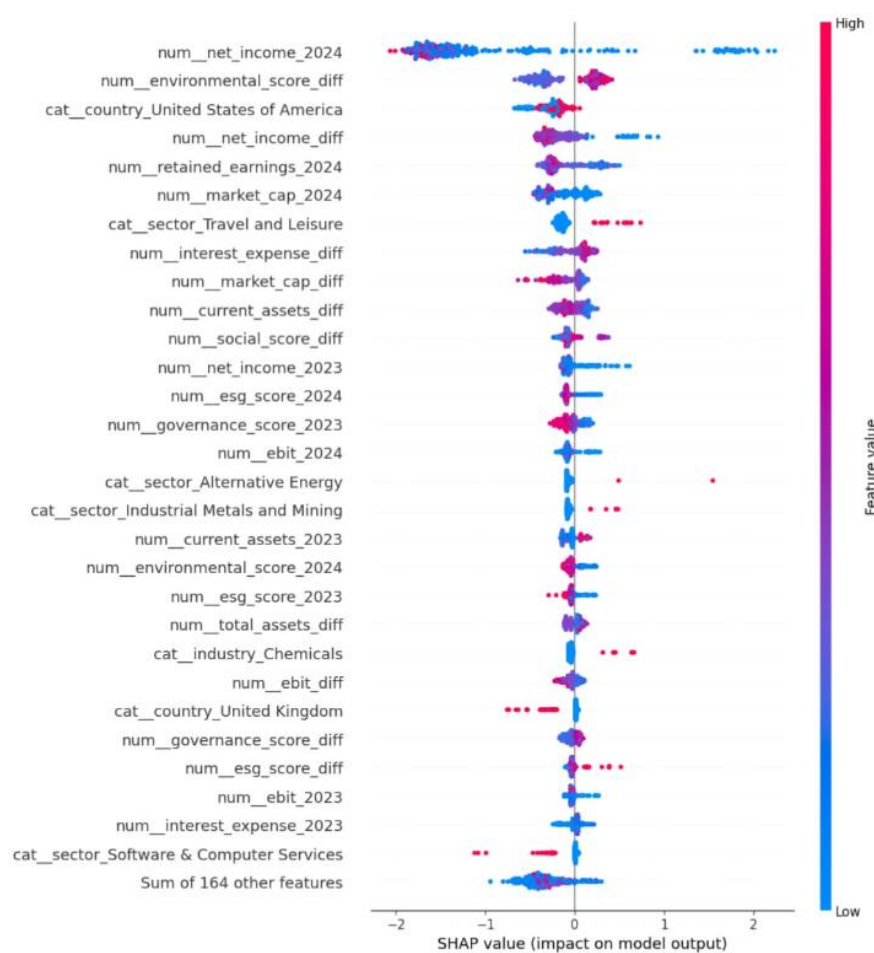


13 pav. Kintamųjų svarba XGBoost (su ESG)

Lyginant svarbius kintamuosius modelyje su ESG ir be ESG, galima pastebėti, kad ir įtraukus ESG svarbiausi, modelio prognozavimo tikslumą gerinantys kintamieji, yra grynasis pelnas 2024 metais ir grynojo pelno pokytis.

Pagal modelių rezultatų 14 ir 15 lenteles galima daryti išvadą, kad ESG duomenų įtraukimas modelio kokybės nepagerina. Tai patvirtina ir Vilkoksono statistikos p-reikšmė lygi 0,43. Remiantis ja, nulinė hipotezė apie medianų skirtumo lygumą nuliui priimama, todėl statistiškai reikšmingo skirtumo tarp šių modelių nėra.

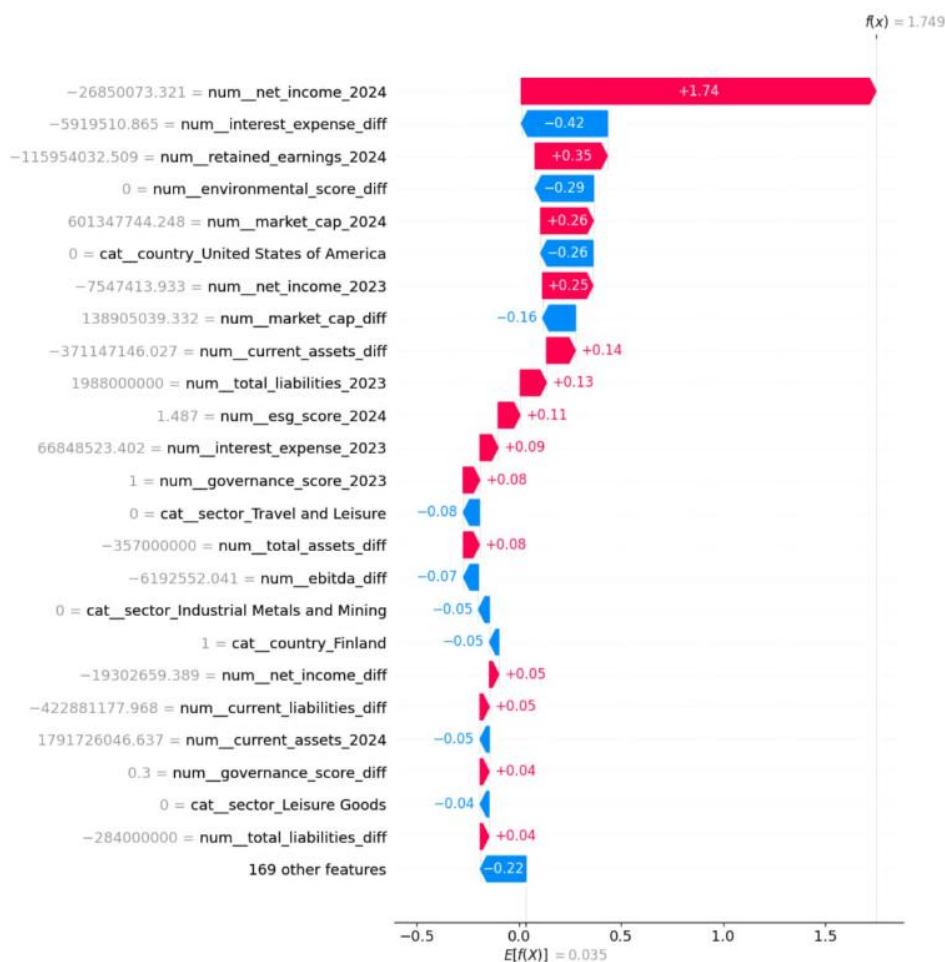
Tolimesnė tyrimo dalis apima šio tyrimo realią praktinę vertę – rezultatų interpretavimą. Interpretavimo metodai pritaikyti tik geriausiajam modeliui su ESG rodikliais – XGBoost. 14 paveiksle pateikiama SHAP metodo metu gauta „SHAP summary plot“ diagrama, parodanti kiekvieno kintamojo Shapley‘aus reikšmes.



14 pav. „SHAP summary plot“ XGBoost modeliui

Remiantis 14 paveikslu, galima daryti išvadą, kad didėjant įmonės grynajam pelnui mažėja tikimybė patirti finansinius sunkumus 2025 metais. Norėtusi išskirti ir antroje vietoje esantį aplinkosauginio rodiklio pokyčio kintamąjį. Kuo mažesnis jo pokytis, tuo mažesnė finansinių sunkumų tikimybė.

Toliau, 15 paveiksle, pateikimas pavyzdys įmonės, kuriai yra prognozuojami finansiniai sunkumai 2025 metais.



15 pav. „SHAP waterfall“ XGBoost modeliui

Šio stebinio atveju, neigiamas grynasis pelnas 2024 metais, modelio prognozę labiausiai priartino prie 1. Neigiamos palūkanų sąnaudos indikuoja mažėjančią finansinę naštą ir tai prisideda prie artėjimo link mažesnės finansinės rizikos.

DiCE metodo pritaikymui pasirinktas pavyzdys pateikiamas 16 paveiksle.

Original – changed columns only:										
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	
0	1.791726e+09	959373760.0	601347776.0	-26850074.0	-115954032.0	1.879505e+09	2.475000e+09	893000000.0	1.704000e+09	
Counterfactuals – changed columns + target:										
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	target
0	4773664.5	1816712.0	4467968.5	105263160.0	-2.035500e+10	1.294000e+09	2.266239e+09	0.0	1.174635e+09	0
0	4773664.5	1816712.0	4467968.5	105263160.0	-2.035500e+10	1.294000e+09	1.711100e+09	0.0	1.174635e+09	0
0	4773664.5	1816712.0	4467968.5	105263160.0	-2.035500e+10	6.314490e+08	8.580881e+06	0.0	1.860174e+06	0
0	4773664.5	1816712.0	4467968.5	105263160.0	-2.035500e+10	0.000000e+00	1.711100e+09	0.0	1.174635e+09	0
0	4773664.5	1816712.0	4467968.5	105263160.0	1.733000e+08	6.314490e+08	8.580881e+06	0.0	1.174635e+09	0

16 pav. DiCE metodo alternatyvūs scenarijai pasirinktai įmonei

Analizuojant sugeneruotus DiCE scenarijus matyti, kad modelio sprendimo pasikeitimas siejamas su keliomis pagrindinėmis finansinių rodiklių kryptimis. Visų pirma, reikšmingą įtaką turi trumpalaikių išpareigojimų mažėjimas, kas rodo, kad mažesnė trumpalaikė finansinė našta yra susijusi su mažesne prognozuojama rizika. Taip pat scenarijuose pastebimas teigiamas grynojo pelno pokytis, kuris iš neigiamos reikšmės tampa teigiamas ir tai rodo, kad pelningumo gerėjimas yra vienas svarbiausių veiksnių, lemiančių mažesnę finansinę riziką. Be to, daugelyje scenarijų mažėja bendras

įsiskolinimas, kas leidžia daryti išvadą, kad skolos mažinimas yra svarbus veiksnys gerinant įmonės finansinę būklę.

Apibendrinat rezultatus tikslo kintamajam, apskaičiuotam pagal grynojo pelno reikšmę („finansiniai rodikliai“ lygūs 1, kai grynasis pelnas yra neigiamas), galima daryti išvadą, kad ESG rodiklių įtraukimas statistiškai reikšmingai **nepagerino** mašininio mokymosi modelių prognozavimo tikslumo. Svarbiausi kintamieji modelių prognozėms yra 2024 metų grynasis pelnas ir pelno pokytis.

### 3.2.4. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą

Klasių pasiskirstymas:

- 126 įmonės patiriančios finansinius sunkumus 2025 metais;
- 1189 įmonės nepatiriančios finansinių sunkumų 2025 metais.

16 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį nebuvo įtraukti ESG rodikliai.

**16 lentelė.** Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifišku mas	AUC	Precizišku mas	F1	PR- AUC
RF	0,34	0,92	0,80	0,93	0,93	0,56	0,66	0,68
RF+W	0,28	0,92	0,88	0,93	0,93	0,56	0,69	0,70
RF+S	0,37	0,92	0,84	0,93	0,92	0,57	0,68	0,68
LGBM	0,44	0,92	0,80	0,94	0,92	0,57	0,67	0,64
XGB	0,45	0,93	0,72	0,95	0,91	0,60	0,65	0,65
LGBM+W	0,72	0,94	0,64	0,97	0,91	0,67	0,65	0,67
XGB+W	0,50	0,93	0,72	0,95	0,91	0,62	0,67	0,62
LGBM+S	0,44	0,92	0,72	0,94	0,90	0,56	0,63	0,66
XGB+S	0,53	0,93	0,68	0,96	0,90	0,63	0,65	0,64
EE	0,43	0,90	0,80	0,91	0,89	0,49	0,61	0,65
LG+S	0,60	0,84	0,60	0,87	0,81	0,32	0,42	0,40
LG	0,82	0,92	0,28	0,98	0,79	0,64	0,39	0,40
SVM +S (SIGMOID)	0,57	0,89	0,28	0,95	0,75	0,39	0,33	0,22
SVM (SIGMOID)	0,10	0,73	0,52	0,76	0,75	0,18	0,27	0,22

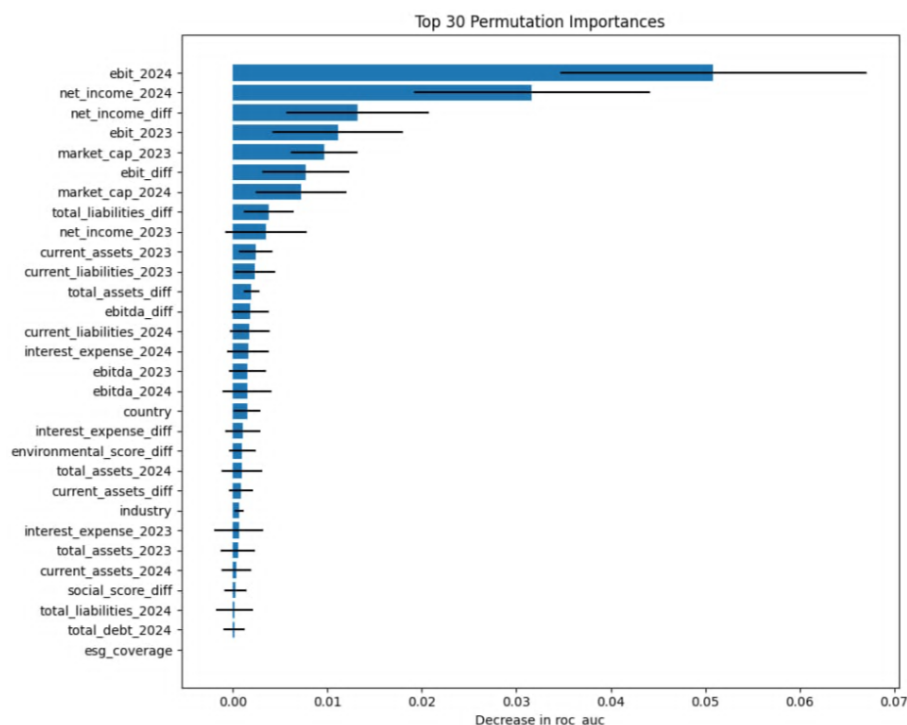
17 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį buvo įtraukti ESG rodikliai.

**17 lentelė.** Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas įmonės negebėjimu padengti skolą

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciškumas	F1	PR-AUC
RF	0,34	0,92	0,80	0,93	0,93	0,56	0,66	0,67
RF+S	0,45	0,93	0,72	0,95	0,92	0,60	0,65	0,68
XGB+S	0,42	0,91	0,72	0,93	0,92	0,53	0,61	0,68
RF+W	0,41	0,94	0,80	0,95	0,92	0,63	0,70	0,66
LGBM	0,52	0,92	0,72	0,94	0,91	0,55	0,62	0,62
LGBM+W	0,64	0,93	0,68	0,96	0,91	0,63	0,65	0,64
XGB	0,41	0,92	0,72	0,94	0,90	0,56	0,63	0,64
LGBM+S	0,43	0,91	0,76	0,92	0,90	0,51	0,61	0,63
XGB+W	0,45	0,92	0,76	0,93	0,90	0,54	0,63	0,64
EE	0,82	0,94	0,52	0,98	0,89	0,76	0,62	0,65
LG	0,70	0,88	0,52	0,92	0,81	0,39	0,45	0,42
LG+S	0,81	0,91	0,32	0,97	0,80	0,57	0,41	0,44
SVM (POLY)	0,14	0,84	0,36	0,90	0,74	0,26	0,31	0,32
SVM+S (SIGMOID)	0,50	0,49	0,76	0,46	0,60	0,13	0,22	0,13

Neįtraukiant ESG duomenų, geriausi modeliai pagal AUC yra RandomForest be jokio disbalanso sprendimo ( $AUC = 0,93$ ) ir RandomForest, kur sprendžiant disbalanso problemą įtraukti svoriai tikslo kintamojo klasėms ( $AUC = 0,93$ ). Nors abiejų modelių AUC lygūs, antrojo modelio su svoriais jautrumas ( $jautrumas = 0,88$ ) ir PR-AUC ( $PR - AUC = 0,7$ ) yra geriausi iš visų lyginamų modelių, tad pasirenkamas RandomForest su svoriais. Įtraukiant ESG duomenis dauguma modelių pagal AUC vertinami ganėtinai gerai, tačiau geriausiai vertinamą galima išskirti ir pasirinkti RandomForest be disbalanso problemos sprendimo, nes jo jautrumas ( $jautrumas = 0,8$ ) yra geriausias tarp visų modelių, o PR-AUC nuo geriausio atsilieka tik per 0,01.

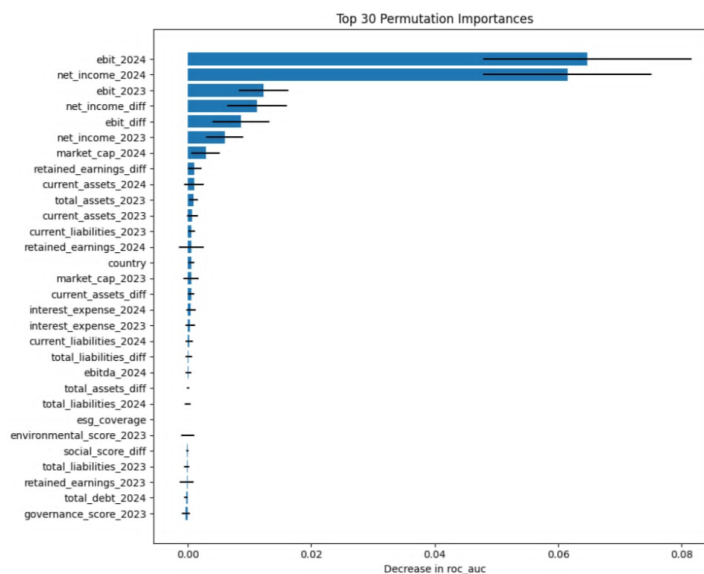
Kintamųjų svarba RandomForest modelyje su svoriais, kai duomenų imtyje nėra ESG rodiklių, pateikiama 18 paveiksle.



17 pav. Kintamų svarba RandomForest modelyje su svoriais (be ESG)

Iš 17 paveiksle pateikiamos diagramos galima daryti išvadą, kad EBIT rodiklis 2024 metais daro didžiausią įtaką modelio prognozavimo tikslumui (sumaišius šio kintamojo reikšmes AUC reikšmė sumažėja per 0,05). Šiek tiek mažiau, tačiau vis tiek svarbus kintamasis yra grynasis pelnas 2024 metais, kurio sumaišymas modelio prognozavimo AUC įvertį sumažina per 0,03.

Kintamųjų svarba XGBoost modelyje, kai duomenų imtyje įtraukti ESG rodikliai, pateikiama 18 paveiksle.

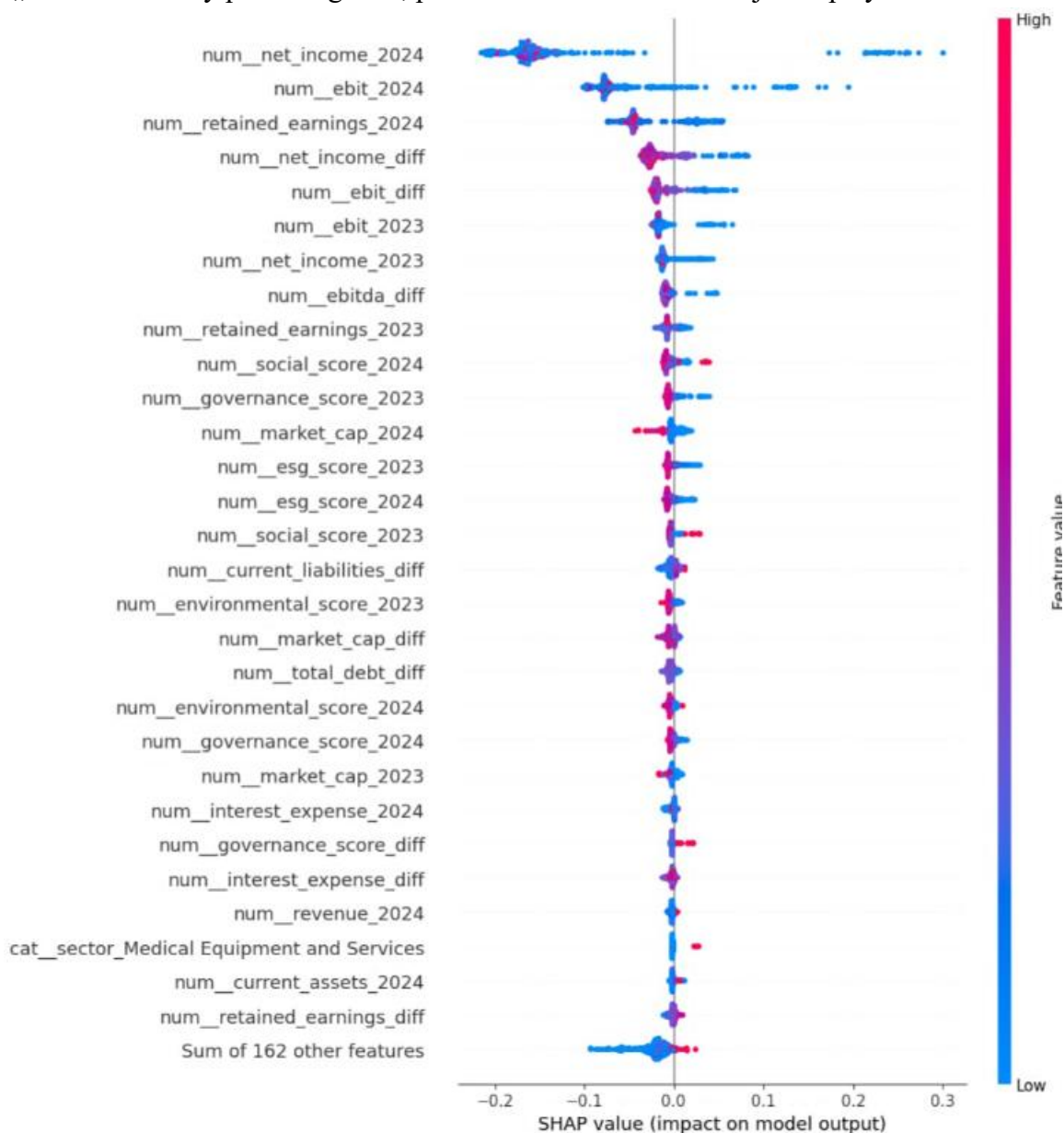


18 pav. Kintamų svarba RandomForest modelyje (su ESG)

Lyginant svarbius kintamuosius modelyje su ESG ir be ESG, galima pastebėti, kad ir įtraukus ESG svarbiausi, modelio prognozavimo tikslumą gerinantys kintamieji, yra EBIT ir grynasis pelnas 2024 metais.

Pagal modelių rezultatų 16 ir 17 lenteles galima daryti išvadą, kad ESG duomenų įtraukimas modelio kokybės nepagerina. Tai patvirtina ir Vilkoksono statistikos p-reikšmė lygi 1. Remiantis ja, nulinė hipotezė apie medianų skirtumo lygumą nuliui priimama, todėl statistiškai reikšmingo skirtumo tarp šių modelių nėra.

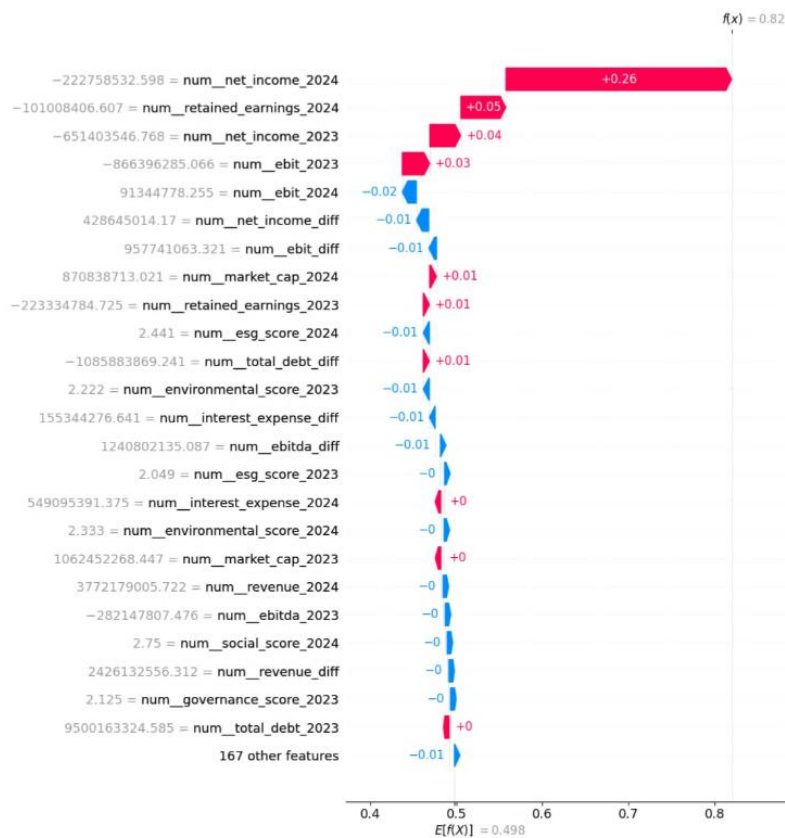
Tolimesnė tyrimo dalis – modelio rezultatų interpretavimas. Interpretavimo metodai pritaikyti tik geriausiajam modeliui su ESG rodikliais – RandomForest. 19 paveiksle pateikiama SHAP metodo metu gauta „SHAP summary plot“ diagrama, parodanti kiekvieno kintamojo Shapley‘aus reikšmes.



19 pav. „SHAP summary plot“ RandomForest modeliui

Remiantis 19 paveikslu, galima daryti išvadą, kad didėjant įmonės grynajam pelnui mažėja tikimybė patirti finansinius sunkumus 2025 metais. Norėtusi išskirti ir bendrą ESG balą tiek 2023, tiek 2024. Kuo mažesnė šio balo reikšmė, tuo didesnė finansinių sunkumų tikimybė.

Toliau, 20 paveiksle, pateikimas pavyzdys įmonės, kuriai yra prognozuojami finansiniai sunkumai 2025 metais.



20 pav. „SHAP waterfall“ RandomForest modeliui

Šio stebinio atveju, neigiamas grynasis pelnas 2024 metais, modelio prognozę labiausiai priartino prie 1. Neigiamas nepaskirstytas pelnas indikuoja, kad įmonė kurį laiką dirbo nuostolingai ir tai dar labiau sustiprina finansinės rizikos tikimybės augimą. Prie finansinių sunkumų reikšmingai prisideda ir grynasis pelnas bei EBIT 2023 metais.

DiCE metodo pritaikymui pasirinktas pavyzdys pateikiamas 21 paveiksle.

Original – changed columns only:										
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	
0	1.625990e+09	1.114617e+09	870838720.0	-222758528.0	-101008408.0	3.772179e+09	1.091050e+10	8.414280e+09	9.854008e+09	
Counterfactuals – changed columns + target:										
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	target
0	4773664.5	1816712.0	4467968.5	49083996.0	-1.385400e+10	8.940585e+08	8.580881e+06	2.973384e+09	3.391425e+09	0
0	4773664.5	1816712.0	4467968.5	49083996.0	-1.385400e+10	8.334406e+09	8.580881e+06	0.000000e+00	1.860174e+06	0
0	4773664.5	1816712.0	4467968.5	49083996.0	-1.385400e+10	0.000000e+00	8.580881e+06	2.973384e+09	1.860174e+06	0
0	4773664.5	1816712.0	4467968.5	49083996.0	-1.385400e+10	0.000000e+00	8.580881e+06	2.460000e+09	1.860174e+06	0
0	4773664.5	1816712.0	4467968.5	49083996.0	-1.385400e+10	0.000000e+00	1.048436e+10	8.767573e+09	3.980653e+09	0

21 pav. DiCE metodo alternatyvūs scenarijai pasirinktai įmonei

Visų pirma, pateiktuose scenarijuose pastebimas perėjimas iš neigiamo į teigiamą grynąjį pelną, o rodo, kad pelningumo atstatymas yra vienas svarbiausių veiksnių mažinant finansinę riziką. Taip pat reikšmingą įtaką turi trumpalaikių įsipareigojimų sumažėjimas, kuris susijęs su mažesne finansine našta ir geresne likvidumo situacija. Be to, scenarijuose dažnai mažėja bendras įsiskolinimas ir tai leidžia daryti išvadą, kad skolos mažinimas yra svarbus veiksnys gerinant įmonės finansinę būklę. Galima pastebėti pokyčius ir nepaskirstytame pelne, kurie iš neigiamos reikšmės tampa teigiami. Tai indikuoja ilgalaikės finansinės būklės gerėjimą ir sukaupto kapitalo atsiradimą.

Apibendrinat rezultatus tikslo kintamajam, apskaičiuotam pagal palūkanų padengiamumą („finansiniai rodikliai“ lygūs 1, kai  $\frac{EBIT}{\text{palūkanų sąnaudos}} < 1$ ), galima daryti išvadą, kad ESG rodiklių įtraukimas statistiškai reikšmingai **nepagerino** mašininio mokymosi modelių prognozavimo tikslumo. Svarbiausi kintamieji modelių prognozėms yra grynasis pelnas ir EBIT.

### 3.2.5. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra paremtas veiklos pelnu

Klasių pasiskirstymas:

- 81 įmonė patirianti finansinius sunkumus 2025 metais;
- 1234 įmonės nepatiriančios finansinių sunkumų 2025 metais.

18 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį nebuvo įtraukti ESG rodikliai.

**18 lentelė.** Modelių rezultatai (be ESG), kai tikslo kintamasis yra paremtas neigiamu veiklos pelnu

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
XGB+W	0,83	0,98	0,81	0,99	0,99	0,81	0,81	0,92
LGBM+S	0,61	0,98	0,88	0,99	0,99	0,82	0,85	0,83
RF+S	0,47	0,97	0,81	0,98	0,99	0,72	0,76	0,83
EE	0,57	0,97	0,94	0,97	0,98	0,68	0,79	0,72
LGBM	0,41	0,97	0,88	0,98	0,98	0,74	0,80	0,85
RF	0,31	0,97	0,81	0,98	0,98	0,76	0,79	0,84
XGB	0,51	0,97	0,94	0,97	0,98	0,65	0,77	0,78
LGBM+W	0,81	0,98	0,75	1,00	0,98	0,92	0,83	0,85
RF+W	0,28	0,97	0,81	0,98	0,97	0,72	0,76	0,80
XGB+S	0,52	0,97	0,94	0,98	0,97	0,71	0,81	0,79
LG+S	0,89	0,95	0,25	0,99	0,85	0,67	0,36	0,35
LG	0,91	0,95	0,25	1,00	0,83	0,80	0,38	0,30
SVM (POLY)	0,06	0,25	0,94	0,20	0,79	0,07	0,13	0,24
SVM+S (POLY)	0,43	0,88	0,56	0,90	0,76	0,27	0,37	0,21

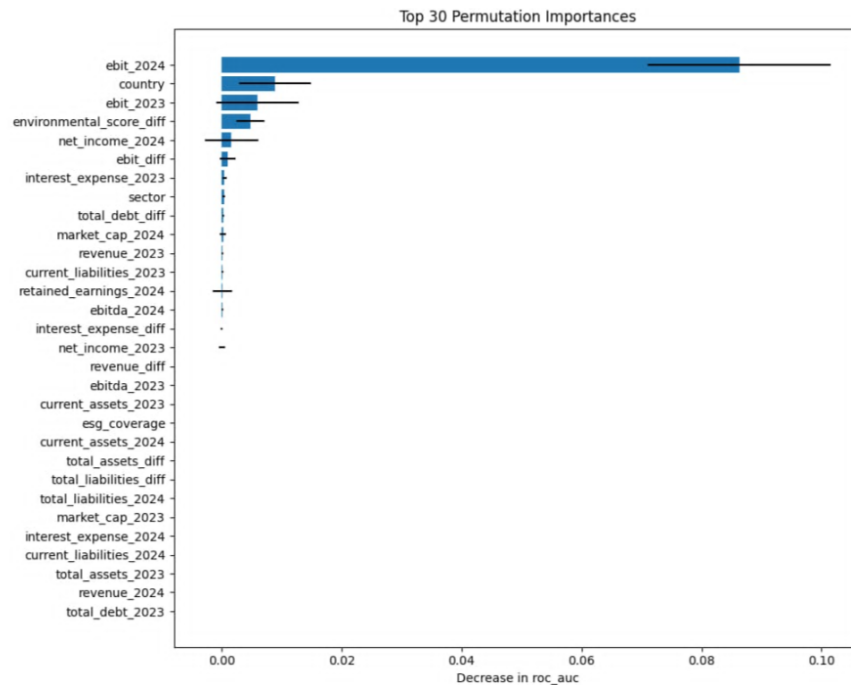
19 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį buvo įtraukti ESG rodikliai.

**19 lentelė.** Modelių rezultatai (su ESG), kai tikslo kintamasis yra paremtas neigiamu veiklos pelnu

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Preciziškumas	F1	PR-AUC
LGBM+S	0,56	0,99	1,00	0,99	0,99	0,84	0,91	0,88
XGB+S	0,59	0,98	0,94	0,98	0,99	0,79	0,86	0,89
LGBM+W	0,68	0,98	0,94	0,98	0,99	0,75	0,83	0,87
XGB	0,81	0,98	0,75	0,99	0,99	0,86	0,80	0,87
RF+W	0,70	0,98	0,75	1,00	0,99	0,92	0,83	0,89
XGB+W	0,39	0,97	0,88	0,98	0,99	0,70	0,78	0,88
RF+S	0,45	0,97	0,88	0,98	0,99	0,74	0,80	0,83
EE	0,61	0,97	0,88	0,98	0,98	0,70	0,78	0,82
LGBM	0,30	0,97	0,88	0,98	0,98	0,70	0,78	0,76
RF	0,29	0,97	0,81	0,98	0,98	0,76	0,79	0,82
LG+S	0,72	0,92	0,50	0,95	0,86	0,38	0,43	0,36
LG	0,76	0,92	0,38	0,95	0,85	0,33	0,35	0,31
SVM (POLY)	0,06	0,26	1,00	0,21	0,84	0,08	0,14	0,25
SVM+S (POLY)	0,68	0,91	0,50	0,94	0,82	0,33	0,40	0,23

Neįtraukiant ESG duomenų dauguma modelių pagal AUC vertinami ganėtinai gerai, tad lieka geriausią modelį rinkti pagal jautrumą. Trys modeliai pasiekia dideles jautrumo reikšmes neprarasdami gero AUC rezultato: EasyEnsamble (*jautrumas* = 0,94), XGBoost be disbalanso sprendimo (*jautrumas* = 0,94) ir XGBoost naudojantis SMOTE (*jautrumas* = 0,94). Kadangi visų trijų modelių AUC labai panašus, jautrumas identiškas, tai pasirenkamas su didžiausia PR-AUC reikšme - XGBoost su SMOTE. Kai įtraukiami ESG duomenys vėl yra daug modelių turinčių gera AUC, tačiau vienas labai išsiskiria iš kitų - LightGBM su SMOTE. Šio modelio jautrumas lygus 1, kiti rezultatai taip pat aukščiau ar vieni aukštesnių, tad kaip geriausias pasirenkamas šis, kuris nepraleidžia nei vieno atvejo kuomet įmonė patirs finansinius sunkumus.

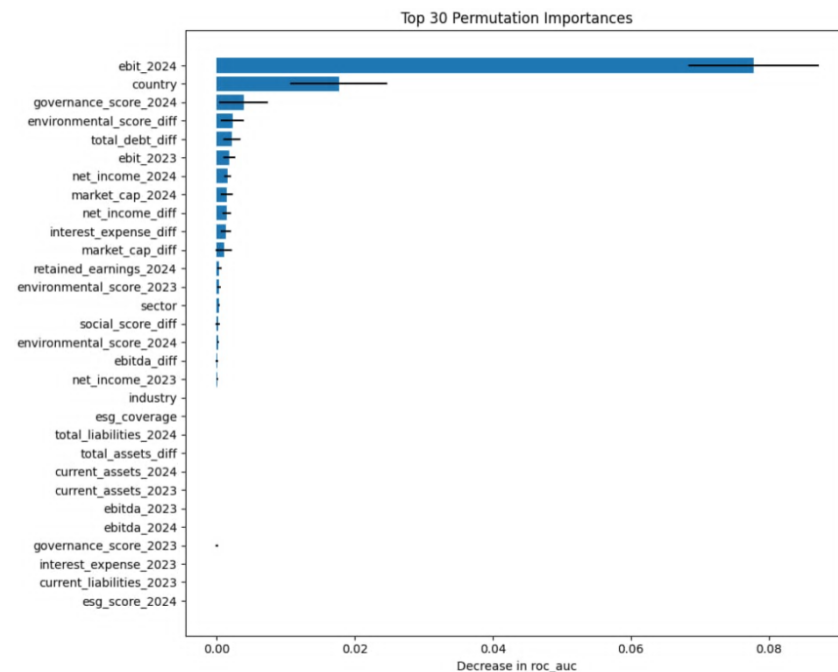
Kintamųjų svarba XGBoost su SMOTE modelyje, kai duomenų imtyje nėra ESG rodiklių, pateikiama 22 paveiksle.



22 pav. Kintamų svarba XGBoost su SMOTE modelyje (be ESG)

Iš 22 paveiksle pateikiamos diagramos galima daryti išvadą, kad EBIT rodiklis 2024 metais daro didžiausią įtaką modelio prognozavimo tikslumui (sumaišius šio kintamojo reikšmes AUC reikšmė sumažėja per 0,09). Kitų kintamųjų, pavyzdžiui, šalies, EBIT 2023, aplinkosauginio ESG rodiklio pokytis, svarba yra labai minimali, šių kintamųjų sumaišymas modelio prognozavimo AUC įvertį vidutiniškai sumažina vos per 0,01.

Kintamųjų svarba LightGBM su SMOTE modelyje, kai duomenų imtyje įtraukti ESG rodikliai, pateikiama 23 paveiksle.

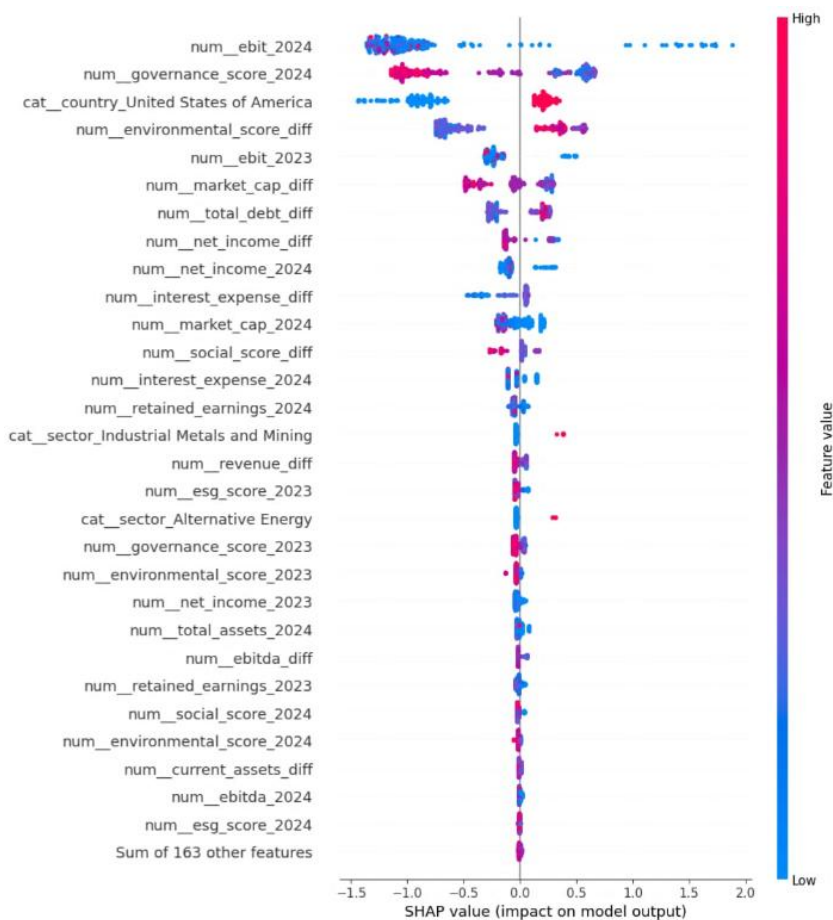


23 pav. Kintamų svarba LightGBM su SMOTE modelyje (su ESG)

Lyginant svarbius kintamuosius modelyje su ESG ir be ESG, galima pastebėti, kad ir įtraukus ESG svarbiausias, pagrindinis, modelio prognozavimo tikslumą gerinantis kintamasis, yra EBIT 2024 metais.

Pagal modelių rezultatų 18 ir 19 lenteles galima daryti išvadą, kad ESG duomenų įtraukimas modelio kokybę išlaiko tokią pačią. Tai patvirtina ir Vilkoksono statistikos p-reikšmė lygi 0,1875. Remiantis ja, nulinė hipotezė apie medianų skirtumo lygumą nuliui priimama, todėl statistiškai reikšmingo skirtumo tarp šių modelių nėra.

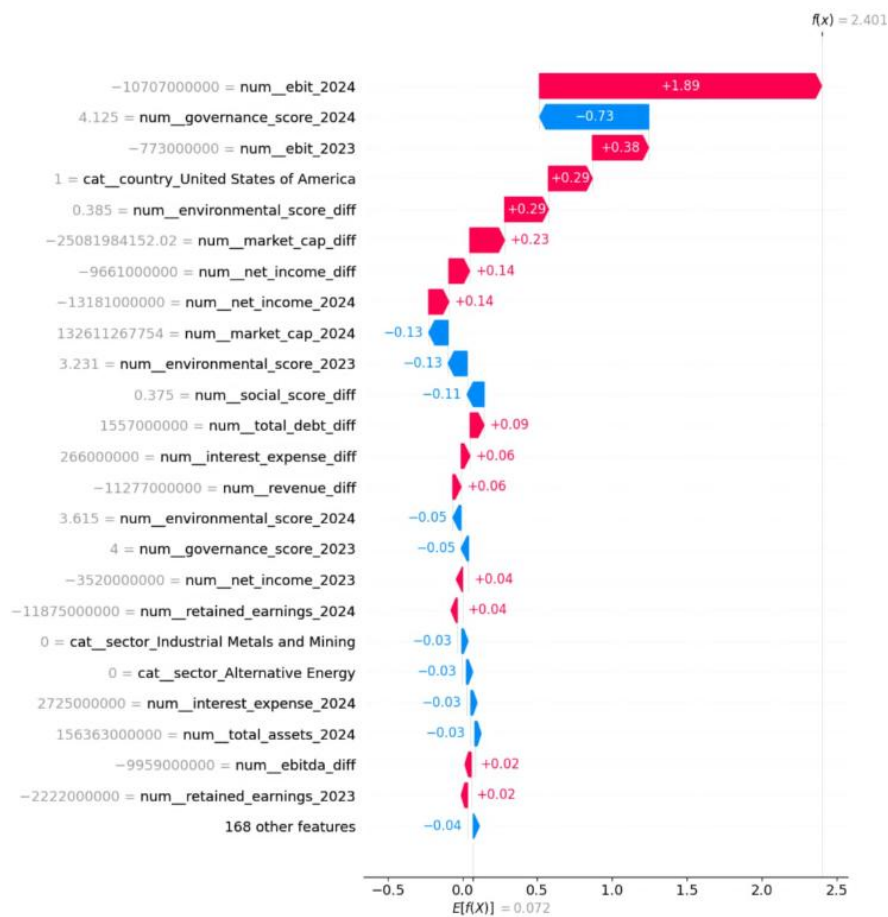
Tolimesnė tyrimo dalis – modelio rezultatų interpretavimas. Interpretavimo metodai pritaikyti tik geriausiajam modeliui su ESG rodikliais – LightGBM su SMOTE. X paveiksle pateikiama SHAP metodo metu gauta „SHAP summary plot“ diagrama, parodanti kiekvieno kintamojo Shapley'aus reikšmes.



24 pav. „SHAP summary plot“ LightGBM su SMOTE modeliui

Remiantis 24 paveikslu, galima daryti išvadą, kad didėjant įmonės EBIT reikšmei mažėja tikimybė patirti finansinius sunkumus 2025 metais. Šiame modelyje išsiskiria valdysenos rodiklis 2024, didėjant jo reikšmei mažėja finansinių sunkumų tikimybė, o tai rodo, kad aiški įmonės valdymo struktūra ir principai prisideda prie finansinės rizikos mažinimo.

Toliau, 25 paveiksle, pateikimas pavyzdys įmonės, kuriai yra prognozuojami finansiniai sunkumai 2025 metais.



25 pav. „SHAP waterfall“ LightGBM su SMOTE modeliui

SHAP analizė rodo, kad didžiausią įtaką šio stebinio prognozei turi veiklos pelningumo rodikliai. Ypač reikšmingas yra neigiamas EBIT 2024 metais, kuris yra pagrindinis veiksnys, lemiantis prognozuojamą finansinę riziką. Taip pat galima pastebėti, kad blogėjantys finansinių rodiklių pokyčiai, tokie kaip grynojo pelno ir skolos didėjimas, dar labiau didina riziką. Kai kurie ESG rodikliai, ypač valdysenos aspektai, prisideda prie rizikos mažinimo, tačiau jų poveikis nėra pakankamas kompensuoti neigiamą pelningumo įtaką.

Šiam modeliui DiCE metodas nepateikė jokių alternatyvių scenarijų, kurie pakeistų modelio prognozę. Taip įvyko dėl to, nes modelis yra labai „užtikrintas“ savo prognozėmis, modelio sprendimas yra stabilus ir norint pakeisti prognozės rezultatą reikėtų atlikti reikšmingus kintamųjų reikšmių pokyčius, kurie viršytų minimalias korekcijas.

Apibendrinat rezultatus tikslo kintamajam, apskaičiuotam pagal veiklos pelno reikšmę („finansiniai rodikliai“ lygūs 1, kai  $EBIT < 0$ ), galima daryti išvadą, kad ESG rodiklių įtraukimas statistiškai reikšmingai **nepagerino** mašininio mokymosi modelių prognozavimo tikslumo. Svarbiausias kintamasis modelių prognozėms yra EBIT. Šiuo atveju EBIT yra ne tik pagrindinis paaiškinantis kintamasis, bet ir tiesiogiai susijęs su pačiu tikslo kintamuoju ir jo apibrėžimu, todėl EBIT dominavimas yra tikėtinas. Dėl šios priežasties ESG rodiklių įtaka prognozavimo tikslumui gali būti neatskleista.

### 3.2.6. Finansinių sunkumų prognozavimas, kai tikslo kintamasis yra kombinuotas

Klasių pasiskirstymas:

- 338 įmonės patiriančios finansinius sunkumus 2025 metais;
- 977 įmonės nepatiriančios finansinių sunkumų 2025 metais.

20 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį nebuvo įtraukti ESG rodikliai.

**20 lentelė.** Modelių rezultatai (be ESG), kai tikslo kintamasis yra kombinuotas

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Precižiškumas	F1	PR-AUC
EE	0,52	0,86	0,74	0,90	0,90	0,72	0,73	0,78
XGB+S	0,49	0,87	0,72	0,92	0,90	0,77	0,74	0,77
LGBM+W	0,46	0,87	0,76	0,91	0,90	0,74	0,75	0,77
RF+W	0,56	0,87	0,72	0,93	0,90	0,78	0,75	0,78
XGB+W	0,66	0,87	0,69	0,94	0,90	0,80	0,74	0,78
LGBM	0,33	0,87	0,78	0,90	0,90	0,74	0,76	0,79
RF	0,44	0,85	0,75	0,88	0,89	0,69	0,72	0,77
LGBM+S	0,33	0,84	0,76	0,87	0,89	0,68	0,72	0,78
RF+S	0,48	0,87	0,74	0,92	0,89	0,77	0,75	0,79
XGB	0,60	0,86	0,71	0,91	0,89	0,74	0,72	0,75
LG+S	0,48	0,84	0,76	0,87	0,88	0,67	0,71	0,75
LG	0,50	0,84	0,75	0,87	0,88	0,66	0,70	0,73
SVM (POLY)	0,26	0,73	0,72	0,73	0,78	0,48	0,58	0,59
SVM+S (POLY)	0,33	0,73	0,75	0,73	0,78	0,49	0,59	0,63

21 lentelėje pateikiami mašininio mokymosi modelių prognozavimo gerumo įvertinimai, kai į duomenų imtį buvo įtraukti ESG rodikliai.

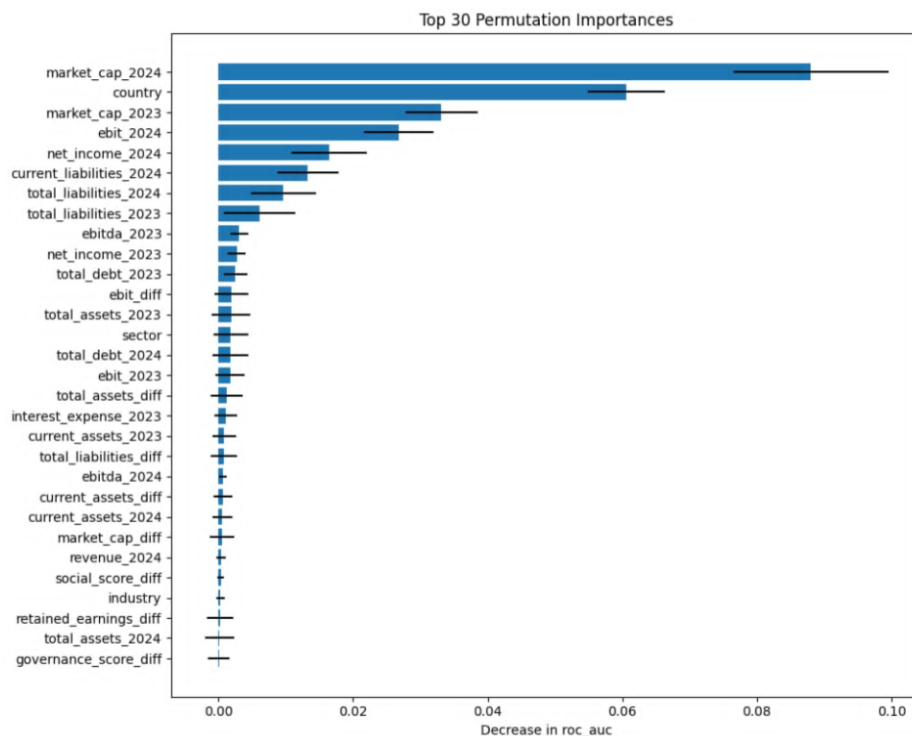
**21 lentelė.** Modelių rezultatai (su ESG), kai tikslo kintamasis yra kombinuotas

Modelis	Sprendimo riba	Tikslumas	Jautrumas	Specifiškumas	AUC	Precižiškumas	F1	PR-AUC
LGBM+W	0,47	0,87	0,76	0,90	0,90	0,73	0,75	0,77
EE	0,51	0,85	0,76	0,88	0,90	0,68	0,72	0,77
RF+W	0,60	0,87	0,68	0,94	0,90	0,81	0,74	0,79
RF+S	0,49	0,87	0,72	0,92	0,90	0,75	0,74	0,80
LGBM+S	0,54	0,86	0,71	0,92	0,89	0,75	0,73	0,78
XGB	0,69	0,87	0,65	0,94	0,89	0,80	0,72	0,76
XGB+W	0,69	0,87	0,65	0,94	0,89	0,80	0,72	0,76

XGB+S	0,42	0,86	0,75	0,89	0,88	0,71	0,73	0,75
LGBM	0,58	0,85	0,72	0,90	0,88	0,71	0,72	0,73
LG	0,48	0,84	0,81	0,86	0,88	0,66	0,73	0,73
RF	0,51	0,85	0,69	0,90	0,88	0,71	0,70	0,77
LG+S	0,51	0,83	0,72	0,87	0,88	0,66	0,69	0,74
SVM (POLY)	0,31	0,81	0,65	0,87	0,81	0,63	0,64	0,63
SVM+S (POLY)	0,42	0,80	0,68	0,84	0,78	0,60	0,63	0,60

Neįtraukiant ESG duomenų dauguma modelių pagal AUC vertinami labai panašiai: didžioji dalis AUC reikšmių yra tarp 0,88 ir 0,90. Iš šios daugumos išsiskiria modelis, turintis geriausią jautrumą (*jautrumas* = 0,78) bei PR-AUC reikšmę (*PR – AUC* = 0,79) - LightGBM be papildomo duomenų disbalanso problemos sprendimo. Kai įtraukiami ESG duomenys taip pat 12 iš 14 modelių turi AUC reikšmes tarp 0,88 ir 0,90. Šiuo atveju renkantis modelį tolimesniam tyrimui reikia pasirinkti vidutinį variantą: nesirenkamas didžiausią jautrumą turinti modelis tam, kad neprarasti per daug preciziškumo. PR-AUC atžvilgiu išsiskiria vienas modelis, turintis geriausią reikšmę - RandomForest su SMOTE (*PR – AUC* = 0,8).

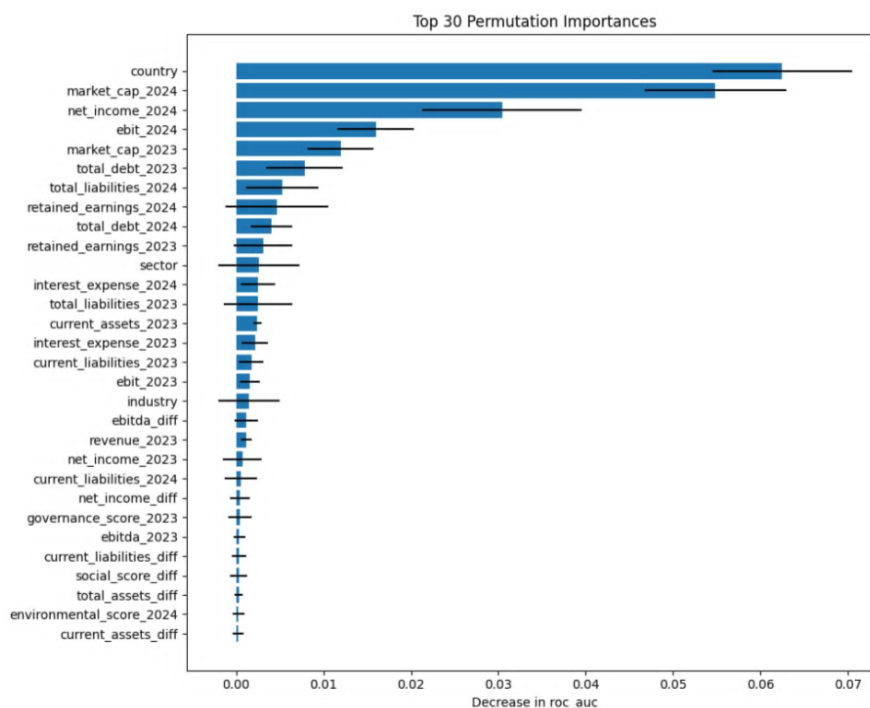
Kintamųjų svarba LightGBM modelyje, kai duomenų imtyje nėra ESG rodiklių, pateikiama 26 paveiksle.



26 pav. Kintamųjų svarba LightGBM modelyje (be ESG)

Iš 26 paveiksle pateikiamos diagramos galima daryti išvadą, kad rinkos kapitalizacijos vertė 2024 metais daro didžiausią įtaką modelio prognozavimo tikslumui (sumaišius šio kintamojo reikšmes AUC reikšmė sumažėja per 0,09). Taip pat, svarbus kintamasis yra šalis, kurio sumaišymas modelio prognozavimo AUC įvertį sumažina per 0,06.

Kintamųjų svarba RandomForest su SMOTE modelyje, kai duomenų imtyje įtraukti ESG rodikliai, pateikiama 27 paveiksle.

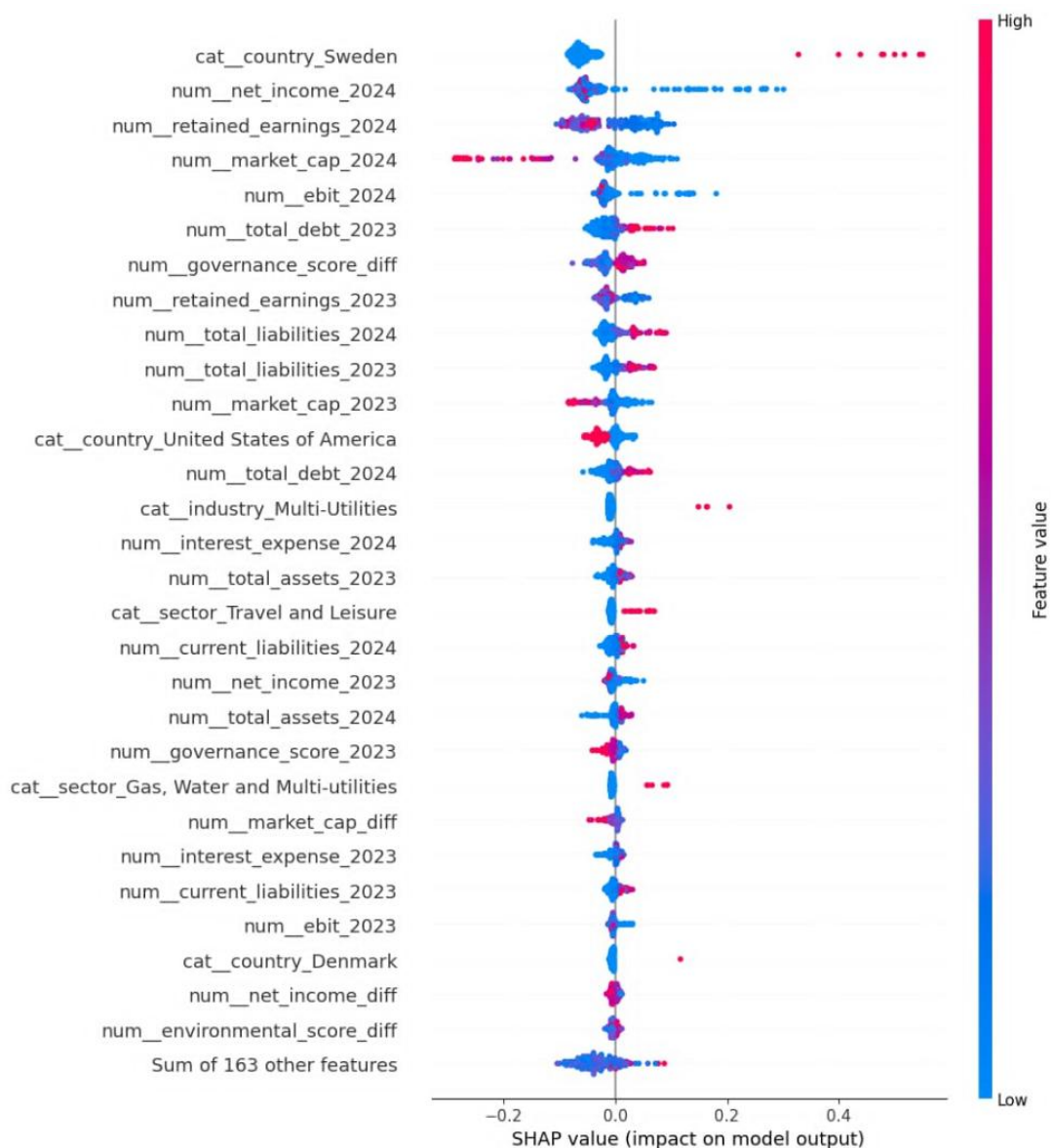


27 pav. Kintamųjų svarba RandomForest su SMOTE modelyje (su ESG)

Lyginant svarbius kintamuosius modelyje su ESG ir be ESG, galima pastebėti, kad ir įtraukus ESG svarbiausias modelio prognozavimo tikslumą gerinantis kintamasis, yra rinkos kapitalizacija ir grynasis pelnas 2024 metais.

Pagal modelių rezultatų 20 ir 21 lenteles galima daryti išvadą, kad ESG duomenų įtraukimas modelio kokybę išlaiko beveik tokią pačią. Tai patvirtina ir Vilkoksono statistikos p-reikšmė lygi 0,4375. Remiantis ja, nulinė hipotezė apie medianų skirtumo lygumą nuliui priimama, todėl statistiškai reikšmingo skirtumo tarp šių modelių nėra.

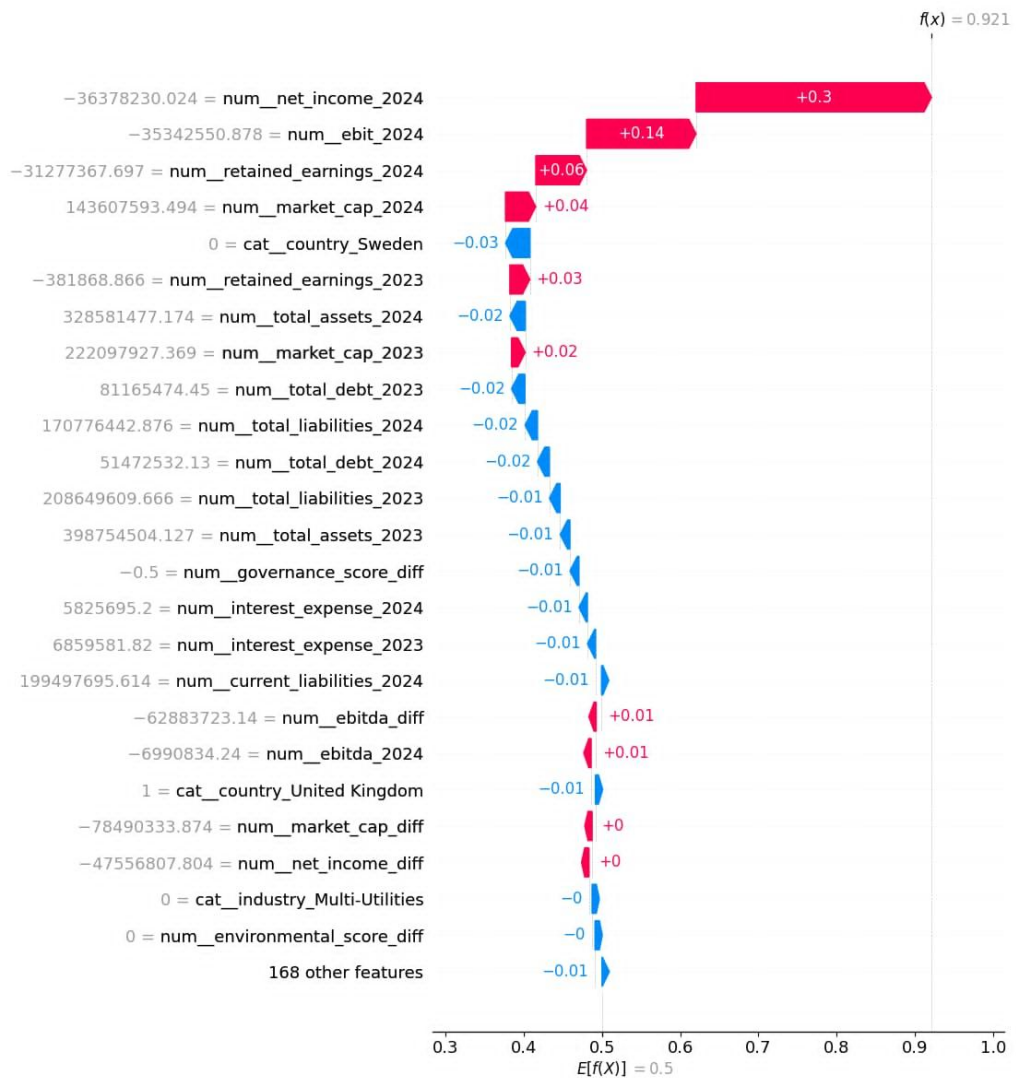
Tolimesnė tyrimo dalis – modelio rezultatų interpretavimas. Interpretavimo metodai pritaikyti tik geriausiajam modeliui su ESG rodikliais – RandomForest su SMOTE. 28 paveiksle pateikiama SHAP metodo metu gauta „SHAP summary plot“ diagrama, parodanti kiekvieno kintamojo Shapley'aus reikšmes.



**28 pav.** „SHAP summary plot“ LightGBM su SMOTE modeliui

Remiantis 28 paveikslu, galima daryti išvadą, kad didėjant įmonės grynojo pelno ir rinkos kapitalizacijos reikšmei mažėja tikimybė patirti finansinius sunkumus 2025 metais. Šiame modelyje ypatingai neišsiskiria nei vienas ESG rodiklis.

Toliau, 29 paveiksle, pateikimas pavyzdys įmonės, kuriai yra prognozuojami finansiniai sunkumai 2025 metais.



29 pav. „SHAP waterfall“ LightGBM su SMOTE modeliui

SHAP analizė rodo, kad didžiausią įtaką šio stebinio prognozei turi grynojo pelno ir EBIT rodikliai. Ypač reikšmingas yra neigiamas grynasis pelnas 2024 metais, kuris yra pagrindinis veiksnys, lemiantis prognozuojamą finansinę riziką.

DiCE metodo pritaikymui pasirinktas pavyzdys pateikiamas 30 paveiksle.

Original – changed columns only:									
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024
0	303195072.0	199497696.0	143607600.0	-36378232.0	-31277368.0	767826624.0	328581472.0	51472532.0	170776448.0

Counterfactuals – changed columns + target:										
	current_assets_2024	current_liabilities_2024	market_cap_2024	net_income_2024	retained_earnings_2024	revenue_2024	total_assets_2024	total_debt_2024	total_liabilities_2024	target
0	4773664.5	1816712.0	4467968.5	18962758.0	-2.035500e+10	0.0	204015472.0	47690168.0	1860174.0	0
0	4773664.5	1816712.0	4467968.5	18962758.0	-2.035500e+10	194423840.0	204015472.0	47690168.0	102166776.0	0
0	4773664.5	1816712.0	4467968.5	18962758.0	-2.035500e+10	215986112.0	10813348.0	0.0	1860174.0	0
0	4773664.5	1816712.0	4467968.5	18962758.0	-4.939097e+07	0.0	10813348.0	0.0	54800000.0	0
0	4773664.5	1816712.0	4467968.5	25398082.0	-2.035500e+10	0.0	496207168.0	0.0	1860174.0	0

30 pav. DiCE metodo alternatyvūs scenarijai pasirinktai įmonei

Visų pirma, pateiktuose scenarijuose pastebimas perėjimas iš neigiamo į teigiamą grynąjį pelną, o rodo, kad pelningumo atstatymas yra vienas svarbiausių veiksnių mažinant finansinę riziką. Be to,

scenarijuose dažnai mažėja bendras įsiskolinimas ir tai leidžia daryti išvadą, kad skolos mažinimas yra svarbus veiksnys gerinant įmonės finansinę būklę.

Apibendrinat rezultatus tikslo kintamajam, apskaičiuotam pagal veiklos pelno reikšmę („finansiniai iššūkiai“ lygūs 1, kai  $Altman'o Z'' < 1,1$  arba  $NetIncome < 0$  arba  $InterestCoverage < 0$ ), galima daryti išvadą, kad ESG rodiklių įtraukimas statistiškai reikšmingai **nepagerino** mašininio mokymosi modelių prognozavimo tikslumo. Svarbiausi kintamieji modelių prognozėms yra rinkos kapitalizacija ir EBIT. ESG rodiklių įtaka prognozavimo rezultatams nedarė ypatingos įtakos.

Išbandžius 5 skirtingus tikslo kintamųjų apibrėžimus, apmokinus keletą pagrindinių dvireikšmio kintamojo klasifikavimo modelių, atlikus tyrimą ir apžvelgus gautus rezultatus vieno atsakymo nėra. Prognozavimo modelių pasisekimas labai priklauso nuo tikslo kintamojo apibrėžimo. Iš pažiūros geriausias modelis buvo su tikslo kintamuoju apibrėžtu remiantis, kad 2025 metų veiklos pelnas bus neigiamas, tačiau pagal šį tikslo kintamąjį tokiai užduočiai trūksta kompleksškumo, nes modelis sprendimus atliko pagrinde remdamasis tik 2023 ir 2024 metų EBIT duomenimis. Gerokai naudingesni ir prasmingesni modeliai buvo tie, kurių tikslo kintamieji buvo apibrėžiami kompleksiškesnėmis sąlygomis: Altman'o Z'', kombinuotu įverčiu. Naudojant pastaruosius tikslo kintamuosius buvo pastebėtas didesnis kiekis nepriklausomų kintamųjų įsitraukimas, o tai leidžia daryti prielaidą, kad tokie modeliai labiau subalansuoti.

Nei vienu atveju tarp geriausio modelio su ESG duomenimis ir geriausio modelio be ESG duomenų nepavyko įrodyti, kad ESG duomenų įtraukimas į modelį duoda statistiškai reikšmingą pokytį, tačiau skirtingais atvejais ESG rodikliai buvo sutinkami SHAP analizės metu kaip suteikiantys papildomos prognostinės informacijos.

## Išvados

1. Atlikta literatūros analizė atskleidė, kad aplinkosauginių, socialinės atsakomybės ir valdysenos rodiklių įtaka finansinei rizikai nėra plačiai iširta, tačiau pastebima, kad tvarumo veiksniai veikia rizikos lygį. Problema šioje srityje yra ESG rodiklių grynumas ir finansinės rizikos matavimas. Priklausomai nuo šaltinio, iš kurio šie tvarumo rodikliai yra gaunami, galima pasiekti skirtingus prognozavimo rezultatus (tik apie 15% rodiklių reikšmių sutampa tarp juos teikiančių agentūrų). Baigiamajame darbe finansinė rizika prognozuojama kaip tikimybė įmonei patirti finansinius sunkumus. Finansiniai sunkumai literatūroje yra apibrėžiami daug skirtingų būdų, todėl šio darbo tyrime pasirinkta analizuoti 5 skirtingus tikslo kintamuosius, o visi aplinkosauginiai, socialinės atsakomybės ir valdysenos rodikliai yra iš to paties duomenų šaltinio.
2. Šio darbo tyrimo duomenų imtyje, pagal visus analizuojamus tikslo kintamuosius, yra klasių disbalansas – tik apie 10% visų įmonių yra priskiriamos, kaip patiriančios finansinius sunkumus 2025 metais. Ši problema išskiriama ir literatūroje, kurioje nurodoma, kad geriausiai šią problemą išsprendžia SMOTE metodo panaudojimas ir gradientinio stiprinimo bei ansambliniai mašininio mokymosi algoritmai. Remiantis šiomis literatūros analizės išvadomis tyrime buvo sukurti modeliai pagal 6 skirtingus mašininio mokymosi metodus (logistinė regresija, atraminiai vektoriai, XGBoost, LightGBM, atsitiktiniai miškai, EasyEnsemble), kiekvienam iš jų atliekant hiperparametrų optimizaciją.
3. Lyginant sukurtų modelių rezultatus pastebėta, jog stabiliausiai veikė ir dažniausiai gerus rezultatus demonstravo XGBoost, LightGBM ir atsitiktinių miškų modeliai, tačiau rezultatų skirtumai priklausomai nuo pasirinkto tikslo kintamojo skyrėsi reikšmingai, pvz. tikslo kintamajam, paremtam neigiamu veiklos pelnu, pavyko sudaryti modelį įtraukiant ESG rodiklius, kuris pasiekė 99% tikslumą ir 100% jautrumą testavimo imtyje, o kombinuotam tikslo kintamajam, geriausias sudarytas modelis pasiekė tik 87% tikslumą ir 72% jautrumą. Šis tyrimas dar labiau išryškino finansinių sunkumų apibrėžimų nevienodumo pasekmes prognozuojant finansinę riziką. Statistinio reikšmingumo testų, pritaikytų prognozavimų tikslumų skirtumams įvertinti, tiek įtraukiant ESG, tiek neįtraukiant ESG rodiklių p-reikšmės visada buvo didesnės nei 0,05. Dėl to, negalima atmesti nulinės hipotezės, kad modelio be ESG rodiklių prognozavimo tikslumas statistiškai reikšmingai nesiskiria nuo modelio, kuriame įtraukiami ESG rodikliai, pavyzdžiui, geriausi modeliai (su ESG ir be ESG rodiklių) pagal kombinuotą tikslo kintamąjį pasiekė tokį patį prognozavimo tikslumą - 87%.
4. Kiekvienam tikslo kintamajam buvo parinktas geriausiai finansinę riziką prognozuojantis modelis į kurį įtraukti aplinkosaugos, socialinės atsakomybės ir valdysenos rodikliai. Pritaikyti interpretavimo metodai (SHAP ir DiCE) kiekvienu atveju leido nustatyti svarbiausius kintamuosius ir jų įtaką finansinei rizikai. Tyrimo metu pastebėta, kad visuose analizuotuose modeliuose didesnės valdysenos rodiklio reikšmės buvo susijusios su mažesne finansinių sunkumų tikimybe (Shapley'aus reikšmė vidutiniškai -0,1). Aplinkosaugos rodiklis taip pat svarbus mažinant finansinių sunkumų tikimybę, tačiau vidutinė šio rodiklio Shapley'aus reikšmė yra -0,05. Nepriklausomai nuo pasirinkto finansinės rizikos apibrėžimo, rinkos kapitalizacija, grynasis pelnas, veiklos pelnas prieš mokesčius ir palūkanas yra pagrindiniai modelio prognozę nulemiantys veiksniai.

## Literatūros sąrašas

1. POLLMAN, E. THE MAKING AND MEANING OF ESG, t. 14 (2024), p. 403–421. Prieiga per internetą: <https://ssrn.com/abstract=4219857>.
2. TETTAMANZI, P.; VENTURINI, G. ir MURGOLO, M. Sustainability and Financial Accounting: a Critical Review on the ESG Dynamics. *Environmental Science and Pollution Research*, t. 29 (2022), p. 16758–16761. Prieiga per internetą: [https://www.researchgate.net/publication/357810986\\_Sustainability\\_and\\_Financial\\_Accounting\\_a\\_Critical\\_Review\\_on\\_the\\_ESG\\_Dynamics](https://www.researchgate.net/publication/357810986_Sustainability_and_Financial_Accounting_a_Critical_Review_on_the_ESG_Dynamics).
3. United Nations Global Compact. *Who Cares Wins: Connecting the Financial Markets to a Changing World?* New York: United Nations, 2004. Prieiga per internetą: [https://www.unglobalcompact.org/docs/issues\\_doc/Financial\\_markets/who\\_cares\\_who\\_wins.pdf](https://www.unglobalcompact.org/docs/issues_doc/Financial_markets/who_cares_who_wins.pdf).
4. MACEY, J. R. *ESG Investing: Why here? Why Now?* Yale Law School, 2021. Prieiga per internetą: <https://ssrn.com/abstract=3942903>.
5. Financial Stability Board. *Recommendations of the Task Force on Climate-Related Financial Disclosures*. Basel, Switzerland: Financial Stability Board, 2017. Prieiga per internetą: <https://www.fsb.org/uploads/P290617-5.pdf>.
6. BOLTON, P. ir KACPERCZYK, M. Do investors care about carbon risk?. *Journal of Financial Economics*, t. 142 (2021), nr. 2, p. 517–549. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0304405X21001902>.
7. KHAN, M.; SERAFEIM, G. ir YOON, A. Corporate Sustainability: First Evidence on Materiality. *The Accounting Review*, t. 91 (2016).
8. SERVAES, H. ir TAMAYO, A. The Impact of Corporate Social Responsibility on Firm Value: The Role of Customer Awareness. *Management Science*, t. 59 (2012).
9. DYCK, A.; LINS, K. V.; ROTH, L. ir WAGNER, H. F. Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics*, t. 131 (2019), nr. 3, p. 693–714. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0304405X18302381>.
10. SHLEIFER, A. ir VISHNY, R. W. A Survey of Corporate Governance. *The Journal of Finance*, t. 52 (1997), nr. 2, p. 737–783. Prieiga per internetą: <https://www.jstor.org/stable/2329497>.
11. METRICK, A.; GOMPERS, P. ir ISHII, J. Corporate Governance And Equity Prices. *The Quarterly Journal of Economics*, t. 118 (2003), p. 107–155.
12. International Sustainability Standards Board (ISSB). *IFRS S1: General Requirements for Disclosure of Sustainability-Related Financial Information*. London, United Kingdom: IFRS Foundation, 2023. Prieiga per internetą: <https://www.ifrs.org/issued-standards/ifrs-sustainability-standards-navigator/ifrs-s1-general-requirements/>.
13. EDMANS, A. The end of ESG. *Financial Management*, t. 52 (2023), nr. 1, p. 3–17. Prieiga per internetą: <https://doi.org/10.1111/fima.12413>.

14. ZAKHMATOV, D.; KOCZAR, J. ir ВАГИЗОВА, В. Tools for considering ESG factors in business valuation. *Procedia Computer Science*, t. 225 (2023) (2023), p. 4245–4253.
15. NG, A. C. ir REZAEI, Z. Business sustainability performance and cost of equity capital. *Journal of Corporate Finance*, t. 34 (2015), p. 128–149. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0929119915000930>.
16. PURI, S. ESG and Corporate Financial Performance - Evidence from New Zealand. *Rere Āwhio - The Journal of Applied Research and Practice* (2022), p. 57–68.
17. XU, Y. ir ZHU, N. The Effect of Environmental, Social, and Governance (ESG) Performance on Corporate Financial Performance in China: Based on the Perspective of Innovation and Financial Constraints. *Sustainability*, t. 16 (2024), p. 3329.
18. NAZAROVA, V. V.; CHURAKOVA, I. ir DMITRIEV, O. Impact of ESG Disclosure on Financial Performance: Mandatory vs. Voluntary Disclosure. *Finance and business*, t. 19 (2023), p. 52–70.
19. AL AZIZAH, U. S. ir HARON, R. The sustainability imperative: evaluating the effect of ESG on corporate financial performance before and after the pandemic. *Discover Sustainability*, t. 6 (2025).
20. GILLAN, S. L.; KOCH, A. ir STARKS, L. T. Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, t. 66 (2021), p. 101889. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0929119921000092>.
21. FRIEDE, G.; BUSCH, T. ir BASSEN, A. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, t. 5 (2015), nr. 4, p. 210–233. Prieiga per internetą: <https://doi.org/10.1080/20430795.2015.1118917>.
22. DINCA, M.; CIOTLĂUȘI, V. ir AKOMEAH, F. Estimating the Impact of ESG on Financial Forecast Predictability Using Machine Learning Models. *International Journal of Financial Studies*, t. 13 (2025).
23. TERZIOGLU, M.; BOZCUK, A.; ÜNAL UYAR, G. F.; KAYA, N.; TUTCU, B. ir kt. Machine Learning Classification of Return on Equity from Sustainability Reporting and Corporate Governance Metrics: A SHAP-Based Explanation. *Sustainability*, t. 18 (2025), p. 194.
24. D'AMATO, V.; D'ECCLESIA, R. ir LEVANTESI, S. Firms' profitability and ESG score: A machine learning approach. *Applied Stochastic Models in Business and Industry*, t. 40 (2024), nr. 2, p. 243–261. Prieiga per internetą: <https://doi.org/10.1002/asmb.2758>.
25. TALAŞ, H.; GÖK, E. N.; AKÇAKANAT, Ö; GÜLTEKIN, G.; TERZIOĞLU, M. ir kt. The Contribution of Sustainability and Governance Signals to Return on Equity Prediction: Evidence from Tree-Based Machine Learning, Bootstrapped Grouped CV and SHAP. *Journal of Risk and Financial Management*, t. 19 (2026), nr. 2.
26. HULL, J. C. *Risk Management and Financial Institutions*. 4th Wiley, 2015..
27. ROSS, S. A. &amp. WESTERFIELD, R. &amp. JAFFE, J. F. ir JORDAN, B. D. *Corporate Finance*. 12th New York: McGraw-Hill Education, 2019..

28. MODIGLIANI, F. ir MILLER, M. H. The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, t. 48 (1958), nr. 3, p. 261–297. Prieiga per internetą: <https://www.jstor.org/stable/1809766>.
29. MODIGLIANI, F. ir MILLER, M. H. Corporate Income Taxes and the Cost of Capital: A Correction. *The American Economic Review*, t. 53 (1963), nr. 3, p. 433–443. Prieiga per internetą: <https://www.jstor.org/stable/1809167>.
30. KRAUS, A. ir LITZENBERGER, R. H. A State-Preference Model of Optimal Financial Leverage. *The Journal of Finance*, t. 28 (1973), nr. 4, p. 911–922. Prieiga per internetą: <https://www.jstor.org/stable/2978343>.
31. SCOTT, J. H. A Theory of Optimal Capital Structure. *The Bell Journal of Economics*, t. 7 (1976), nr. 1, p. 33–54. Prieiga per internetą: <https://www.jstor.org/stable/3003189>.
32. DAMODARAN, A. *Investment Valuation: Tools and Techniques for Determining the Value of any Asset, University Edition*. Wiley, 2012. Prieiga per internetą: <https://books.GoogleScholar.lt/books?id=ciJwGJfjHGkC>.
33. BRAUNSBERGER, C. ir ASCHAUER, E. Corporate Failure Prediction: A Literature Review of Altman Z-Score and Machine Learning Models Within a Technology Adoption Framework. *Journal of Risk and Financial Management*, t. 18 (2025), nr. 8.
34. RADEBE, N. E.; NOMLALA, B. C. ir MATENDA, F. R. Systematic Review of Financial Distress Prediction Models for Municipalities: Key Evaluation Criteria and a Framework for Model Selection. *Journal of Risk and Financial Management*, t. 18 (2025), nr. 11.
35. NAMAKI, A.; EYVAZLOO, R. ir RAMTINNIA, S. A Systematic Review of Early Warning Systems in Finance. 2023. Prieiga per internetą: <https://api.semanticscholar.org/CorpusID:263334602>.
36. SHI, Y. ir LI, X. An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. *Intangible Capital*, t. 15 (2019), p. 114.
37. RAHMAN, M. J. ir ZHU, H. Predicting financial distress using machine learning approaches: Evidence China. *Journal of Contemporary Accounting & Economics*, t. 20 (2024), nr. 1, p. 100403. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S1815566924000031>.
38. LOKANAN, M. ir RAMZAN, S. Predicting financial distress in TSX-listed firms using machine learning algorithms. *Frontiers in artificial intelligence*, t. 7 (2024), p. 1466321.
39. WU, D.; MA, X. ir OLSON, D. L. Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems*, t. 159 (2022), p. 113814. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0167923622000859>.
40. YANG, H.; LI, E.; CAI, Y.; LI, J. ir YUAN, G. The extraction of early warning features for the predicting financial distress based on XGboost model and shap framework. *International Journal of Financial Engineering*, t. 08 (2021), p. 2141004.
41. SUN, J.; LI, H.; HUANG, Q. ir HE, K. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches.

- Knowledge-Based Systems*, t. 57 (2014), p. 41–56. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S0950705113003869>.
42. LIU, W.; SUZUKI, Y. ir DU, S. Ensemble learning algorithms based on easyensemble sampling for financial distress prediction. *Annals of Operations Research*, t. 346 (2025), nr. 3, p. 2141–2172. Prieiga per internetą: <https://doi.org/10.1007/s10479-025-06494-y>.
43. SUN, J.; LI, H.; FUJITA, H.; FU, B. ir AI, W. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion*, t. 54 (2020), p. 128–144. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S156625351830856X>.
44. NGUYEN, M.; NGO, T.; NGUYEN, B. ir HONG, S. *Using Machine Learning and Counterfactual Explanations for Financial Distress Prediction*. , 2024..
45. GAO, R.; CUI, S.; WANG, Y. ir XU, W. Predicting financial distress in high-dimensional Imbalanced datasets: a multi-heterogeneous self-paced ensemble learning framework. *Financial Innovation*, t. 11 (2025), p. 1–34.
46. BERG, F.; KÖLBEL, J.,F. ir RIGOBON, R. Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, t. 26 (2022), nr. 6, p. 1315–1344. Prieiga per internetą: <https://doi.org/10.1093/rof/rfac033>.
47. KOTSANTONIS, S. ir SERAFEIM, G. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, t. 31 (2019), p. 50–58.
48. BILLIO, M.; COSTOLA, M.; HRISTOVA, I.; LATINO, C. ir PELIZZON, L. Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, t. 28 (2021), nr. 5, p. 1426–1445. Prieiga per internetą: <https://doi.org/10.1002/csr.2177>.
49. SONG, Y.; JIANG, M.; LI, S. ir ZHAO, S. Class-imbalanced financial distress prediction with machine learning: Incorporating financial, management, textual, and social responsibility features into index system. *Journal of Forecasting*, t. 43 (2024), nr. 3, p. 593–614. Prieiga per internetą: <https://doi.org/10.1002/for.3050>.
50. CITTERIO, A. ir KING, T. The role of Environmental, Social, and Governance (ESG) in predicting bank financial distress. *Finance Research Letters*, t. 51 (2023), p. 103411.
51. BONACORSI, L.; CERASI, V.; GALFRASCOLI, P. ir MANERA, M. ESG Factors and Firms' Credit Risk. *Journal of Climate Finance*, t. 6 (2024), p. 100032. Prieiga per internetą: <https://www.sciencedirect.com/science/article/pii/S2949728024000026>.
52. ALTMAN, E. I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, t. 23 (1968), nr. 4, p. 589–609. Prieiga per internetą: <https://www.jstor.org/stable/2978933>.
53. ALTMAN, E. Predicting Financial Distress Of Companies: Revisiting The Z-Score And Zeta. *Handbook of Research Methods and Applications in Empirical Finance*, t. 5 (2000). Prieiga per internetą: [https://www.researchgate.net/publication/2413921\\_Predicting\\_Financial\\_Distress\\_Of\\_Companies\\_Revisiting\\_The\\_Z-Score\\_And\\_Zeta](https://www.researchgate.net/publication/2413921_Predicting_Financial_Distress_Of_Companies_Revisiting_The_Z-Score_And_Zeta).

54. OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, t. 18 (1980), nr. 1, p. 109–131. Prieiga per internetą: <http://www.jstor.org/stable/2490395>.
55. ASQUITH, P.; GERTNER, R. ir SCHARFSTEIN, D. Anatomy of Financial Distress: An Examination of Junk-Bond Issuers. *The Quarterly Journal of Economics*, t. 109 (1994), nr. 3, p. 625–658. Prieiga per internetą: <http://www.jstor.org/stable/2118416>.
56. WHITAKER, R. B. The early stages of financial distress. *Journal of Economics and Finance*, t. 23 (1999), nr. 2, p. 123–132. Prieiga per internetą: <https://doi.org/10.1007/BF02745946>.
57. CAMPBELL, J. Y.; HILSCHER, J. ir SZILAGYI, J. In Search of Distress Risk. *The Journal of Finance*, t. 63 (2008), nr. 6, p. 2899–2939. Prieiga per internetą: <https://doi.org/10.1111/j.1540-6261.2008.01416.x>.
58. PLATT, H. D. ir PLATT, M. B. Predicting corporate financial distress: Reflections on choice-based sample bias. *Journal of Economics and Finance*, t. 26 (2002), nr. 2, p. 184–199. Prieiga per internetą: <https://doi.org/10.1007/BF02755985>.
59. BLAGUS, R. ir LUSA, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, t. 14 (2013), nr. 1, p. 106. Prieiga per internetą: <https://doi.org/10.1186/1471-2105-14-106>.
60. ČEKANA VIČIUS, V. ir MURAUSKAS, G. *Taikomoji regresinė analizė socialiniuose tyrimuose*. Vilniaus universiteto leidykla, 2014..
61. CORTES, C. ir VAPNIK, V. Support-vector networks. *Machine Learning*, t. 20 (1995), nr. 3, p. 273–297. Prieiga per internetą: <https://doi.org/10.1007/BF00994018>.
62. HASTIE, T.; TIBSHIRANI, R. ir FRIEDMAN, J. *The Elements of Statistical Learning*. 2nd New York, NY: Springer.
63. FRIEDMAN, J. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, t. 29 (2000).
64. CHEN, T. ir GUESTRIN, C. *XGBoost: A Scalable Tree Boosting System*. , 2016..
65. *XGBoost Documentation - Parameters*. [2024. Prieiga per internetą: <https://xgboost.readthedocs.io/>.
66. KE, G.; MENG, Q.; FINLEY, T.; WANG, T.; CHEN, W. ir kt. *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. , 2017..
67. BREIMAN, L. Random Forests. *Machine Learning*, t. 45 (2001), nr. 1, p. 5–32. Prieiga per internetą: <https://doi.org/10.1023/A:1010933404324>.
68. LIU, X.; WU, J. ir ZHOU, Z. Exploratory Undersampling for Class-Imbalance Learning. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, t. 39 (2009), p. 539–550.
69. LUNDBERG, S. ir LEE, S. *A Unified Approach to Interpreting Model Predictions*. , 2017..

70. MOTHILAL, R.; SHARMA, A. ir TAN, C. *Explaining machine learning classifiers through diverse counterfactual explanations.* , 2020..

71. WILCOXON, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, t. 1 (1945), nr. 6, p. 80–83. Prieiga per internetą: <http://www.jstor.org/stable/3001968>.

## Priedai

### 1 priedas

#### Duomenų struktūra ir kintamųjų aprašymai

Kintamojo pavadinimas	Paaiškinimas
ric	Refinitiv Instrument Code
company_name	Įmonės pavadinimas
country	Šalis
sector	Įmonės veiklos sektorius
industry	Įmonės pramonės rūšis
esg_coverage	ESG padegimas
current_assets_2023	Trumpalaikis turtas 2023 metais
current_assets_2024	Trumpalaikis turtas 2024 metais
current_liabilities_2023	Trumpalaikiai įsipareigojimai 2023 metais
current_liabilities_2024	Trumpalaikiai įsipareigojimai 2024 metais
ebit_2023	Pelnas prieš palūkanas ir mokesčius 2023 metais
ebit_2024	Pelnas prieš palūkanas ir mokesčius 2024 metais
ebitda_2023	Pelnas prieš palūkanas, mokesčius, nusidėvėjimą ir amortizaciją 2023 metais
ebitda_2024	Pelnas prieš palūkanas, mokesčius, nusidėvėjimą ir amortizaciją 2024 metais
environmental_score_2023	Agreguotas aplinkosaugos balas 2023 metais
environmental_score_2024	Agreguotas aplinkosaugos balas 2024 metais
esg_score_2023	Bendras agreguotas ESG balas 2023 metais
esg_score_2024	Bendras agreguotas ESG balas 2024 metais
governance_score_2023	Agreguotas valdysenos balas 2023 metais
governance_score_2024	Agreguotas valdysenos balas 2024 metais
interest_expense_2023	Palūkanų sąnaudos 2023 metais
interest_expense_2024	Palūkanų sąnaudos 2024 metais
market_cap_2023	Rinkos kapitalizacija 2023 metais
market_cap_2024	Rinkos kapitalizacija 2024 metais
net_income_2023	Grynas pelnas 2023 metais
net_income_2024	Grynas pelnas 2024 metais
retained_earnings_2023	Nepaskirstytas pelnas 2023 metais
retained_earnings_2024	Nepaskirstytas pelnas 2024 metais
revenue_2023	Pajamos 2023 metais
revenue_2024	Pajamos 2024 metais
social_score_2023	Agreguotas socialinės atsakomybės balas 2023 metais
social_score_2024	Agreguotas socialinės atsakomybės balas 2024 metais

<b>Kintamojo pavadinimas</b>	<b>Paaiškinimas</b>
total_assets_2023	Bendras turtas 2023 metais
total_assets_2024	Bendras turtas 2024 metais
total_debt_2023	Bendras įsiskolinimas 2023 metais
total_debt_2024	Bendras įsiskolinimas 2024 metais
total_liabilities_2023	Bendri įsipareigojimai 2023 metais
total_liabilities_2024	Bendri įsipareigojimai 2024 metais
financial_d_altzzz_2025	Finansiniai sunkumai 2025 metais apibrėžiami pagal Altman'o Z''
financial_d_net_income_2025	Finansiniai sunkumai 2025 metais apibrėžiami neigiamu grynuoju pelnu
financial_d_interest_coverage_2025	Finansiniai sunkumai 2025 metais apibrėžiami įmonės nesugebėjimu padengti skolą
financial_d_ebit_2025	Finansiniai sunkumai 2025 metais apibrėžiami neigiamu veiklos pelnu
financial_d_combined_2025	Finansiniai sunkumai 2025 metais apibrėžiami daugialype sąlyga
current_assets_diff	Trumpalaikio turto skirtumas tarp 2024 ir 2023 metų
current_liabilities_diff	Trumpalaikių įsipareigojimų skirtumas tarp 2024 ir 2023 metų
ebit_diff	Pelno prieš palūkanas ir mokesčius skirtumas tarp 2024 ir 2023 metų
ebitda_diff	Pelno prieš palūkanas, mokesčius, nusidėvėjimą ir amortizaciją skirtumas tarp 2024 ir 2023 metų
environmental_score_diff	Agreguoto aplinkosaugos balo skirtumas tarp 2024 ir 2023 metų
esg_score_diff	Bendro agreguoto ESG balo skirtumas tarp 2024 ir 2023 metų
governance_score_diff	Agreguoto valdysenos balo skirtumas tarp 2024 ir 2023 metų
interest_expense_diff	Palūkanų sąnaudų skirtumas tarp 2024 ir 2023 metų
market_cap_diff	Rinkos kapitalizacijos skirtumas tarp 2024 ir 2023 metų
net_income_diff	Grynojo pelno skirtumas tarp 2024 ir 2023 metų
retained_earnings_diff	Nepaskirstyto pelno skirtumas tarp 2024 ir 2023 metų
revenue_diff	Pajamų skirtumas tarp 2024 ir 2023 metų
social_score_diff	Agreguoto socialinės atsakomybės balo skirtumas tarp 2024 ir 2023 metų
total_assets_diff	Bendro turto skirtumas tarp 2024 ir 2023 metų
total_debt_diff	Bendro įsiskolinimo skirtumas tarp 2024 ir 2023 metų
total_liabilities_diff	Bendrų įsipareigojimų skirtumas tarp 2024 ir 2023 metų