



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

# **Kreditingumo reitingo pokyčio krypties prognozavimas**

Baigiamasis magistro projektas

---

**Paulius Patalavičius**  
Projekto autorius

Prof. dr. Evaldas Vaičiukynas

Vadovas

Doc. dr. Šviesa Leitonienė

Vadovė

---

**Kaunas, 2020**



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

# **Kreditingumo reitingo pokyčio krypties prognozavimas**

Baigiamasis magistro projektas

Didžiųjų verslo duomenų analitika (6213AX001)

---

**Paulius Patalavičius**

Projekto autorius

**Prof. dr. Evaldas Vaičiukynas**

Vadovas

**Doc. dr. Šviesa Leitonienė**

Vadovė

**Doc. dr. Kristina Šutienė**

Recenzentė

**Doc. dr. Aušrinė Lakštutienė**

Recenzentė

**Kaunas, 2020**



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

## **Kreditingumo reitingo pokyčio krypties prognozavimas**

### Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Pauliaus Patalavičiaus, baigiamasis projektas tema „Kreditingumo reitingo pokyčio krypties prognozavimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

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

---

(vardą ir pavardę įrašyti ranka)

---

(parašas)



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

## **Baigiamojo magistro projekto užduotis**

Projekto tema Kreditingumo reitingo pokyčio krypties prognozavimas

Reikalavimai ir sąlygos  
(tikslinti pavadinimą  
pagal poreikį)

Vadovas / Vadovė \_\_\_\_\_  
(vadovo pareigos, vardas, pavardė, parašas) (data)

Vadovas / Vadovė \_\_\_\_\_  
(vadovo pareigos, vardas, pavardė, parašas) (data)

Patalavičius, Paulius. Kreditingumo reitingo pokyčio krypties prognozavimas. Magistro baigiamasis projektas / vadovai: prof. dr. Evaldas Vaičiukynas ir doc. dr. Šviesa Leitonienė; Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Pagrindinė kryptis: Taikomoji matematika (Matematikos mokslai), papildančios kryptys: Ekonomika (Socialiniai mokslai) ir Informatika (Informatikos mokslai)

Reikšminiai žodžiai: kreditingumo reitingas, detekcija, klasifikavimas, išlikimo analizė .

Kaunas, 2020. 98 p.

## Santrauka

Kreditingumo reitingo agentūros yra atsakingos už kreditingumo reitingo valstybėms ir įmonėms sudarymą. Šis reitingas vertina skolininko galimybes gražinti skolą laiku bei tikimybę, jog prisiimti įsipareigojimai nebus vykdomi. Maždaug 95% kreditingumo reitingo industrijos yra valdoma 3 didžiausių kreditingumo reitingo agentūrų (dažnai vadinamų „Didžiojo trejeto“ agentūromis): „Standard & Poor`s“, „Moody`s“ ir „Fitch“. Todėl konkurencija kreditingumo reitingo industrijoje yra labai nedidelė. „Didžiojo trejeto“ agentūros nuolatos susilaukia kritikos dėl priimamų sprendimų. Pavyzdžiui, prieš 2007 – 2008 metų finansinę krizę, agentūros nesugebėjo tinkamai įvertinti finansinių priemonių rizikos ir padarė didelę įtaką kilusiam pasauliniam ekonominiam nuosmukiui. Šiomis dienomis kritikos agentūros susilaukia ir už tai, jog jų priimami neproporcingai griežti kreditingumo reitingai brangina skolinimąsi ir stabdo valstybių ekonominę progresą, skatina žmones „susiveržti diržus“.

Baigiamajame magistro darbe ištirtas kreditingumo agentūrų sudarytų reitingų pokyčių sąryšis su realia finansine bankų situacija. Baigiamojo darbo tikslas - naudojant istorinius bankų finansinius duomenis ir mašininio mokymo algoritmus, ištirti „Didžiojo trejeto“ kreditingumo reitingo agentūrų sudarytus reitingo pasikeitimus ir išsiaiškinti, ar agentūrų sudaryti reitingai remiasi realiais finansiniais rodikliais, ar jie sudaromi pasitelkiant medijų kanalus bei politinius sprendimus.

Darbe buvo naudojami realūs duomenys, kurie gauti naudojant „Bloomberg“ programinės įrangos sistemą. Mašininio mokymo modeliai sukonstruoti pritaikant 89 finansinius rodiklius. Prognozėms sudaryti buvo atlikti detekcijos, išlikimo analizės ir klasifikavimo uždaviniai.

Detekcijos uždavinyje prognozuota, ar įvykus tam tikrų finansinių rodiklių pokyčiui, per 1 metus laiko tiriamo banko kreditingumo reitingo įvertis mažės, ar nemažės. Detekcijos uždaviniui atlikti naudoti logistinės regresijos, atsitiktinių miškų bei atramos vektorių modeliai. Tiksliausiai buvo prognozuojami „Moody`s“ agentūros reitingo pokyčiai (ROC AUC - 0.92).

Klasifikavimo uždavinyje analizuota, ar po finansinių rodiklių pasikeitimo, per 1 metus laiko tiriamo banko kreditingumo reitingas mažės, nesikeis arba didės. Klasifikavimo uždavinys spęstas taikant atsitiktinių miškų ir atramos vektorių klasifikatorius. Patikimiausiai visas 3 klases atskyrė atsitiktinių miškų modelis analizuodamas „Fitch“ agentūros rezultatus. Bendras modelio tikslumas siekė 65.336 %. Tačiau aukščiausias Kappa įvertis pasiektas tiriant „Moody`s“ agentūros duomenis.

Išlikimo analizės uždavinyje apskaičiuota vidutinė kiekvienos dienos tikimybė, jog kreditingumo reitingas sumažės. Analizuojamas laikotarpis iki 1 metų po susijusių finansinių rodiklių pokyčių.

Išlikimo analizės uždaviniui spręsti panaudoti išlikimo miškų bei Cox modeliai. Išlikimo analizės algoritmai visų 3 agentūrų reitingų pokyčius vertino labai panašiai.

Patalavičius Paulius. Predicting direction of change in credit rating. Master's Final Degree Project / supervisors: prof. Evaldas Vaičiukynas and assoc. prof. Šviesa Leitonienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): applied mathematics, informatics and economics.

Keywords: credit rating, detection, classification, survival analysis.

Kaunas, 2020. 98 pages.

### **Summary**

Credit rating agencies are responsible for issuing credit ratings for countries and companies. This rating assesses the debtor's ability to repay the debt on time and probability that the obligations will not be met. Approximately 95% of the credit rating industry is managed by the 3 largest credit rating agencies (often referred to as the Big Three): Standard & Poor's, Moody's and Fitch. As a result, competition in the credit rating industry is very limited. The "Big Three" agencies are constantly criticized for their decisions. For instance, before the financial crisis of 2007 - 2008, agencies failed to properly assess the risks of financial instruments. This led to a significant impact on the global economic downturn. These days, agencies are also criticized for disproportionately managing their credit ratings, making borrowing more expensive and encouraging people to "tighten their belts."

The Master's Final degree project examines the relationship between changes in ratings issued by credit agencies and the real financial situation of banks. The aim of the thesis is to study the changes in ratings made by the Big Three credit agencies using historical corporate financial data and machine learning algorithms. Thesis also analyzes whether the ratings made by the credit agencies are based on real financial indicators or whether they are made through media channels and political decisions.

The work used real data obtained from the Bloomberg software system. Machine training models were constructed by applying 89 financial variables. Detection, classification, and survival analysis tasks were performed to make informed predictions.

The detection models predict whether the credit rating of the investigated bank will decrease or will not due to a change in certain financial indicators over a period of 1 year. Logistic regression, random forest and support vector models were used to perform the detection task. The changes in Moody's rating were most accurately predicted. A random forest model with a ROC AUC of 0.92 was used for this prediction.

The classification task analyzes whether, after a change in financial ratios, the credit rating of the analyzed bank will decrease, remain unchanged or increase over a period of 1 year. The classification problem is solved by applying classifiers of random forests and support vectors. All 3 classes were most reliably separated by the random forest model when analyzing Fitch's results. The overall accuracy of the model was 65.336%.

Using the survival analysis method, the average daily probability that the credit rating would decline was calculated. We analyze the period up to 1 year after the changes in the related financial indicators. Survival forest and Cox models were used to solve the problem. The survival analysis models evaluated the changes in the ratings of all 3 credit agencies very similarly.

## Turinys

<b>Lentelių sąrašas .....</b>	<b>10</b>
<b>Paveikslų sąrašas .....</b>	<b>11</b>
<b>Santrumpų ir terminų sąrašas .....</b>	<b>12</b>
<b>Įvadas.....</b>	<b>13</b>
<b>1. Problemos ir literatūros analizė .....</b>	<b>15</b>
1.1. Kreditingumo reitingavimas .....	15
1.1.1. Kreditingumo reitingo apibrėžimas .....	15
1.1.2. Kreditingumo reitingų tipai ir jų teikiama nauda .....	17
1.2. Kreditingumo reitingo taikymo procesas finansinėse institucijose .....	19
1.2.1. 1 žingsnis: Nuspręsti į kokį segmentą bus orientuojamasi .....	21
1.2.2. 2 žingsnis: Pasirinkti naudojamą kreditingumo reitingo tipą .....	21
1.2.3. 3 žingsnis: Suprojektuoti kreditingumo reitingavimo modelius .....	23
1.2.4. 4 žingsnis: Ištestuoti, paruošti naudojimui, prižiūrėti kreditingumo reitingavimo modelius .....	25
1.3. Kreditingumo reitingų agentūrų sprendimų priėmimo kriterijai .....	26
1.4. Kreditingumo reitingų agentūros ir pavyzdinė išlikimo analizė .....	27
1.5. Apibendrinimas .....	31
<b>2. Taikomų metodologijų ir sprendimų apžvalga.....</b>	<b>32</b>
2.1. „Bloomberg“ terminalas .....	32
2.2. Klasifikavimo ir detekcijos uždaviniai .....	32
2.2.1. Klasifikavimo ir detekcijos uždaviniams spręsti skirti algoritmai .....	32
2.3. Išlikimo analizė .....	35
2.3.1. Žvalgomoji išlikimo analizė .....	38
2.3.2. Regresiniai išlikimo modeliai .....	39
2.4. Geriausio modelio pasirinkimas .....	41
2.5. Mašininio mokymo modelių parametrų derinimas .....	47
2.5.1. Rankinis parametrų derinimas .....	47
2.5.2. Tinklelio paieška (angl. Grid search) .....	47
2.5.3. Atsitiktinė paieška (angl. Random search) .....	48
2.5.4. Bajesinis optimizavimas .....	48
<b>3. Tyrimo objektas, realizacija .....</b>	<b>53</b>
3.1. Darbo priemonės .....	53
3.2. Duomenų gavimas .....	53
3.3. Duomenų apdorojimas, paruošimas darbui .....	54
3.4. Duomenų apžvalga .....	56
3.5. Modeliavimas .....	59
3.5.1. Detekcijos uždavinio rezultatai .....	60
3.5.2. Klasifikavimo uždavinio modelių rezultatai .....	67
3.5.3. Išlikimo analizės uždavinio modelių rezultatai ir palyginimas su detekcija .....	72
3.6. Naudojamų finansinių rodiklių svarbos analizė .....	74
3.7. Tyrimo rezultatų apibendrinimas .....	77
<b>4. Darbo rezultatai ir išvados .....</b>	<b>79</b>
<b>5. Literatūros sąrašas .....</b>	<b>80</b>
<b>6. Priedai.....</b>	<b>82</b>
1 priedas. Tyrime naudojamų finansinių rodiklių aprašas (anglų kalba) .....	82





## Lentelių sąrašas

<b>1.1 lentelė.</b> Duomenų šaltiniai modeliuojant kreditingumo įverčius fiziniams asmenims .....	16
<b>1.2 lentelė.</b> Kreditingumo reitingų tipai .....	17
<b>1.3 lentelė.</b> Svarbiausių kredito agentūrų reitingų nominalai .....	18
<b>1.4 lentelė.</b> Organizacinių kreditingumo vertinimo įverčių privalumai ir galimi iššūkiai .....	19
<b>1.5 lentelė</b> Tiriamų įmonių kiekiai verslo sektoriuose .....	28
<b>1.6 lentelė</b> Tiriamų įmonių kiekiai valstybėse .....	29
<b>3.1 lentelė.</b> Kreditingumo reitingų pokyčių duomenys .....	53
<b>3.2 lentelė.</b> „Standard & Poor`s“ kreditingumo reitingo pokyčiai .....	56
<b>3.3 lentelė.</b> „Fitch“ kreditingumo reitingo pokyčiai .....	56
<b>3.4 lentelė.</b> <i>Moody`s</i> kreditingumo reitingo pokyčiai .....	56
<b>3.5 lentelė.</b> „ <i>Standard &amp; Poor`s</i> “ duomenų rinkinio prognozavimo modelių palyginimas .....	62
<b>3.6 lentelė.</b> <i>Fitch</i> duomenų rinkinio prognozavimo modelių palyginimas .....	64
<b>3.7 lentelė.</b> „ <i>Moody`s</i> “ duomenų rinkinio prognozavimo modelių palyginimas.....	65
<b>3.8 lentelė.</b> „ <i>Standard &amp; Poor`s</i> “ duomenų rinkinio klasifikavimo modelių palyginimas .....	68
<b>3.9 lentelė.</b> „ <i>Fitch</i> “ duomenų rinkinio klasifikavimo modelių palyginimas .....	70
<b>3.10 lentelė.</b> „ <i>Moody`s</i> “ duomenų rinkinio klasifikavimo modelių palyginimas.....	71
<b>3.11 lentelė.</b> Išlikimo miško modelio įverčiai .....	72
<b>3.12 lentelė.</b> Cox modelio įverčiai .....	72

## Paveikslų sąrašas

1.1 pav. Reitingavimo sistemos vystymo laipsniai .....	21
1.2 pav. Kreditingumo reitingo tipo pasirinkimas .....	23
1.3 pav. Išlikimo funkcija skirtinguose sektoriuose.....	29
1.4 pav. Išlikimo funkcija skirtingose valstybėse .....	30
1.5 pav. Kaupiamąjo pavojingumo funkcija bankų ir kituose sektoriuose .....	31
2.1 pav. Atsitiktiniai miškai .....	34
2.2 pav. SVM – atramos vektorių modelis.....	34
2.3 pav. Branduolio triukas atramos vektorių modelyje .....	35
2.4 pav. Išlikimo analizė, pavyzdys .....	36
2.5 pav. Išlikimo funkcija.....	37
2.6 pav. Pavojingumo ir kaupiamąjo pavojingumo funkcijos .....	38
2.7 pav. 2 skirtingų grupių išlikimas .....	39
2.8 pav. Šališkumo ir dispersijos kompromisas .....	41
2.9 pav. Sumaišymo matrica .....	43
2.10 pav. ROC kreivė.....	45
2.11 pav. Idealaus modelio ROC kreivė .....	46
2.12 pav. Pradinių modelio parametrų grupavimas .....	50
2.13 pav. Parzeno medžio vertinimas 4 stebiniams .....	51
3.1 pav. Duomenų fiksavimo laikas, metų ketvirtis.....	57
3.2 pav. Tiriamų „Standard & Poor`s“ įmonių registravimo šalys .....	58
3.3 pav. Tiriamų „Fitch“ įmonių registravimo šalys.....	58
3.4 pav. Tiriamų „Moody`s“ įmonių registravimo šalys.....	59
3.5 pav. „Standard & Poor`s“ agentūros prognozių ROC kreivė .....	61
3.6 pav. „Standard & Poor`s“ agentūros prognozių PR kreivė.....	62
3.7 pav. „Fitch“ agentūros prognozių ROC kreivė .....	63
3.8 pav. „Fitch“ agentūros prognozių PR kreivė .....	63
3.9 pav. „Moody`s“ agentūros prognozių ROC kreivė.....	64
3.10 pav. „Moody`s“ agentūros prognozių PR kreivė .....	65
3.11 pav. Kreditingumo agentūrų prognozių ROC kreivė .....	66
3.12 pav. Kreditingumo agentūrų prognozių PR kreivė .....	67
3.13 pav. „Standard & Poor`s“ reitingo pokyčio klasifikavimo sumaišymo matrica; atsitiktinis miškas .....	68
3.14 pav. „Standard & Poor`s“ duomenų rinkinio sumaišymo matrica; atramos vektorių modelis .....	68
3.15 pav. „Fitch“ duomenų rinkinio sumaišymo matrica; atsitiktinis miškas .....	69
3.16 pav. „Fitch“ duomenų rinkinio sumaišymo matrica; atramos vektorių modelis .....	70
3.17 pav. „Moody`s“ duomenų rinkinio sumaišymo matrica; atsitiktinis miškas .....	71
3.18 pav. „Moody`s“ duomenų rinkinio sumaišymo matrica; atramos vektorių modelis .....	71
3.19 pav. „Standard & Poor`s“ detekcijos ir išlikimo analizės palyginimas.....	73
3.20 pav. „Fitch“ detekcijos ir išlikimo analizės palyginimas.....	73
3.21 pav. „Moody`s“ detekcijos ir išlikimo analizės palyginimas .....	74
3.22 pav. Finansinių rodiklių svarba sudarant „Standard & Poor`s“ prognozes.....	75
3.23 pav. Finansinių rodiklių svarba sudarant „Fitch“ prognozes .....	76
3.24 pav. Finansinių rodiklių svarba sudarant „Moody`s“ prognozes .....	77

## Santrumpų ir terminų sąrašas

### Terminai:

**Multikolinearumas** – reiškinys, kai modelio nepriklausomi kintamieji tarpusavyje koreliuoja. Situacija gali būti vadinama multikolinearia, jei egzistuoja tiesinės priklausomybės tarp keleto arba visų modelio nepriklausomų kintamųjų.

**Reguliarizacija** – regresijos forma, kai modelio koeficientų įverčiai sumažinami link 0. Regularizacija naudojama siekiant neapmokyti labai sudėtingo ir lankstaus (prisitaikančio prie triukšmo) modelio, tokiu būdu išvengiant persimokymo (angl. overfitting).

**Visuminė kliento vertė** – pajamos, kurias įmonė gauna per visą jos ir kliento bendradarbiavimo laikotarpį. Rodiklis parodo kiek ir kokių produktų klientas perka ir kiek yra pasiryžęs tam išleisti pinigų.

**Brier įgūdžių balas** – Brier balo išplėtimas, leidžiantis interpretuoti vidutines balo vertes. Apskaičiuojant Brier įgūdžių balą analizuojama prognozė yra lyginama su kita prognoze. Jei lyginimo rezultatas yra 1, analizuojama prognozė yra geresnė. Jei gaunamas 0 – abiejų prognozių galimybės yra lygios.

**C indeksas (angl. C-index)** – ROC AUC įverčio apibendrinimas, kuris atsižvelgia į *cenžūruotas* reikšmes. Įvertis aprašo modelio klasių atskyrimo gebėjimus teisingai pateikti išlikimo reitingą, kuris yra pagrįstas individualiais rizikos balais.

**Aplikacijų programavimo sąsaja (angl. API)** – kompiuterio sistemos suteikiama sąsaja, kurią naudodami programuotojai gali naudotis sistemos funkcionalumu ir apsikeisti duomenimis tarp sistemų.

**Klasių disbalansas** – klasifikavimo ir detekcijos uždavinių problema, kai didžioji dalis stebinių priklauso vienai ar kelioms klasėms. Likusioms klasėms atstovauja gerokai mažiau stebinių, susidaro nelygė. Esant dideliame klasių disbalansui gali būti gaunami iškreipti prognozavimo rezultatai.

**Vienetinis kodavimas (angl. one-hot encoding)** – kategorinių kintamųjų formato pakeitimas į mašininio mokymo algoritmams suprantamą formą. Kiekvienai kategorinio kintamojo vertei  $x$  sukuriami po naują kintamąjį (stulpelį)  $x$ . Jei stebinyje fiksuojama ši vertė  $x$ , naujame stulpelyje rašomas „1“, jei nefiksuojama – „0“. Užkodavus visus kategorinius kintamuosius, pradiniai kategoriniai stulpeliai yra pašalinami iš duomenų rinkinio ir paliekami vienetinio kodavimo metu gauti nauji kintamieji.

## Įvadas

Paskutiniais metais itin sparčiai vystosi kompiuterinės technikos infrastruktūros pajėgumai bei gebėjimai apdoroti vis didesnius informacijos kiekius, atlikti tikslesnes ir spartesnes kompiuterines prognozes. Todėl įmonės, norėdamos išlikti patrauklios rinkoje, stengiasi efektyviai ir sparčiai išnaudoti mašininio mokymo naudojimo teikiamas naudas. Įmonės, kurios šiandien investuoja į technologijas bei veikloje taiko kompiuterinę duomenų analizę, ateityje turės išvystytą analitikos bei verslo įvykių prognozavimo modelių infrastruktūrą. Paruošta infrastruktūra leis būti dideliu žingsniu priekyje prieš konkurentus, kurie pavėluos ir nespės laiku pasinaudoti duomenų analizės ir mašininio mokymo algoritmų teikiamais privalumais.

Pritaikant mašininio mokymo metodiką, iš duomenų gaunama loginė informacija, kurią galima panaudoti interpretuojant norimus uždavinius ir / ar juos išspręsti (pavyzdžiui optimizuojant pardavimų procesą, mažinant išlaidas, pritraukiant klientus, atliekant natūralios kalbos analizę). Viena svarbiausių mašininio mokymo taikymo krypčių – kreditingumo rizikos prognozavimas. Pasaulyje didėjant gyventojų skaičiui, didėja paskolų teikimo kiekiai ir sumos, auga ir efektyvių kreditingumo vertinimo sprendimų poreikis. Kreditingumo rizikos vertinimo poreikis ypatingai išryškėjo įvykus 2007 - 2008 metų finansinei krizei. Krizė įsibėgėjo bankrutavus investiciniam bankui „Lehman Brothers“, kuris nesugebėjo suvaldyti prisiimtų finansinių rizikų ir išryškino kiek įmanoma tikslesnio finansinių rodiklių pokyčių prognozavimo svarbą. Finansinės krizės laikotarpis sutapo su žinių duomenyse radimo technologijos implementavimu daugelyje veiklų. Kompiuterinis modeliavimas pradėjo dar sparčiau vystytis bankuose, medicinoje, prekyboje, finansinėse rinkose. Mašininio mokymo metodai tapo viena reikšmingiausių technologijų – kompiuterinių modelių nustatomos prognozės dažnai pasirodė esančios gerokai tikslesnės net už savo srities ekspertų, žmonių prognozes. Todėl kreditingumo reitingo vertinimo sistemos tapo svarbiu įrankiu apskaičiuojant kreditingumo riziką bei atliekant rizikos pokyčių stebėseną.

Kreditingumo reitingo vertinimo sistemų tikslas – suklasifikuoti kredito paraiškų teikėjus į 2 grupes. 1 grupei priskiriami patikimi klientais, kuriems apsimoka skolinti. Manoma, jog jie dažniausiai sugebės gražinti paskolą laiku ir paskolos teikėjas uždirbs numatytas palūkanas. 2 grupei priskiriami nepatikimi klientai: prognozuojama, jog suteikti jiems paskolą bus nuostolinga. Dažniausiai finansinės institucijos skolininko (fizinio ar juridinio asmens) riziką nustato sudarydamos jų kreditingumo reitingą.

*Darbo tikslas* – naudojant istorinius įmonių finansinius duomenis ir mašininio mokymo algoritmus, ištirti 3 didžiausių kreditingumo reitingo agentūrų reitingų pasikeitimus.

Darbe siekiama nustatyti ar agentūrų sudaryti kreditingumo reitingai didės, mažės ar išliks tokie patys. Siekiama išsiaiškinti, ar agentūrų sudaryti reitingai remiasi realiais finansiniais rodikliais, ar jie sudaromi pasitelkiant medijų kanalus bei politinius sprendimus. Tyrimui atlikti naudojama reali informacija gauta iš vienos didžiausių pasaulyje finansinių ir verslo naujienų agentūrų bei informacijos apie finansines rinkas tiekėjų „Bloomberg L.P“. Duomenys apima bankų reitingų pokyčius 2007 metų sausį – 2019 metų gruodį bei 89 finansinių rodiklių kintamųjų reikšmes bei jų pokyčius bėgant laikui.

Darbo tikslui pasiekti buvo išsikelti toliau išvardinti *uždaviniai*:

1. Atlikti problemos analizę: apibrėžti kreditingumo reitingą, jo tipus ir naudojimo svarbą. Išanalizuoti, kaip kreditingumo reitingas naudojamas finansinėse institucijose, ištirti kreditingumo reitingo taikymo procesą.
2. Apžvelgti kreditingumo reitingo pokyčio nustatymui tinkamus uždavinius: detekciją, klasifikavimą, išlikimo analizę.
3. Aprašyti geriausio modelio parinkimo metodikas. Apžvelgti kaip jos yra apskaičiuojamos, kodėl ir kada jas verta naudoti. Apžvelgti mašininio mokymo modelių parametrų derinimo procedūrą. Išanalizuoti 4 populiariausius parametrų derinimo būdus: rankinį parametrų derinimą, tinklelio paiešką, atsitiktinę paiešką bei Bajesinį optimizavimą. Išanalizuoti darbe naudojamas Bajesinio optimizavimo metodikas: Parzeno medį ir simuliuojamą atkaitinimą.
4. Atlikti kreditingumo reitingo agentūrų reitingų bei tiriamų bankų duomenų gavybą iš „Bloomberg“ duomenų bazių. Atlikti gautų duomenų paruošimą. Apžvelgti importuotus duomenis, ištirti jų panaudojimo galimybes modeliavime. Atlikti žvalgomąją duomenų analizę.
5. Naudojant mašininio mokymo algoritmus, sumodeliuoti kreditingumo reitingų pokyčių krypties prognozes svarbiausioms kreditingumo reitingo agentūroms („Standard & Poor`s“, „Fitch“ ir „Moody`s“).
6. Palyginti algoritmų rezultatus, įvertinti jų patikimumą. Ištirti, kaip sekėsi prognozuoti kiekvienos „Didžiojo trejeto“ kreditingumo reitingo agentūros reitingo pokyčius.

## 1. Problemos ir literatūros analizė

### 1.1. Kreditingumo reitingavimas

Kreditingumo reitingavimas yra viena pačių sėkmingiausių statistikos ir modeliavimo sričių finansų ir bankų sektoriuose. Analitikai, nustatantys kreditingumo reitingą užima labai svarbią vietą finansinėse institucijose ir jų svarba kiekvienais metais tik didėja. Kreditingumo reitingo skaičiavimas bei taikymas buvo viena pirmų finansinės rizikos vertinimo priemonių. Pinigų ir kitų vertybių skolinimas siekiant gauti naudos egzistuoja jau daugiau nei 5000 metų, tačiau kreditingumo vertinimas pradėtas taikyti tik prieš 80–90 metų. Kreditingumo reitingavimas apibrėžiamas kaip būdas duomenų rinkinyje atskirti tam tikras grupes, kai nėra konkrečių charakteristikų, kurios aprašo vieną ar kitą grupę, tačiau galima rasti panašumų ir skirtumų tarp kelių grupių. Klasifikavimo uždavinys buvo pirmą kartą pristatytas 1936 metais britų statistiko ir biologo Ronaldo Fišerio, straipsnyje „*The use of multiple measurements in taxonomic problems*“. Straipsnyje aprašytas garsusis Iris gėlių duomenų rinkinys, kuris yra sudarytas iš 3 skirtingų klasių duomenų. Klasifikacija atliekama panaudojant 4 turimus parametrus. Tyrime taikomas linijinis atskirties modelis, kuris pagal fizinius augalų duomenis išskiria juos į 3 klases. Prasidėjus Antrajam pasauliniam karui finansinės institucijos susidūrė su kreditingumo rizikos valdymo problemomis. Iki tol kreditingumą vertinę analitikai buvo siunčiami į kariuomenę ir nebegalėjo atlikti savo darbo. Juos reikėjo kažkaip pakeisti. Dėl šios priežasties buvo sudaromi kredito išdavimo taisyklių rinkiniai, kurie turėjo pakeisti kreditingumo ekspertus. Taisyklių rinkiniai buvo sudaromi taip, kad ir analitinių įgūdžių neturintys asmenys galėtų nustatyti ar finansinė institucija turėtų išduoti paskolą, ar ne. Šie rinkiniai tapo viena pirmųjų pasaulyje ekspertinių sistemų.

Pasibaigus Antrajam pasauliniam karui, buvo gana greitai nuspręsta sujungti skolinimosi sprendimų automatizavimą bei statistinius klasifikavimo metodus. Klientų kreditingumo reitingavimas pradėtas naudoti Jungtinių Amerikos Valstijų perpardavinėtojų bei prekių platintojų 6-ajame dešimtmetyje, siekiant minimizuoti prisiimamą riziką. Kreditingumo reitingavimas taip pat padėjo pamatus šiandien labai svarbiai sričiai – duomenų gavybai. Didelės svarbos duomenų gavybos metodikos: vartotojų segmentavimas, jų polinkių modeliavimas ir klasterizavimas iš pradžių buvo labai sėkmingai taikomi kreditingumo reitingo modeliavimui. Kreditingumo reitingų pritaikymas padėjo įmonėms sumažinti nemokių klientų srautus 50 % ir daugiau, lyginant su ankstesniais ekspertiniais vertinimais. Nuo 9-ojo dešimtmečio kreditingumo reitingavimas pradėtas naudoti ne tik išduodant paskolas įmonėms, bet ir teikiant vartojimo paskolas, būsto paskolas, paskolas verslui. Taip pat reitingavimas pradėtas naudoti reklaminėse kampanijose: nustatant kaip sėkmingai reklama pritrauks naujus vartotojus, kokią įtaką reklama darys išlaikant esamus vartotojus, vertinant kuriems klientams verta siųsti reklaminius atvirlaiškius bei kokie jie turėtų būti. Kompiuterinių technologijų vystymasis leido panaudoti vis daugiau ir daugiau kreditingumo reitingo prognozavimo modelių kūrimo būdų: pradėta naudoti logistinė regresija, atsitiktiniai miškai, neuroniniai tinklai. Atsirado galimybės dirbti su didžiulėmis ir netvarkingomis duomenų imtimis. Taip natūraliai išaugo ir sričių, kuriose naudojamas matematinis modeliavimas, kiekis (Thomas, Crook, & Edelman, 2002).

#### 1.1.1. Kreditingumo reitingo apibrėžimas

Reitingas – skaitinis įvertis, priskiriamas vartotojui. Naudojant reitingą siekiama numatyti, kaip smarkiai individas bus linkęs atlikti vieną ar kitą veiksmą ateityje. Reitingas skaičiuojamas lyginant vartotojo veiksmo atlikimo tikimybę su kitų vartotojų to paties veiksmo atlikimo tikimybėmis.

Prognozės atliekamos darant prielaidą, jog klientas elgsis panašiai, kaip praeityje elgėsi panašiomis charakteristikomis ir elgsena pasižymintys kiti vartotojai. Atlikus istorinę vartotojų elgsenos ir charakteristikų analizę yra sudaromos tendencijos. Jos padeda geriau suprasti potencialius naujus vartotojus bei atlikti jų elgsenos prognozes (pavyzdžiui, įvertinus charakteristikas ir praeities klientų veiksmus, pasakyti, kaip uoliai naujas vartotojas bus linkęs gražinti paskolą). Finansinės institucijos (bankai, paskolų išdavimo įmonės) naudoja kreditingumo reitingo prognozavimo modelius, siekdamos geriau įvertinti prisiimamą finansinę riziką (pavyzdžiui, vertina tikimybę, jog klientas sugebės gražinti pasiimtą paskolą).

Kreditingumo reitingas sudaromas remiantis kreditingumo istorija: banko sąskaitų kiekiu, išiskolinimų lygiu, skolų gražinimo istorija. Aukštesnį kreditingumo reitingą turintys asmenys ir įmonės gauna geresnes skolinimosi sąlygas ir mažesnes palūkanas, todėl už tą pačią paskolą sumoka mažiau, nei individai, turintys žemesnį kreditingumo reitingą. Mašininio mokymo modeliai analizuoja vartotojų įpročius ir siekia atpažinti tendencijas, kurios padėtų numatyti jų ateities veiksmus. Šiems dirbtinio intelekto modeliams tinkamai funkcionuoti privaloma turėti patikimus duomenų šaltinius, kurie padėtų atpažinti logines sąsajas bei išgauti kokybiškus rezultatus. Naudojami pagrindiniai 5 duomenų šaltiniai, fizinių asmenų atveju, yra vaizduojami 1.1 lentelėje. Kreditingumui nustatyti svarbu atsižvelgti į asmens socialinius ir demografinius duomenis, elgseną, įvertinti kitų finansinių institucijų duomenis (kitas paskolos, komunalinių mokesčių mokėjimo grafikus, „Sodros“ duomenis). Kreditingumui nustatyti dažnai naudinga atkreipti dėmesį ir į psichometrinius įverčius, ankstesnės duomenų analizės metu gautus pastebėjimus.

**1.1 lentelė.** Duomenų šaltiniai modeliuojant kreditingumo įverčius fiziniams asmenims

Duomenų šaltinis	Aprašymas
Socialiniai ir demografiniai duomenys	Vartotojo lytis, amžius, šeimyninė padėtis, užimtumas, darbinė patirtis, būsto tipas, kiek laiko gyvena tam tikroje vietoje.
Elgsenos duomenys	Ankstesni vartotojo santykiai su kredito įmone. Gali būti vertinama ar klientas anksčiau vėlavo gražinti paskolą, kiek daugiausiai dienų klientas yra vėlavęs atlikti mokėjimą, kiek kartų vėlavo gražinti ir kiti aspektai.
Duomenys iš kitų finansinių institucijų	Duomenys surinkti iš trečiųjų šalių įmonių: asmeninė kredito istorija, duomenys apie apmokėtas / neapmokėtas vandens, elektros, telefono, interneto sąskaitas.
Psichometriniai duomenys	Duomenys gaunami klientams atsakius į specialius klausimynus. Klausimais vertinamas vartotojo išsilavinimas, finansinės žinios, asmenybė, požiūris į gyvenimiškas situacijas.
Didžiųjų duomenų analizė	Duomenys iš išorės organizacijų ir socialinių tinklų.

Galima teigti, jog kreditingumo reitingo taikymas visiškai pakeitė skolinimo procesą. Anksčiau, kai reitingai nebuvo sudarinėjami, skolinimo procesas buvo paprastesnis: skolintojai remdavosi tik savo patirtimi ir nuojauta. Tačiau abi pusės nevisada gebėjo įvertinti, ar jų prisiimama rizika atitinka individualią finansinę situaciją. Atsiradus galimybėms sudaryti kreditingumo reitingus ir prognozuoti rizikas, naudos gavo visi. Skolininkai nebegalėjo prisiimti neproporcingos rizikos ir pradėjo atsargiau vertinti savo finansines galimybes. Skolintojai suprato, kad kiekvienas atvejis yra unikalus ir tik tiksliai įvertindami riziką jie gali konkuruoti paskolų rinkoje.



### 1.1.2. Kreditingumo reitingų tipai ir jų teikiama nauda

Kreditingumo reitingai yra charakterizuojami naudojamu duomenų šaltiniu, kredito išdavimo proceso etapu, kredito išdavimo institucija ir reitingo sukūrimo būdu (MICROFINANCE CREDIT RISK MANAGEMENT TOOL GUIDES). 1.2 lentelėje vaizduojami kreditingumo reitingo analizėje naudojami reitingo tipai ir jiems priskiriamos charakteristikos. Atrinkimo, surinkimo bei atnaujinimo reitingai yra priskiriami kredito išdavimo proceso etapo charakteristikai. Šie reitingai naudojami skirtinguose kredito svarstymo etapuose: vertinant naujų klientų atnešamą riziką, skirstant esamus klientus į rizikos grupes bei optimizuojant paskolų išdavimo procesą. Bendrinis ir pritaikomas reitingai charakterizuojami prisitaikymu prie finansinės institucijos. Bendrinis reitingas sudaromas naujiems klientams, apie kuriuos dar neturima finansinių žinių. Pritaikomas reitingas naudoja specifinę vertinimo metodiką bei istorinius duomenis, šis reitingas laikomas itin tikslu. Taip pat išskiriami ir analitinių charakteristikų reitingų tipai: statistinis ir ekspertinis. Statistinis sudaromas remiantis matematiniu modeliavimu bei mašininio mokymo metodais. Ekspertinis reitingo tipas sudaromas atsižvelgiant į specialistų patirtį ir patarimus.

1.2 lentelė. Kreditingumo reitingų tipai

Reitingo tipas	Charakteristika	Aprašymas
Atrinkimo reitingas	Kredito išdavimo proceso etapas	Finansinėse institucijose naudojamas vertinant naujų klientų riziką (didelė, vidutinė, maža). Dažniausiai remiasi socialiniais, demografiniais ir psichometriniais duomenimis. Atsižvelgdamos į prognozuojamą riziką, finansinės institucijos gali parinkti tolimesnes kreditingumo tikrinimo procedūras ir / arba paskolos suteikimo sąlygas.
Surinkimo reitingas	Kredito išdavimo proceso etapas	Naudojamas nustatyti tikimybę, jog esamas klientas laikysis įsipareigojimų ir skolas padengs laiku. Finansinės institucijos, atsižvelgdamos į šį reitingą, gali sudaryti klientų grupes. Jeigu tam tikra grupė vėluoja gražinti skolą, įmonė gali imtis efektyviausių tai grupei veikslių: išsiųsti pranešimą, paskambinti, susitikti asmeniškai. Surinkimo reitingo naudojimas lemia padidėjusį efektyvumą ir sumažėjusias finansinės institucijos išlaidas.
Atnaujinimo reitingas	Kredito išdavimo proceso etapas	Naudojamas optimizuojant klientų atnaujinimo procesą: pritraukiant jau esamus klientus, kurie laiku apmokėjo skolas, sudarant optimalius procesus ir sąlygas didesnės rizikos klientams. Taip pat atnaujinimo reitingas naudojamas gerų klientų atpažinimui bei išlaikymui: siūlant nuolaidas ar specialius pasiūlymus.
Bendrinis (įprastas) reitingas	Prisitaikymas prie finansinės institucijos	Reitingas, kuris dažniausiai apskaičiuojamas naujai įmonei, asmeniui, produktui ar regionui, dar neturint platesnių finansinių institucijų žinių apie juos.
Pritaikomas reitingas	Prisitaikymas prie finansinės institucijos	Reitingas apskaičiuojamas naudojant specifinę finansinės institucijos vertinimo metodiką. Rekomenduojama, kad tokio tipo reitingui sudaryti būtų naudojami bent dviejų metų istoriniai duomenys.
Statistinis reitingas	Analitinis nustatymas	Duomenų tarpusavio sąryšiai identifikuojami naudojant matematinius modelius ir įvairius statistinius paketus (pavyzdžiui, SPSS, R ir Python programavimo kalbų bibliotekos). Statistiniam reitingui sudaryti dažniausiai naudojami logistinės regresijos, sprendimų medžių bei neuroninių tinklų metodai.

Ekspertinis reitingas	Analitinis nustatymas	Naudoja srities ekspertų žinias ir praeities patirtį, atliekant vartotojų elgsenos prognozę.
-----------------------	-----------------------	----------------------------------------------------------------------------------------------

(Caire, et al., 2006)

Dažniausiai finansinės institucijos bei paskolų išdavimo bendrovės kreditingumo reitingų vertes atvaizduoja raidžių nominalais. 2020 metais didžiausios ir svarbiausios kredito agentūros yra *Moody`s*, *Standard and Poor`s*, *Fitch* ir *DBRS*. 1.3 lentelėje pateikiami agentūrų reitingų nominalai nuo aukščiausio – didžiausia tikimybė įsiskolinimą gražinti laiku, iki žemiausio – mažiausia tikimybė skolą gražinti laiku (<https://corporatefinanceinstitute.com/resources/knowledge/finance/credit-rating/>, 2020). Matome, jog svarbiausios kredito agentūros kreditingumo reitingus aprašo labai panašiai: geriausi žymimi keliomis A raidėmis, blogiausi – C ar net D raidėmis. Kartais išskyrimui naudojami + ir – simboliai, didžiosios ir mažosios raidės.

### 1.3 lentelė. Svarbiausių kredito agentūrų reitingų nominalai

Moody`s	S&P	Fitch	DBRS
Aaa	AAA	AAA	AAA
Aa1	AA+	AA+	AA (high)
Aa2	AA	AA	AA
Aa3	AA-	AA-	AA (low)
A1	A+	A+	A (high)
A2	A	A	A
A3	A-	A-	A (low)
Baa1	BBB+	BBB+	BBB (high)
Baa2	BBB	BBB	BBB
Baa3	BBB-	BBB-	BBB (low)
Ba1	BB+	BB+	BB (high)
Ba2	BB	BB	BB
Ba3	BB-	BB-	BB (low)
B1	B+	B+	B (high)
B2	B	B	B
B3	B-	B-	B (low)
Caa1	CCC+	CCC+	CCC (high)
Caa2	CCC	CCC	CCC
Caa3	CCC-	CCC-	CCC (low)
-	D	D	D

(<https://corporatefinanceinstitute.com/resources/knowledge/finance/credit-rating/>, 2020)

Finansinės institucijos, naudodamos kreditingumo reitingą ir atlikdamos jo pokyčių analizę, gauna apčiuopiamos naudos keletu aspektų:

1. Kokybės užtikrinimas. Reitingavimas užtikrina rizikingumo išmatavimą, randama, kurie klientai nepatenka į norimus rizikos režius.

2. Nuoseklus sprendimų priėmimas. Nebereikia remtis vien tik už paskolų išdavimą atsakingo darbuotojo suvokimu ir įžvalgomis, kurios kartais gali nukrypti nuo standartų. Kreditingumo reitingo prognozavimo algoritmai prognozes išlaiko maksimaliai nuoseklias ir integalias.
3. Proceso optimizavimas ir kontrolė. Užtikrinamas aukšto lygio skaidrumas, darbuotojai nebegali paprastai piktnaudžiauti savo padėtimi išduodant kreditus.
4. Glaudesnis ryšys su klientais, klientų pasitenkinimo paslaugomis didinimas. Vartotojų skirstymas į grupes padeda pritaikyti personalizuotus sprendimus, rasti efektyviausius bendradarbiavimo būdus ir pritaikyti optimaliausias sąlygas.
5. Kreditingumo institucijų veiklos efektyvumo gerinimas. Sumažėja prisiimama rizika; procesas tampa spartesnis; didėja darbuotojų produktyvumas (dalis darbo atliekama naudojant algoritmą); dėl didesnio vartotojų pasitenkinimo bei gerų klientų skatinimo išlaikomi skolas laiku grąžinantys klientai.

Kreditingumo reitingai gali būti sudaromi remiantis skirtingomis metodikomis. Nėra sukurto optimalaus būdo sudaryti kreditingumo reitingus. Finansinės institucijos pasirenka jų manymu svarbiausius kriterijus, šie deriniai gali ir stipriai skirtis vienas nuo kito. Svarbu suprasti, jog tie kreditingumo reitingai, kurie buvo sudaryti remiantis realiomis finansinėmis žiniomis daug tiksliau parodo realią situaciją, nei reitingai, kurie remiasi išoriniais ir nepagrįstais rodikliais ar politinio poveikio priemonėmis.

## 1.2. Kreditingumo reitingo taikymo procesas finansinėse institucijose

1.4 lentelėje vaizduojami potencialūs kreditingumo reitingavimo privalumai ir iššūkiai, su kuriais susiduria skirtingo tipo finansinės organizacijos. Organizacijos išskiriamos kaip: turinčios centralizuotą arba decentralizuotą kreditingumo vertinimo valdymą, turinčios didelę arba ribotą prognozavimo modeliutojų komandą, turinčios arba neturinčios išvystytas IT sistemas. Apgalvotas ir nuoseklus kreditingumo reitingavimas gali duoti naudos visoms jį taikančioms organizacijoms. Bandoma paaiškinti (žr. 1.4 lentelę), kaip kreditingumo reitingo sistemos yra sukuriamos, pritaikomos bei palaikomas jų funkcionalumas (Caire, et al., 2006).

**1.4 lentelė.** Organizacinių kreditingumo vertinimo įverčių privalumai ir galimi iššūkiai

Faktoriai, į kuriuos reikėtų atsižvelgti prieš kuriant kreditingumo reitingavimo sprendimą			
Organizaciniai įverčiai		Privalumai	Iššūkiai
Kreditingumo vertinimo valdymas	Decentralizuotas	Geriausi ir prasčiausi vartotojai rečiau susiduria su informacijos trūkumu. Standartizuojami skolinimo sprendimai, kaina.	Užtikrinti bendradarbiavimą tarp atskirų filialų. Užtikrinti vienodą kreditingumo reitingo naudojimą bei vieną politiką. Reitingavimo duomenų valdymas.
	Centralizuotas	Sparčiai priimami sprendimai. Modelius mato ir naudoja tik nedidelė analitikų grupė.	Reikalingas patikimas bendradarbiavimas tarp filialų ir kredito centro. Mažesnis proceso skaidrumas.

Žmogiškieji ištekliai	Aukšto lygio modeliavimo/IT specialistai	Sumažėjusios projekto kainos. Modelis gali būti nuolat atnaujinamas ir pritaikomas prie besikeičiančios situacijos. Programavimo darbai gali būti atliekami bet kuriuo metu.	Reikia įsigyti statistinio modeliavimo programinę įrangą. Specialistai gali turėti tik teorinių statistikos žinių ir jokios modeliavimo patirties. IT specialistai gali būti užsiėmę kitais svarbiais projektais.
	Žemesnio lygio modeliavimo / IT specialistai	Išorinės konsultacijos gali suteikti naujų žinių ir patirties. Nereikia pirkti statistinio modeliavimo programinės įrangos.	Išorinis modeliavimas padidina kainas. Priklausomybė nuo išorės programuotojų. Intelektinės nuosavybės dalybos. Pardavėjai nori, kad projektui vadovautų kompetentingas asmuo.
IT sistemos	Išvystytos ir lanksčios IT sistemos	Į jau išvystytas IT sistemas lengviau integruoti reitingavimą. Galima saugoti ir pasiekti istorinius reitingavimo duomenis. Galimybė dirbti su duomenimis, kurti ataskaitas.	Gali būti neįmanoma redaguoti išeities kodo.
	Besivystančios IT sistemos (mažesnis funkcionalumas / lankstumas)	Atsiranda postūmis automatizuoti dokumentavimo procesą.	Didėja kainos, nes reikia vystyti ir IT sistemą. Reikia skirti papildomą laiką IT sistemos vystymui. Gali prireikti atnaujinti ir fizinę įrangą.
Kreditingumo reitingavimo patirtis	Didelė patirtis atliekant kreditingumo reitingavimą	Vadovybė geriau supranta kūrimo procesą bei gali efektyviau organizuoti darbus. Optimaliau įvertinamos rizikų kainos, naudojamas vartotojų segmentavimas lojalumo programoms.	Gali išsiskirti nuomonės dėl kreditingumo reitingavimo tipų. Per didelės viltys gauti gerus rezultatus net ir turint labai prastus duomenis.
	Nedidelė patirtis atliekant kreditingumo reitingavimą	Pritaikius išorinę pagalbą galima išvystyti tinkamiausią reitingavimo sprendimą.	Neįvertinamos visos kreditingumo reitingavimo galimybės. Netinkamose vietose atliekami kompromisai. Lėtesnis kūrimo procesas.

(Caire, et al., 2006)

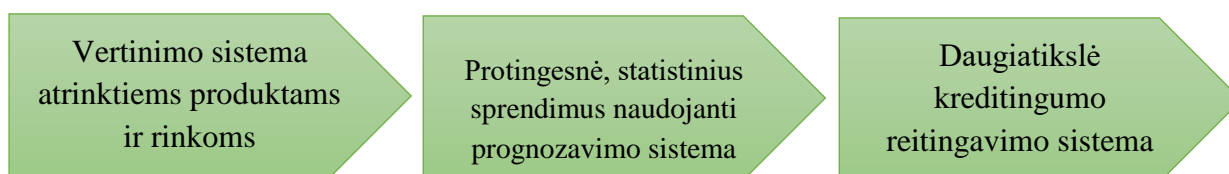
Praktikoje dažnai naudojama 4 žingsnių sistema, skirta sukurti ir įgyvendinti nuosavą kreditavimo reitingavimą. Visos sistemos dalys yra lanksčios ir apima skirtingas metodologijas, todėl gali būti taikomos bet kuriai finansų institucijai, turinčiai aiškų planą ir patirtį skolinimo srityje.

### 1.2.1. 1 žingsnis: Nuspręsti į kokį segmentą bus orientuojamasi.

Pirmas žingsnis kreditavimo reitingo nustatyme yra vartotojų ir finansinių paslaugų, kurioms bus taikomi modeliai nustatymas. Šį sprendimą turėtų priimti ne kažkurio skyriaus vadovas ir ne 1 skyrius. Įmonė turi sudaryti darbo grupę iš visų funkcinių sričių, kurioms kreditavimo reitingas yra aktualus (kreditavimo rizika, kredito išdavimas, marketingas, IT, konsultacijos). Ši darbo grupė turi prižiūrėti reitingo sudarymo ir įgyvendinimo ypatybes. Priimtas sprendimas turi tenkinti visų funkcinių sričių poreikius (Caire & Kossman, Credit Scoring: Is it Right for Your Bank?, 2003).

Jeigu su kreditavimo reitingavimu įmonė susiduria pirmą kartą, ekspertai rekomenduoja reitingavimo sistemą vystyti ir pritaikyti palaipsniui. 1.1 paveikslėlyje vaizduojami reitingavimo sistemos vystymo laipsniai. Pirmiausiai atrinktiems produktams ir rinkoms sudaroma vertinimo sistema. Vėliau sudarytai sistemai pritaikomi statistiniai sprendimai ir vertinimo sistema tampa prognozavimo sistema. Galiausiai prognozės pritaikomos kelių tikslų kreditavimo reitingavimui atlikti. Sistemos kūrimas atliekamas pažingsniui, pavyzdžiui:

1. Rankiniu būdu sukuriami kreditavimo sprendimų priėmimo rezultatų kortelė.
2. Kortelė perdaroma į automatizuotą sprendimą – programą (ši nustato tikimybę, jog paskola nebus gražinta).
3. Įvertinami statistiškai besivystantys veiksniai (įvertinama tikimybė, kad paskola bus gražinta vėluojant).
4. Rinkinių vertinimas (įvertinama tikimybė, kad paskola, kurią jau vėluojama gražinti x dienų, vėluos dar y dienų).
5. Sukuriamas kreditavimo reitingus naudojantis apgavysčių prevencijos algoritmas.



1.1 pav. Reitingavimo sistemos vystymo laipsniai

(Caire, et al., 2006)

### 1.2.2. 2 žingsnis: Pasirinkti naudojamą kreditavimo reitingo tipą

Naudojami 3 abstraktūs kreditavimo reitingų tipai:

- **Statistinis:** rezultatai yra išvedami empiriškai iš ankstesnių paskolų duomenų.
- **Ekspertinis:** sudaromas remiantis ekspertų nuomone ir institucine patirtimi.
- **Hibridinis:** gaunamas apjungus statistinį bei ekspertinį metodus.

Statistiniai kreditavimo reitingo modeliai, naudojami prognozuojant tikimybę, kad individas nesugebės laiku gražinti paskolos. Ekspertinio tipo kreditavimo reitingo modeliai išsiskiria tuo, kad pasiekiamas didelis prognozavimo tikslumas. Todėl šis tipas vertinamas kaip pats geriausias atliekant

rizikos valdymą, kainos nustatymą bei paskolos parengimą. Ekspertiniai ir hibridiniai kreditingumo reitingo modeliai naudojami vertinant kelių pasirinkimų santykinę riziką. Svarbus ekspertinių ir hibridinių modelių privalumas – galimybė atlikti atgalinį testavimą naudojant istorines situacijas. Tokiu būdu kiekvienam kreditingumo vertinimui galima nustatyti istorinę tikimybę, jog paskola nebus gražinta.

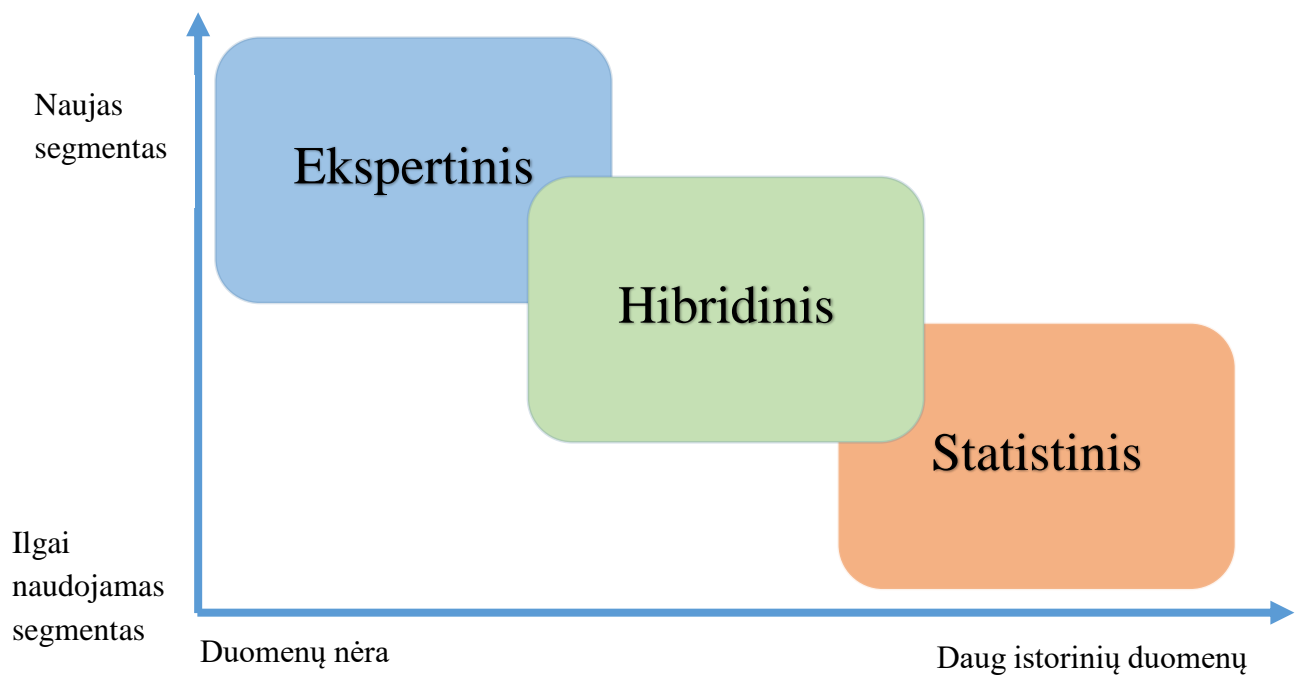
Realybėje visi kreditingumo vertinimai naudoja statistinio ir ekspertinio vertinimų junginį. Ekspertų vertinimas gali būti naudojamas ten, kur trūksta duomenų statistiniam modeliavimui. Tokiu atveju srities ekspertų įžvalgos dažniausiai būna tikslesnės nei mašininio mokymo algoritmų. Tačiau atgalinis modelių rezultatų ir parametrų testavimas vis tiek turi būti atliekamas. Hibridinių kreditingumo reitingo koeficientai apskaičiuojami sujungiant statistiškai išvestus koeficientus su koeficientais gautais ekspertiniu būdu.

Kreditingumo reitingo tipas parenkamas atsižvelgiant į šias dedamąsias:

1. Istorinių duomenų apie geras ir prastas paskolas kiekis ir kokybė.
2. Skirtumai tarp naujų ir senų klientų duomenų.
3. IT komandos modeliavimo gebėjimai, arba kaina samdant išorės modeliuotojų komandą.
4. Naudojamų IT sistemų apribojimai.

Svarbiausias faktorius – istorinių duomenų kiekis ir kokybė. Turint mažos apimties arba nekokybiškus duomenis, sunku taikyti tiek statistinį tiek ekspertinį metodus ir dažnai pradedama tiesiog spėlioti. Antras svarbiausias faktorius – kaip smarkiai įmonė tikisi, kad jos ateities modeliavimo duomenys bus susiję su praeities duomenimis. Jeigu duomenys bus panašūs, prognozavimo kokybė modeliuojant išaugs. Ir priešingai, jei ateities duomenys smarkiai skirsis nuo naudotų praeityje, tai pagal praeities duomenis sudaryti modeliai atliks nepagrįstas ir netikslias prognozes.

Paveikslėlyje 1.2 iliustruojamas optimalaus reitingo tipo pasirinkimo sprendimo priėmimas. X ašis parodo turimų istorinių duomenų kiekį, o Y ašis vaizduoja kaip smarkiai duomenų segmentas skiriasi nuo ankstesnių (istorinių) duomenų. 3 stačiakampiai viduryje parodo rekomenduojamų modelių tipus skirtingose situacijose. Kaip matome, jei naudojamas naujas duomenų segmentas ir neturime jokių istorinių duomenų, verta rinktis ekspertinį kreditingumo reitingo sudarymo tipą. Priešingu atveju – jei turime daug istorinių duomenų ir duomenų segmentas yra naudojamas ilgą laiką, rinksimės statistinį kreditingumo reitingo sudarymo tipą. Tarpiniais atvejais turėtume bandyti sujungti abiejų tipų koeficientus ir formuoti hibridinius metodus (Caire, et al., 2006).



1.2 pav. Kreditingumo reitingo tipo pasirinkimas

### 1.2.3. 3 žingsnis: Suprojektuoti kreditingumo reitingavimo modelius

Kreditingumo reitingo projektavimas yra atliekamas 3 etapais:

- **Apibrėžimas:** Nustatoma kokie kriterijai apibrėžia pelningą paskolos ir „gero“ skolininko sąvoką bei kokie kriterijai aprašo nuostolingą paskolą ir „blogą“ skolininką.
- **Tiriamoji analizė:** Įvertinama, kurie parametrai daro didžiausią įtaką prisiimamos rizikos įverčiui.
- **Vystymas:** Apskaičiuojama kurių faktorių kombinacijos leidžia pasiekti geriausius rezultatus atgaliniame testavime.

#### Apibrėžimo etapas

Kiekviena finansinė institucija, užsiimanti paskolų išdavimu turi apibrėžti ką laiko „blogu“ klientu. „Blogas“ dažniausiai yra tas, kurio bus stengiamasi vengti ir neišduoti paskolos. Atliekant statistinio modelio projektavimą svarbus tiksliai ir išmatuojamai apibrėžti kas yra „blogas“ klientas. Tai atliekama, nes modelis sukuria sąryšius tarp rizikos vertinimų ir netinkamų paskolų. Ekspertinis modelis gali būti sukuriamas ir neaprašius tikslios „blogas“ reikšmės, bet ją vis tiek reikia aprašyti modelį patvirtinant ir testuojant. Mikrofinansų institucijos dažnai pradeda klientus laikyti „blogais“, jei šie vėluoja grąžinti paskolą daugiau nei 15 dienų. Komerciniai bankai laukia net iki 90 ar dar daugiau dienų, prieš pradėdami klientą laikyti nemokiu. Visą laikotarpį paskolas išduodantys bankai perskaičiuoja tikimybes, kad klientas taps nemokiu. Norint išvengti 2007-2008 metų bankų krizės scenarijaus, labai svarbu, jog įmonių - skolintojų nustatytas „blogo“ skolininko apibrėžimas atitiktų jų prisiimamą riziką.

#### Tiriamosios analizės etapas

Įžvalgų analizės metu siekiama išskirti tas charakteristikas, kurios bus naudojamos kaip modeliavimo parametrai. Statistinio modelio parametrum nustatymui analizuojami visi prieinami

besiskolinančiojo asmens duomenys. Lyginama, kurie kintamieji yra naudingi ir kurie neneša papildomos vertės. Parametrų išrinkimui gali būti naudojami įvairūs programinės įrangos paketai: *STATA*, *SAS*, *SPSS*, taikomosios *Python*, *R*, *MATLAB* kalbų bibliotekos ir t.t. Tiriamosios analizės metu turėtų būti gaunamas charakteristikų, kurias vertėtų įtraukti į modelį, sąrašas. Taip pat atlikus tiriamąją analizę turėtų išaiškėti charakteristikų tarpusavio ryšys ir kaip jos susijusios su paskolos grąžinimu. Tiriamaoji analizė dažnai atliekama srities ekspertų pasiūlytus faktorius papildant jų statistiniais įverčiais. Analizuojamų charakteristikų kiekis smarkiai priklauso nuo kiekvienos organizacijos veiklos: taikomo duomenų rinkimo, analizės būdų bei vietinės rinkos situacijos.

Dažniausiai modeliavime naudojamų (didžiausią skaičiavimo vertę generuojančių) kreditingumo klientų charakteristikos:

- Demografiniai įverčiai: lytis, gimimo metai, šeimyninė padėtis, išsilavinimas.
- Informacija apie namų ūkį: gyventojų skaičius namų ūkyje, namo valdytojo statusas, kiek metų gyvenama namų ūkyje, kambarių kiekis, automobilių kiekis.
- Verslo demografiniai įverčiai: šalis, sektorius, verslo tipas, įmonės veiklos trukmė, darbuotojų skaičius (taikoma, kai paskola suteikiama įmonei).
- Finansiniai įverčiai: verslo įstaigos pajamos, namų ūkio pajamos, nuomos kaina.
- Turtas: bendras turtas, ankstesni įsiskolinimai, nuosavo kapitalo kiekis.
- Paskolų grąžinimo istorija: paskolų kiekis, ilgiausias įsiskolinimas.
- Kredito biuro informacija: kredito biuro finansinis reitingas, buvimas „juodajame sąraše“.
- Subjektyvūs įverčiai: sugebėjimas vadovauti įmonei, pinigų srauto kintamumas, verslo perspektyvos (taikoma, kai paskola suteikiama įmonei).
- Prašomos paskolos charakteristikos: paskolos suma, paskolos tipas, paskolos gavėjo indėlis į finansavimą (taikoma, kai paskola suteikiama įmonei).

(Schreiner, 2003)

### **Vystymo etapas**

Vystymo etape, sudarant kreditingumo reitingo nustatymo modelį, kiekvienam faktoriui priskiriamas tam tikras svoris. Svoriai statistiniam metodui yra gaunami panaudojant statistinių metodų išeišgas (pavyzdžiui naudojant regresijos skaičiavimą). Ekspertiniuose modeliuose svoriai nustatomi rankiniu būdu, remiantis nuspėjama individualių faktorių svarba bei jų tarpusavio sąveika. Nepriklausomai nuo to, kokio tipo modelis yra kuriamas, svarbu sistemingai dokumentuoti vystymo proceso žingsnius. Vystymo proceso dokumentavimas susideda iš informacijos apie:

- Pagrindinius priimtus sprendimus.
- Taikytus duomenų atrinkimo metodus.
- Duomenų kokybės užtikrinimo iššūkius.
- Duomenų tvarkymo / manipuliavimo eigą.
- Informacijos apie atliktą atgalinį testavimą.

Paskolą išduodančios įmonės darbuotojai, turėdami tokio tipo dokumentaciją, gali atpažinti įvykusias klaidas, jas ištaisyti bei atlikti kitus būtinausius pakeitimus (Caire, et al., 2006).



#### **1.2.4. 4 žingsnis: Ištestuoti, paruošti naudojimui, prižiūrėti kreditingumo reitingavimo modelius**

##### **Atgalinis testavimas**

Pirmas svarbus kreditingumo modelio išbandymas yra parodyti tikslus rezultatus atgalinio testavimo metu. Atgalinis testavimas atliekamas panaudojant istorinius duomenis ir išsiaiškinant kaip tiksliai modelis būtų veikęs praityje. Tokiu būdu atliekama analizė įvertina riziką kredito įstaigoms neprarandant realių pinigų. Atgalinio testavimo metu pasiekti rezultatai yra svarbiausias parametras parodantis modelio tikslumą. Labai svarbu neužteršti testavimo duomenų imties duomenimis naudotais modelio apmokymui ir parametru parinkimui, nes modelis gali prisitaikyti prie vienu ar kitu duomenų ir „persimokyti“ (Caire, et al., 2006).

##### **Pilotinis testavimas**

Pilotinio testavimo tikslas yra išbandyti modelį praktikoje: nustatyti kokiam intervale gaunami reitingų rezultatai? Ar reitingavimo rezultatai atitinka prisiimtą rizikos lygį? Ar iškyla neplanuotų keblumų renkant duomenis modeliavimui? Atsakymai į šiuos klausimus gali pakeisti reitingavimo modelio taikymą bei kitas procedūras.

Pilotinis testavimas gali trukti ilgesnį laiko tarpą, pavyzdžiui 3-6 mėnesius arba tol kol bus įvertintas tam tikras užklausų dėl paskolų kiekis, pavyzdžiui 500 užklausų. Pilotinis testavimas gali būti atliekamas tuo pačiu metu kaip ir kitos procedūros. Kredito išdavimo ekspertai nuolatos stebi ir jei reikia atnaujina prognozavimo modelį: tiek prieš išduodant paskolą, tiek ir po jos išdavimo. Nauji modeliai ir kreditingumo vertinimo sprendimai taip pat gali būti išbandomi atskirais atvejais tik tam tikruose įmonės padaliniuose. Pavyzdžiui, mieste A kreditingumas vertinamas taikant modelį A, o mieste B - taikant modelį B. Remiantis atvejų analizėmis, nėra griežtų pilotinio testavimo reikalavimų ar trukmės suvaržymų. Tiksliausias paskolos išdavimo sprendimas turėtų priklausyti nuo duomenų tipo, kiekio ir sprendimų priėmimo proceso (centralizuoto arba decentralizuoto) (Caire, et al., 2006).

##### **Darbuotojų apmokymas**

Pilotinis testavimas turi apimti ir darbuotojų paruošimą naudotis modeliu ir jo rezultatais. Modelio naudotojai turėtų suprasti kaip buvo sukurtas modelis, kokie yra organizacijos modelio naudojimo tikslai bei perprasti visas su modeliu susijusias procedūras. Labai svarbu testavimo periodu analizuoti tiesioginių modelio vartotojų atsiliepimus, taip randant potencialius modelio trūkumus. Vadovai turi būti apmokyti suprasti pagrindinius principus, kaip veikia modelis ir į ką atsižvelgiant skaičiuojamas kreditingumas. Rizikos vertintojai ir IT departamentas turi išanalizuoti modelio parametrus ir suprasti kaip jie gali būti modifikuojami. Žmogiškųjų išteklių departamentas privalo užtikrinti, kad, pasikeitus modeliavimo procesui, už paskolų išdavimą atsakingi asmenys gautų reikalingą apmokymą (Caire, et al., 2006).

##### **Integravimas su informacinėmis sistemomis, ilgalaikė priežiūra**

Finansinėse institucijose reitingavimo modelis dažniausiai kuriamas kaip jau egzistuojančios IT sistemos modulis. Svarbu šį modulį tinkamai integruoti į sistemą, atidžiai apgalvoti sistemos architektūrą bei duomenų srautų judėjimą. Kuo sudėtingesnis prognozavimo modelis, tuo sunkiau jis

gali būti integruojamas į IT sistemas. Dėl šios priežasties kartais yra kuriamos atskiros modelio aplikacijos, kurios yra susiejamos su pagrindine duomenų baze.

Kreditingumo modeliai naudojami ne vien tik sprendžiant paskolos atsipirkimo klausimą. Jie gali atlikti ir svarbų vaidmenį duomenų valdyme. Pavyzdžiui naudinga saugoti vartotojų duomenis, kuriems iš pradžių buvo suteikta paskola, bet vėliau ji buvo atšaukta. Taip pat verta įsiminti atvejus, kai modelio sprendimas buvo keičiamas įsikišus atsakingiems darbuotojams. Tokio tipo ir panaši informacija turėtų būti saugoma vėlesnei analizei bei galimam esamo modelio tobulinimui. Siekiant užtikrinti aukštą kokybę, pravartu nuolat atnaujinti šiuos vartotojų elgsenos modelius.

Modelių valdymas bei priežiūra yra ilgalaikis procesas, kuris turi nesibaigti ties modelio sukūrimu ir realizavimu produkcinėje aplinkoje. Norint užtikrinti kokybišką prognozavimo modelio veikimą svarbu nuolat rinkti naujus duomenis, juos saugoti, stebėti duomenų bei galimų sprendimų pasikeitimus. Situacija realiame pasaulyje keičiasi kasdien: atsiranda naujos technologijos, finansinės krizės, mados, nauji poreikiai, nauji vartotojų segmentai. Todėl svarbu, kad ir modeliai nepasentų bei išmoktų teisingai atsakyti į kasdien besikeičiančius klausimus.

(Caire, et al., 2006)

Poskyryje išanalizavome kreditingumo reitingų nustatymo ir taikymo procesą finansų institucijose. Toliau išanalizuojame kreditingumo reitingų agentūrų sprendimų priėmimo kriterijus.

### **1.3. Kreditingumo reitingų agentūrų sprendimų priėmimo kriterijai**

Kiekviena kreditingumo reitingo agentūra turi savo unikalią metodologiją, kuri įvertina skirtingus šaltinius, situacijas bei tiriamus faktorius. Tačiau agentūrų taikomi metodai turi tam tikrų bendrų bruožų. Šiame poskyryje pateikiami kriterijai, kuriais remiasi svarbiausios kreditingumo reitingo agentūros sudarydamos bankų kreditingumo reitingus. Faktorių grupėmis įvertinamos bankų galimybės gražinti skolas be išorinės pagalbos. Grupės yra sudarytos iš makro faktorių analizės bei individualių bankų veiksmų.

Pirmas veiksnys, kurį įvertina kreditingumo reitingų agentūros – bankų makroekonominė aplinka. Moksliskai įrodyta, jog makroekonominių faktorių analizė gali padėti numatyti bankų žlugimą. Analizė yra sudaroma iš 2 dalių. Pirmoje dalyje įvertinamos makroekonominės rizikos bankų veiklos šalyse. Antroje dalyje analizuojamos specifinės charakteristikos ir bankų sektoriaus situacija atitinkamose valstybėse. Kreditingumo reitingų agentūros, remdamosios makroekonominiais faktoriais, priskiria rizikos balą šalims, kuriose veikia vertinami bankai. Jeigu bankai veikia keliuose valstybėse, balas yra skaičiuojamas nustatant tiriamų šalių svorinį rizikos vidurkį. Įvertinamos valstybių ekonominės struktūros, jų stabilumas, ekonominės politikos tendencijos ir šalies ekonomikos dalyvių rizikos charakteristikos. Analizuojant ekonominį stabilumą atsižvelgiama į bendrojo vidaus produkto augimą ir jo standartinį nuokrypį per paskutinį vertinimo ciklą. Vertindamos šalies institucinę situaciją, agentūros atsižvelgia į viešųjų įstaigų vientisumą ir teisinės sistemos garantijas. Kreditingumo reitingų agentūros, vertindamos pramonės rizikas, atsižvelgia į 3 struktūrinius bruožus:

1. Bankų industrijos reglamentų kokybė ir veiksmingumas.
2. Konkurencinė dinamika.
3. Vyriausybės ir centrinių bankų įtaka ekonomikai.

Antras veiksnys, kurį įvertina kreditingumo reitingų agentūros – individuali nagrinėjamo banko situacija. Nagrinėjami faktoriai gali būti dalinami į 4 grupes: turto riziką, kapitalą ir pelningumą, finansavimo struktūrą bei kokybinius veiksnius. Kreditingumo agentūros, vertindamos bankų turto portfelius, nuolatos stebi bankų investicijas ir prisiimamų rizikų spektrą. Vertinama bankų rizikos koncentracija ir jų diversifikacija. Analizuojant bankų kapitalą ir pelningumą tikrinama ar bankai atitinka visus keliamus norminius reikalavimus, ar geba atkurti kapitalą per nepaskirstytą pelną. Finansavimo struktūros analizės metu yra lyginama banko finansavimo sudėtis bei vidutinė įstaigų finansavimo sudėtis vietinėje rinkoje. Taip pat atsižvelgiama į bankų galimybes valdyti jų likvidumo poreikius esant nepalankioms sąlygoms. Kokybiniai veiksniai susideda iš sunkiau išmatuojamų bei vertinamų aspektų. Šie veiksniai gali apimti bankų „stabilumą“, verslo diversifikavimą, valdymo kokybę ir ilgilaikę strategiją. Kokybiniai veiksniai dažnai yra susiję tarpusavyje, todėl agentūros paprastai juos sujungia.

Kreditingumo reitingų agentūros taip pat įvertina ir potencialią pagalbą kurią bankams gali suteikti jų filialai arba šalių vyriausybės. Atliekant bankų filialų pagalbos vertinimą atsižvelgiama į filialų tarpusavio santykius ir filialų galimybes suteikti finansinę pagalbą (analizuojama ar filialai geba suteikti pagalbą, ar galbūt jiems patiems gali būti reikalinga pagalba). Potenciali vyriausybių pagalba labia smarkiai priklauso nuo valstybių kreditingumo įverčių. Svarbiausi faktoriai į kuriuos atsižvelgia kreditingumo agentūros: ar bankai yra vyriausybių laikomi „sisteminė svarba“ ir, ar anksčiau, esant sudėtingoms finansinėms situacijoms, tiriamų šalių vyriausybės suteikdavo pagalbą bankams.

Poskyryje išanalizavome kreditingumo reitingų agentūrų sprendimų priėmimo kriterijus. Toliau apžvelgiamo darbe tiriamas kreditingumo reitingų agentūras bei reitingo dinamiškumo sąvoką. Reitingų pokyčių analizė yra naudinga ieškant koreliacijų tarp finansinių rodiklių ir kreditingumo reitingų pasikeitimų.

(Fernandez & Vila, 2015)

#### **1.4. Kreditingumo reitingų agentūros ir pavyzdinė išlikimo analizė**

Kreditingumo reitingų agentūra, tai įmonė, kuri yra atsakinga už kreditingumo reitingo taikymą ne fiziniams asmenims, bet valstybėms bei įmonėms. Reitingu vertinamos skolininkų galimybės laiku padengti įsiskolinimus.

Kreditingumo reitingo sudarymo sektorius yra valdomas 3 didžiausių pasaulyje kreditingumo reitingų agentūrų: „Standard & Poor`s“, „Moody`s“ ir „Fitch“. Jos dažnai vadinamos „Didžiojo trejeto“ kreditingumo agentūromis. „Standard & Poor`s“ bei „Moody`s“ kreditingumo reitingų agentūros valdo maždaug po 40 % rinkos dalies ir „Fitch“ – apie 15 %.

Nuolatos abejojama kreditingumo reitingo sudarymo pagrįstumu. Pavyzdžiui, milijonai vertybinių popierių, kuriems buvo skiriami aukščiausi kreditingumo reitingai, visiškai nuvertėjo per 2007 – 2008 metų finansinę krizę. Iš karto po finansinės krizės (2011m. – 2013m.) Europos Sąjunga aptarė griežtesnes taisykles dėl kreditingumo reitingo agentūrų. Buvo kritikuojama, jog kreditingumo reitingų agentūros, priskirdamos žemus kreditingumo reitingus, sudaro sunkumų po krizės atsigaunančioms valstybėms gauti paskolas. Buvo reikalaujama, jog kreditingumo reitingų agentūros tiksliau pagrįstų savo nustatomus reitingus ir negalėtų valstybėms teikti politinių rekomendacijų.

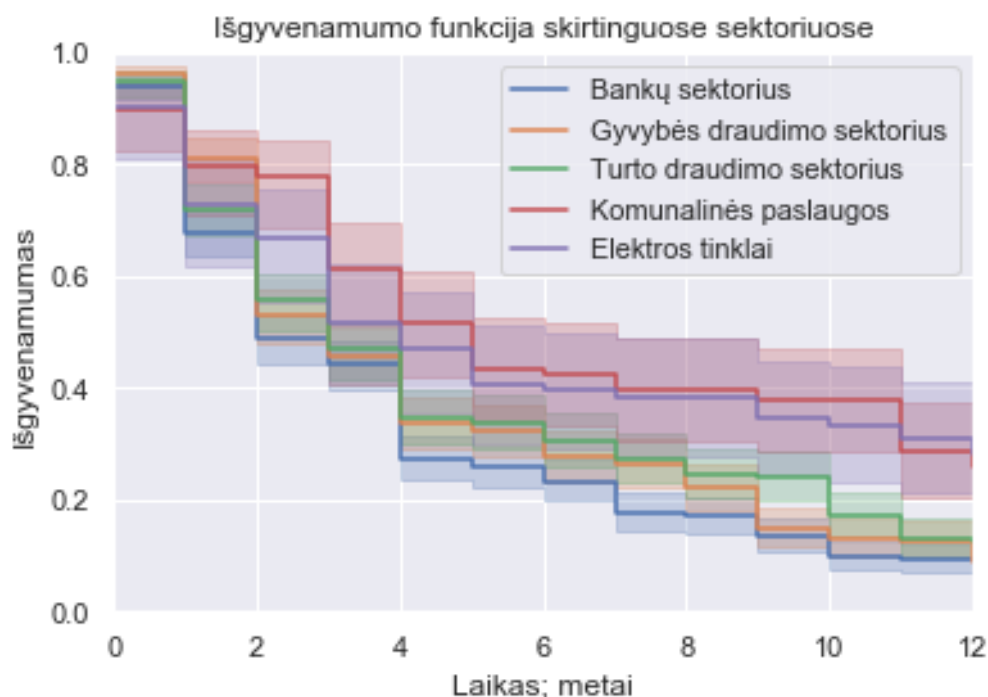
Vienas iš būdų kreditingumo reitingų sudarymo pagrįstumo analizei atlikti – išlikimo analizė. Šiame poskyryje pateikiamas pavyzdinis išlikimo analizės taikymas kreditingumo reitingo pokyčiams nustatyti.

Apžvelgiamas pavyzdinis darbe naudojamas kreditingumo reitingų pokyčių duomenų rinkinys. Kaip pavyzdys pateikiami „Standard & Poor`s” kreditingumo reitingo agentūros duomenys. Šiame rinkinyje fiksuojami visų sektorių (ne vien tik bankų) kreditingumo reitingų pokyčiai 2007 – 2019 metais. Lentelėje (žr. 1.5 lentelę) matome 5 verslo sektorius, turinčius daugiausiai tiriamų įmonių. Daugiausia pateikta duomenų apie bankų sektoriaus (514 unikalios įmonių), gyvybės draudimo sektoriaus (386 unikalios įmonės) bei turto draudimo sektoriaus (365 unikalios įmonės) įmones. Duomenų apie kitų sektorių įmonių kreditingumo pokyčių gauta kelis kartus mažiau.

**1.5 lentelė** Tiriamų įmonių kiekiai verslo sektoriuose

Verslo sektorius	Įmonių kiekis
Bankų sektorius	514
Gyvybės draudimo sektorius	386
Turto draudimo sektorius	365
Komunalinės paslaugos	108
Elektros tinklai	81

Ištiriamas kreditingumo reitingų pokyčių duomenų rinkinio išlikimas. Ieškoma kiek praėjo laiko nuo tyrimo pradžios (2007 sausio 1 dienos), kol kreditingumo reitingo reikšmė pirmą kartą mažėjo tam tikrame sektoriuje. Pavyzdyje (žr. 1.3 pav.) vaizduojamas kreditingumo reitingo išlikimas skirtinguose sektoriuose. Nagrinėjant išlikimo funkcijas skirtinguose sektoriuose pastebimos kelios tendencijos. Po pirmų metų visuose sektoriuose nustatytas labai panašus išlikimas. Vėliau išlikimas išsiskyrė į 3 grupes. Ilgiausiai kreditingumo reitingas nekrito komunalinių paslaugų ir elektros tinklų sektoriuose (praėjus 5 metams apie pusę šių sektorių įmonių vis dar nebuvo patyrusios kreditingumo reitingo kritimo). Šiek tiek dažniau kreditingumo reitingas mažėjo gyvybės draudimo ir turto draudimo sektoriuose (praėjus 5 metams maždaug trečdalis šių sektorių įmonių nepatyrė kreditingumo reitingo kritimo). Dažniausiai kreditingumo reitingas krito bankų sektoriuje (per 5 metus tik ketvirtadalis bankų nepatyrė reitingo sumažėjimo). Panašios tendencijos išsilaikė visą laikotarpį: bankų sektoriuje visą laiką fiksuotas mažiausias išlikimas.



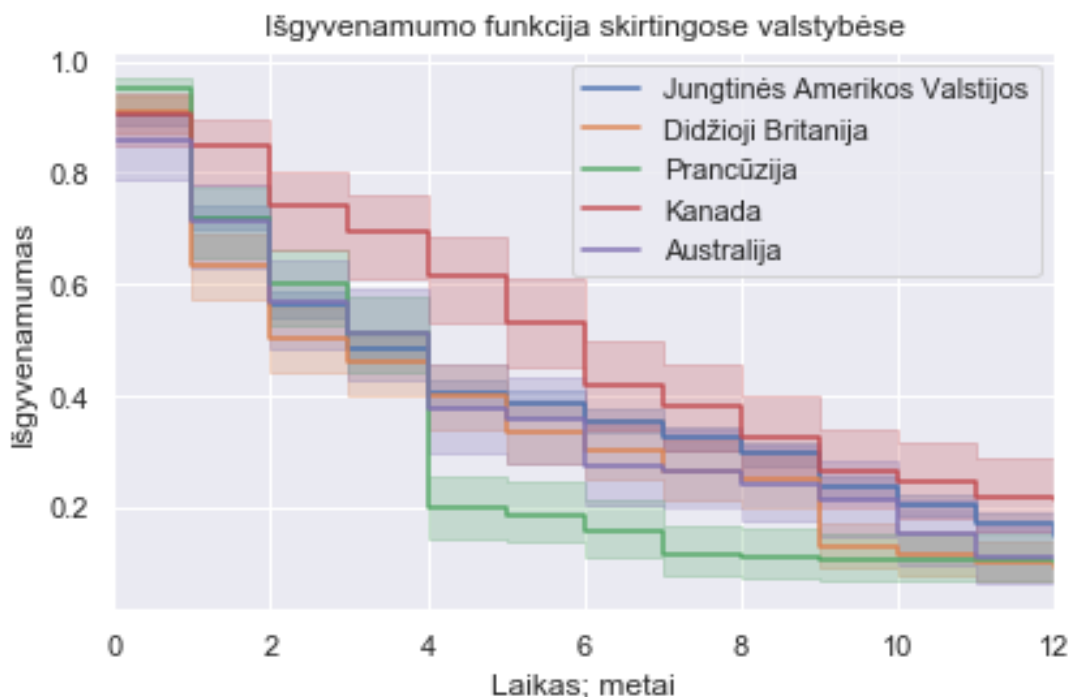
**1.3 pav.** Išlikimo funkcija skirtinguose sektoriuose

Nustačius dažniausiai besikeičiančius verslo sektorius, nuspręsta ištirti, apie kurių šalių įmones „Bloomberg“ pateikė daugiausiai informacijos ir kur anksčiausiai buvo fiksuojami kreditingumo reitingo kritimai. Lentelėje (žr. 1.6 lentelę) matome 5 šalis, apie kurių įmonių kreditingumo reitingų pokyčius yra daugiausiai informacijos. Lentelėje pirmą poziciją užima Jungtinės Amerikos Valstijos, kur tiriamos net 1802 įmonės, apie antroje pozicijoje esančios Didžiosios Britanijos įmonės turime daugiau nei 6 kartus mažiau informacijos – tiriami 271 įmonės kreditingumo reitingo pokyčiai.

**1.6 lentelė** Tiriamų įmonių kiekiai valstybėse

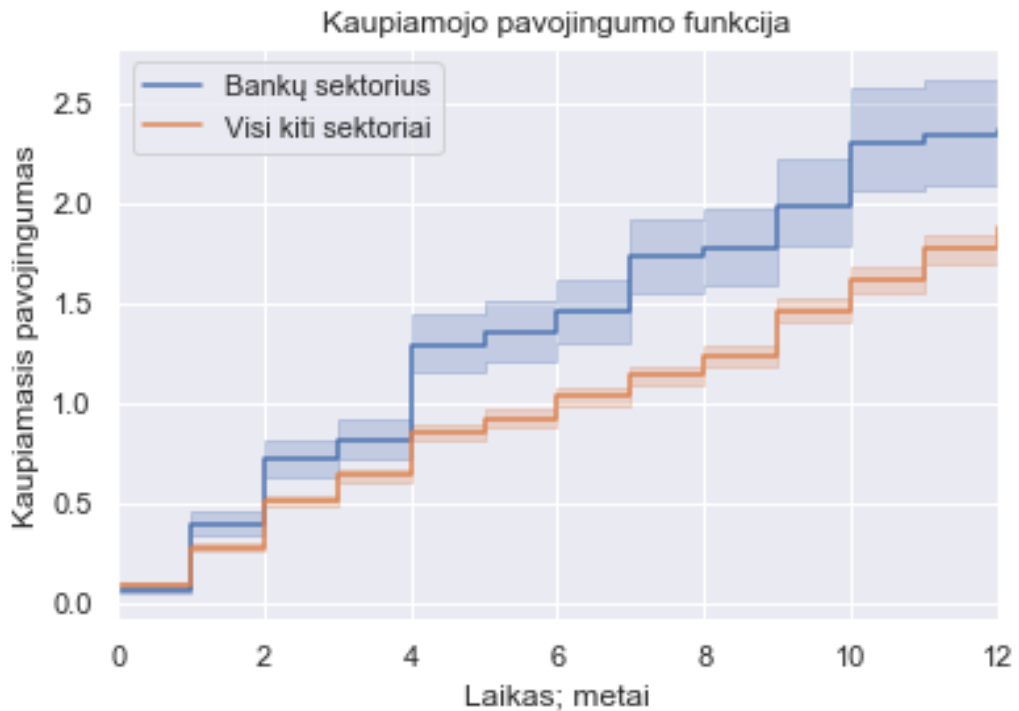
Šalis	Įmonių kiekis
Jungtinės Amerikos Valstijos	1802
Didžioji Britanija	271
Prancūzija	197
Kanada	150
Australija	146

Toliau išsiaiškinama, po kiek laiko pirmą kartą sumažėjo kreditingumo reitingų įverčiai skirtingose valstybėse. Pavyzdyje (žr. 1.4 pav.) vaizduojama išlikimo funkcija 5 dažniausiai duomenyse sutinkamose valstybėse. Akivaizdu, kad 1 – 4 metais Didžiosios Britanijos įmonės dažniausiai susiduria su kreditingumo reitingo kritimo problemomis. Tačiau būtent 4 metais net apytiksliai trečdaliui Prancūzijos įmonių fiksuojamas kreditingumo reitingo sumažėjimas. Visą tiriamą laikotarpį Kanados įmonėms nustatytas gerokai mažesnė kreditingumo reitingo sumažėjimo tikimybė, nei kitoms tiriamoms valstybėms.



**1.4 pav.** Išlikimo funkcija skirtingose valstybėse

Kadangi tolimesniame darbe analizuojami tik bankų sektoriaus kreditingumo reitingai, nuspręsta palyginti bankų ir kitų sektorių pavojo funkcijas. Kaupiamajam pavojo funkcijai palyginti nubrėžtos funkcijos (žr. 1.5 pav.) Tiek bankų, tiek ir kitų sektorių pavojo funkcijos rodo nuolatinį kreditingumo reitingo sumažėjimo pavojumą. Tačiau nuolatinis bankų sektoriaus pavojo funkcija yra aiškiai didesnė. Pavyzdyje matome, kad didžiausias skirtumas tarp funkcijų pavojo funkcijų matomas 4, 7, 10 ir 11 metais. 4, 7 ir 10 metais fiksuojamas didžiausias bankų sektoriaus kreditingumo reitingo kritimo pavojumas. Kitų sektorių kreditingumo reitingo kritimo pavojumas labiausiai padidėja 1, 2 ir 9 metais.



1.5 pav. Kaupiamojo pavojingumo funkcija bankų ir kituose sektoriuose

## 1.5. Apibendrinimas

Akivaizdu, kad kreditingumo reitingo nustatymas bei jo taikymai padarė didžiulę įtaką pinigų skolinimo procesui bei padėjo skolinančioms įstaigoms pasiekti gerokai tikslesnius rezultatus. Vertinimo procesas yra sudėtingas ir turėtų būti efektyviai įvertinamas keliais lygmenimis: kreditingumo vertinimo valdymo veikla, žmogiškaisiais ištekliais, IT sistemų išsivystymo verte bei patirtimi išduodant paskolas.

Finansinės institucijos, norėdamos maksimizuoti pelną bei pasiūlyti geresnes sąlygas, nei konkurentai, vis aktyviau kreditingumo vertinimui naudoja kompiuterinį modeliavimą, paremtą mašininio mokymo metodais. Modeliavimas yra sudarytas iš problemos apibrėžimo (nustatoma kokią riziką finansinė institucija gali prisiimti), tiriamosios analizės (vykdomas įrankių bei charakteristikų pasirinkimas) bei modelio vystymo žingsnių. Galima tikėtis, jog ateityje kompiuterinis modeliavimas ir prognozavimas bus taikomas vis plačiau, modeliai taps sudėtingesni, jie bus apjungiami tarpusavyje ir jų tikslumas tik didės. Svarbu neužmiršti, kad tinkamas prognozavimo modelio naudojimas neapsiriboja tik programinės įrangos darbu. Įmonės turi pasiekti geriausią įmanomą balansą tarp vartotojų poreikių, darbuotojų žinių bei naujų programinės įrangos galimybių.

Kreditingumo reitingai gali būti sudaromi remiantis skirtingomis metodikomis. Siekiant tiksliai išmatuoti rizikas, svarbu visus reitingus vertinti kritiškai, analizuoti jų pagrįstumą. Tolimesniame darbe pristatomi reitingų dinamikos tyrimo būdai taikant mašininio mokymo algoritmus. Taip pat analizuojami „Didžiojo trejeto“ agentūrų reitingų pokyčiai.

## 2. Taikomų metodologijų ir sprendimų apžvalga

Pabaigę analizuoti kreditingumo sąvoką, kreditingumo reitingo sudarymą, bei pagrindines problemas su kuriomis susiduria finansinės institucijos atlikdamos reitingavimą, pereiname prie industrijoje taikomų metodologijų ir sprendimų apžvalgos. Šiame skyriuje pateikiami dažnai naudojami kreditingumo reitingavimo metodai, jų privalumai ir trūkumai. Aprašomos mašininio mokymo algoritmų rezultatų vertinimo metodikos. Apžvelgiami mašininio mokymo metodų parametrų derinimo būdai.

### 2.1. „Bloomberg“ terminalas

„Bloomberg“ terminalas, tai įmonės „Bloomberg L.P.“ sukurta kompiuterinių programų sistema. Sistema suteikia vartotojams galimybę stebėti ir analizuoti finansinių rinkų duomenis realiu laiku. Terminalas pateikia finansines naujienas, istorinius pasaulinių įmonių finansų rodiklius, kreditingumo reitingus, leidžia prekiauti įmonių akcijomis internetinėje investavimo platformoje. Pirmoji „Bloomberg“ terminalo versija išleista 1982 metų gruodį. Įmonės, norinčios naudoti „Bloomberg“ terminalą turi įsigyti 2 metų trukmės licenciją. Šiuo metu pasaulyje yra apie 325 tūkstančiai aktyvių terminalo prenumeratorių. Lietuvoje „Bloomberg“ terminalus akademiniams tikslams turi Kauno technologijos universitetas bei ISM vadybos ir ekonomikos universitetas.

### 2.2. Klasifikavimo ir detekcijos uždaviniai

Klasifikavimas, tai viena dažniausiai sutinkamų užduočių priimant įvairiausius sprendimus. Klasifikavimo uždavinys sprendžiamas tada, kai norima, remiantis turimais objekto parametrais, jį priskirti vienai iš apibrėžtų klasių. Kadangi grupės, į kurias yra klasifikuojama yra žinomos iš anksto, klasifikavimo uždavinys priskiriamas mokymosi su mokytoju grupei. Klasifikavimo uždavinys dažnai naudojamas garsų atpažinime, simbolių atpažinime, medicininių ligų atpažinime, kreditingumo reitingavime ir kitose srityse. Klasifikacijos uždavinys, kai stebiniai yra priskiriami vienai iš 2 klasių, vadinamas *detekcija*.

Vienas svarbiausių sprendimų, kurių priima finansinės institucijos – nuspręsti ar suteikti paskolą jos prašančiam asmeniui ar įmonei. Dar ne taip seniai šis sprendimas buvo priimamas atsižvelgiant vien tik į kliento užpildytą prašymo formą bei ekspertų patarimus. Tačiau patobulėjus kompiuterinei infrastruktūrai: atsiradus galimybei saugoti didžiulius duomenų kiekius, juos sparčiai pasiekti, analizuoti, atlikti mašininį apmokymą bei gauti prognozes, kredito išdavimo procesas pasikeitė labai stipriai. Šiuolaikiniai prognozavimo modeliai paskolos prašytojams priskiria kreditingumo reitingus, kurie labai tiksliai įvertina prisiimamas rizikas. Jei šis reitingas sutampa su finansinės institucijos nustatyta riba arba viršija ją, paskola yra suteikiama, o jei reitingas yra žemesnis – nesuteikiama.

Klasifikavimo uždaviniams spręsti skirti mašininio mokymo modeliai yra labai aktualūs verslui ir todėl tobulėja didžiuliu greičiu. Beveik kasdien mokslinėje literatūroje ir straipsniuose skelbiami nauji mašininio mokymo metodų taikymai. Metodai nuolatos lyginami su senesniais, bandoma atrasti algoritmus, kurių pagalba pasiekiamas didžiausias prognozavimo tikslumas ar kitos modelio vertinimo charakteristikos.

#### 2.2.1. Klasifikavimo ir detekcijos uždaviniams spręsti skirti algoritmai

Darbe nuspręsta išanalizuoti ir pritaikyti 3 populiarius ir plačiai naudojamus detekcijos algoritmus: logistinę regresiją, atsitiktinius miškus bei atramos vektorių modelį. Tuo tarpu klasifikavimo



uždaviniui spręsti nuspręsta naudoti 2 iš jų: logistinę regresiją ir sudėtingesnę – atsitiktinių miškų algoritmą. Toliau yra pristatomi visi paminėti algoritmai.

### **Logistinė regresija**

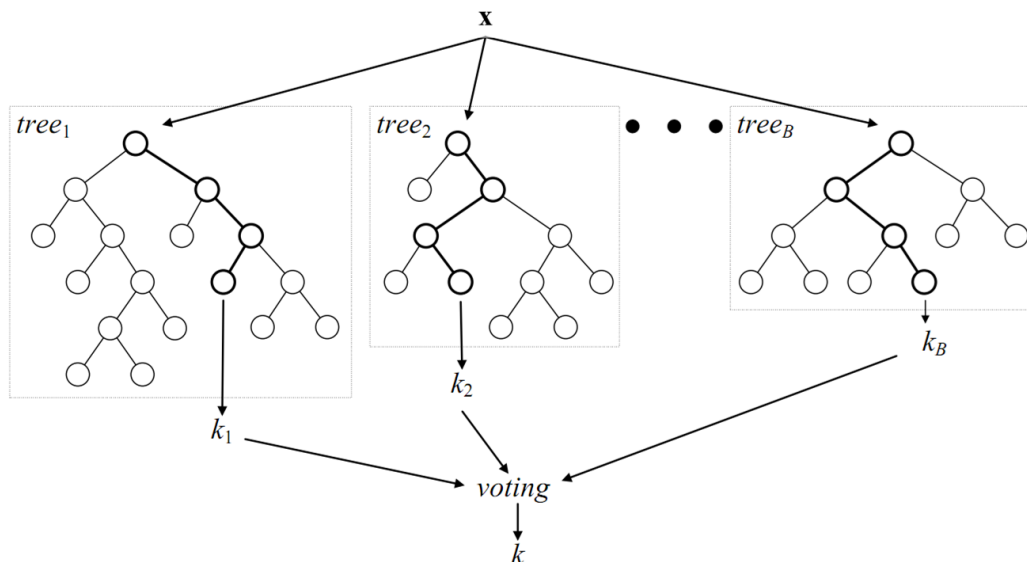
Logistinė regresija pradėta naudoti biologijos moksle XX amžiaus pradžioje. Istoriskai logistinė regresija yra laikoma dažniausiai industrijoje naudojamu algoritmu. Labai svarbus logistinės regresijos privalumas yra tas, kad jos prognozės rezultatas nėra diskretus. Atliekant logistinę regresiją, kiekvienam stebiniui yra gaunama tikimybė patekti į kažkurią klasę. Logistinė regresija gali būti naudojama, kai prognozuojama reikšmė yra kategorinė. Pavyzdžiui, ji gali būti taikoma, kai sprendžiama, ar klientas grąžins skolą laiku (žymima „1“), ar negrąžins laiku (žymima „0“). Atliekant prognozavimą logistinės regresijos būdu, turi būti nustatomas sprendimo slenkstis, pagal kurį reikšmės yra priskiriamos į vieną ar kitą klasę. Tarkime, jei regresijos modelio prognozuojama reikšmė yra didesnė arba lygi 0.5, laikoma, kad klientas padengs įsiskolinimus laiku. Jei prognozuojama reikšmė yra mažesnė už 0.5, sakoma, kad klientas vėluos grąžinti skolas. Sprendimo riba gali būti linijinė arba nelinijinė. Siekiant prisitaikyti prie verslo poreikių, gali būti parenkamos aukštesnės daugianario eilės, formuojama sudėtingesnė sprendimo priėmimo riba. Daugelis kitų algoritmų neturi taip lanksčiai taikomos sprendimų priėmimo ribos.

Taip pat logistinė regresija yra efektyvi: veikia greitai ir smarkiai neapkrauna kompiuterio atminties. Algoritmas yra atsparus nedideliems triukšmams bei nedideliam multikolinearumui. Esant dideliam multikolinearumui taikoma L2 tipo regularizacija. Logistinė regresija veikia prasčiau, kai duomenys yra sudaryti iš didelio kiekio savybių, arba kai duomenyse yra daug praleistų reikšmių. Dideli kiekiai kategorinių duomenų taip pat kenkia logistinės regresijos rezultatams.

### **Sprendimo medžiai ir atsitiktiniai miškai**

Sprendimo medis leidžia sukonstruoti medžio struktūros formos klasifikavimo modelį. Medis sudaromas logiškai išskaidant duomenų rinkinį į vis mažesnes dalis. Procesas atliekamas palaipsniui. Galutinis rezultatas – medis su sprendimo priėmimo šakomis ir prognozuojamomis klasėmis vietoj lapų. Sprendimų medžiai yra tinkama priemonė klasifikuoti kategorinius ir skaitinius duomenis. Sprendimo medžių sudaryti modeliai yra intuityvūs ir juos paprasta paaiškinti užsakovams. Didžiausias sprendimų medžio trūkumas – polinkis persimokyti.

Atsitiktiniai miškai (žr. 2.1 pav.) yra daugelio sprendimų medžio junginiai, kurie yra suformuojami iš kelių sprendimų medžių ir pateikia dažniausiai prognozuojamos klasės pavadinimą. Paveikslėlio viršuje matome algoritmui paduodamus duomenis  $x$ . Atsitiktinis miškas kiekvienam iš  $B$  medžių išveda rezultatą, žymimą  $k_i$  raide. Turint  $B$  medžių rezultatus atliekamas balsavimas: stebinyms priskiriamas tai klasei, už kurią balsavo daugiausiai atskirų sprendimo medžių. Atsitiktiniai miškai gali būti naudojami ir regresijoje ar kituose uždaviniuose. Metodas yra sudėtingas bei pasiekia labai tikslių rezultatų ir klasifikavime, ir regresijoje. Dažnas atsitiktinių miškų metodo trūkumas – persimokymas naudojant apsimokymo duomenų imtį. Taip pat, skirtingai nei sprendimo medžiai, atsitiktinių miškų modeliai nėra intuityvūs ir juos sunku suprasti, nes nebeužtenka interpretuoti vieno atskiro medžio. Taikant atsitiktinius miškus reikia skirti daug laiko apmokymui bei rezultatams gauti. Algoritmas naudoja daugiau kompiuterio resursų, nei taikant kitus metodus.

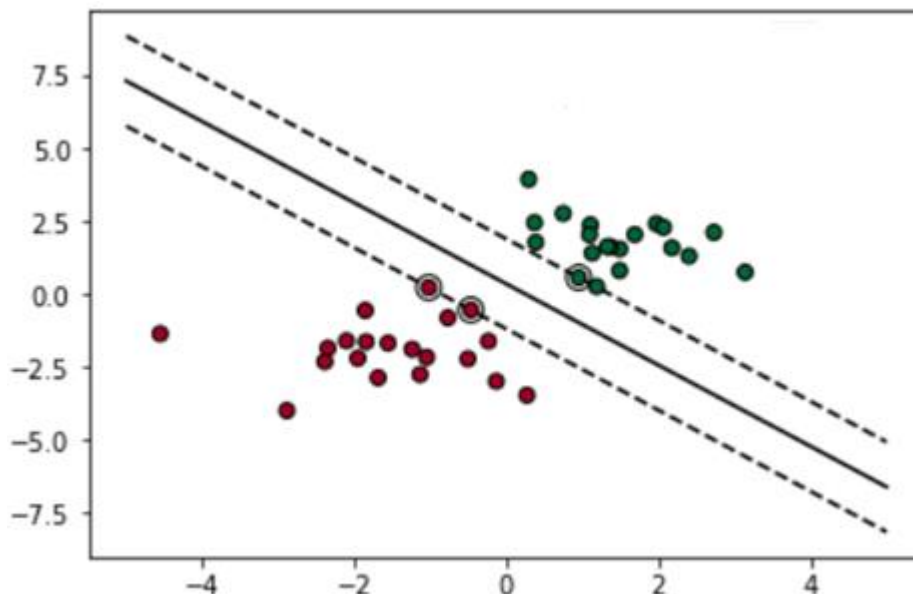


2.1 pav. Atsitiktiniai miškai

(Minelga, et al., 2014)

### SVM – atramos vektorių modelis

Atramos vektorių modelis (žr. 2.2 pav.) yra skirtas rasti sprendimo ribą, kuri atskirtų skirtingas klases ir parinktų maksimalią paraštę tarp jų. Maksimali paraštė randama nustatant kiekvieno taško statmeną atstumą iki sprendimo ribos. Šis mažiausias atstumas ir žymi paraštės ribą. Paveikslėlyje sprendimo riba žymima nepertraukiama linija, o paraštės iš abiejų pusių yra atskirtos punktyrinėmis linijomis. Raudona ir žalia spalva žymimi skirtingų klasių stebiniai. Apibraukti stebiniai, esantys arčiausiai sprendimo priėmimo ribos (pagal juos buvo nustatyta maksimali paraštė).

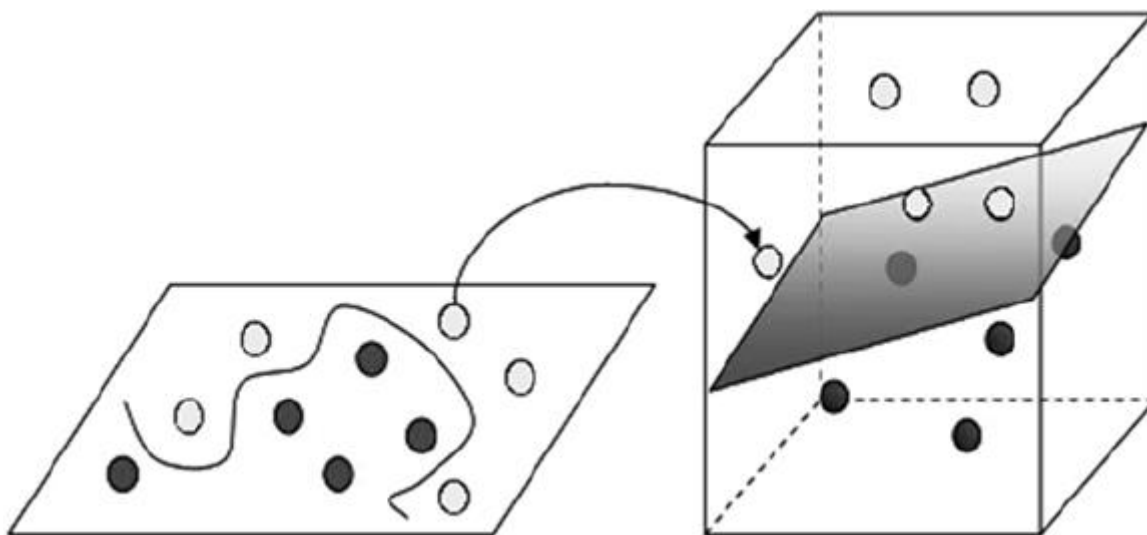


2.2 pav. SVM – atramos vektorių modelis

(Chen, 2019)

Jeigu klasės negali būti tiesiškai atskiriamos, klasifikavimui atlikti naudojamos „minkštos“ paraštės arba branduolio gudrybės metodikos. „Minkštos“ paraštės būdas leidžia daliai taškų patekti į kitos klasės pusę. Tuomet bandoma rasti balansą tarp paraštės dydžio ir neteisingai klasifikuojamų reikšmių kiekio. Paraštės minkštumas (neteisingai klasifikuojamų reikšmių kiekis bei svarba) yra reguliuojamas naudojant metodo parametą  $C$ . Esant didelei parametro  $C$  reikšmei klaidos yra netoleruojamos arba beveik netoleruojamos. Kai parametro  $C$  reikšmė yra nedidelė, klaidos tampa toleruojamos, paraštė siaurėja ir mažėja metodo naudojamų atramos vektorių kiekis.

Branduolio triuko metodika leidžia atskyrimui tarp klasių išnaudoti naują dimensiją. Pavyzdžiui, nesant optimalios atskyrimo ribos dvimatėje erdvėje, galima duomenis atvaizduoti trimatėje erdvėje, kur klasifikavimo uždavinį atlikti yra daug paprasčiau. Iliustracijoje (žr. 2.3 pav.) vaizduojama branduolio gudrybės pritaikymas. Kairėje pusėje matome erdvę, kurioje pažymėti dviejų klasių taškai (balti ir juodi), tačiau šių klasių neįmanoma tiesiškai atskirti. Kaip rodoma dešinėje paveikslėlio pusėje, pridėjus 3 dimensiją, klasės atskiriamos gerokai paprasčiau.



**2.3 pav.** Branduolio triukas atramos vektorių modelyje

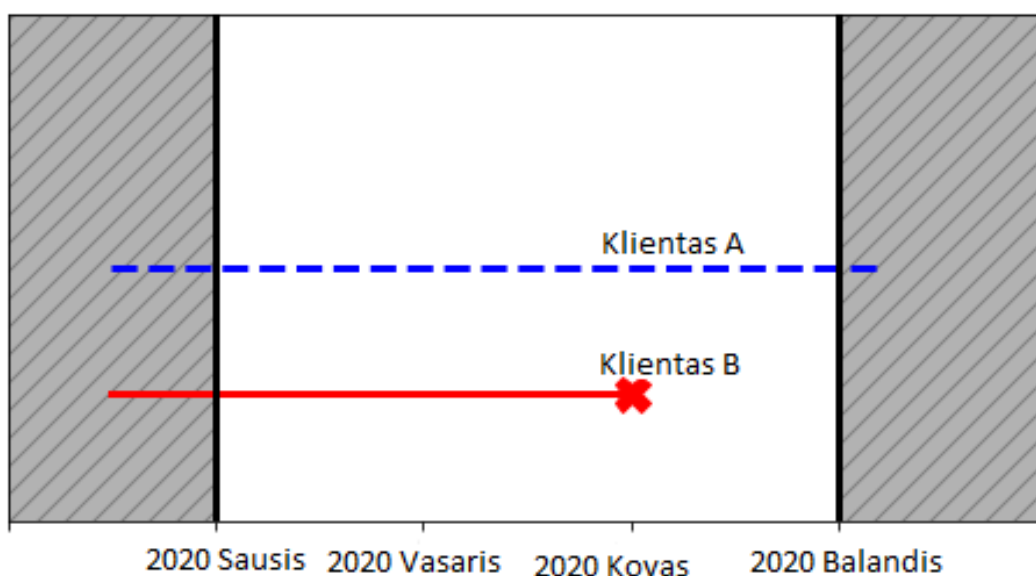
(Zararsiz & Cosgun, 2014)

### 2.3. Išlikimo analizė

Išlikimo analizės terminas ir uždavinys buvo suformuotas medicinos tyrėjų, siekiant nustatyti potencialią sergančių pacientų gyvenimo trukmę. Tačiau išlikimo analizė gali būti taikoma ne vien tik analizuojant pacientų gimstamumą ir mirštamumą, bet ir atliekant kitų procesų analizę laike. Pavyzdžiui:

- Bankai, skolinimo įstaigos ir kitos finansinės institucijos naudoja išlikimo analizę bandydami nustatyti paskolų gražinimo tempą bei laiką, kada skolininkai taps nemokiais.
- Verslai naudoja išlikimo analizę skaičiuodami visuminę kliento vertę bei atlikdami klientų išlaikymo analizę.
- Įmonės taiko išlikimo analizę, prognozuodamos, kiek laiko darbuotojai dirbs jų įmonėje.
- Inžinieriai, gamintojai prognozuoja kiek laiko tam tikras inžinerinis įrenginys veiks nepatirdamas gedimų.

Išlikimo analizė nuo kitų metodikų išsiskiria tuo, jog gauna prasmingą informaciją net ir tose situacijose, kai tiriamas įvykis neįvyksta analizuojamu laikotarpiu. Pavyzdžiui, 2.4 pav. matome 2 klientų bendradarbiavimo su įmone grafiką 2020 metų sausio – balandžio mėnesiais. Klientas A pradėjo bendradarbiauti su įmone anksčiau, nei 2020 metų sausis ir vis dar bendradarbiavo balandžio mėnesį. Klientas B bendradarbiavimą irgi pradėjo prieš užfiksuotą datą grafike, tačiau kovo mėnesį nustojo vykdyti bendrą veiklą. Situacija su klientu B yra aiški, tačiau neturime informacijos apie klientą A, išskyrus tai, kad jis išliko klientu per visą matomą laikotarpį. Kliento A situacija išlikimo analizėje vadinama *cenžūravimu*.



2.4 pav. Išlikimo analizė, pavyzdys

(Fotso, 2019)

Tiriamu atveju regresijos modeliai galėtų prognozuoti klientų bendradarbiavimo pabaigą. Tačiau regresijos algoritmai neįvertina *cenžūruotų* stebinių ir praranda daug naudingos informacijos. Išlikimo analizė įvertina ir *cenžūravimą*. Šio tipo analizėje, vietoje to, kad būtų prognozuojama kada klientas nustos bendradarbiauti su įmone, prognozuojama tikimybė, kad bendradarbiavimas baigsis tam tikru laiku.

Išlikimo analizė apibūdinama 3 vienodo ilgio objektais: X, E ir T. Prieš apmokant išlikimo analizės modelį, reikia nustatyti visus 3 elementus:

- X yra daugiamatis kintamųjų vektorius, susidedantis iš įvairiausių stebinio savybių.
- E yra įvykių indikacijos vektorius. Jeigu analizuojamu laikotarpiu įvykis įvyksta,  $E_i$  įgauna reikšmę „1“, o jei užfiksuojamas *cenžūravimas*,  $E_i$  prilyginamas „0“.
- T vektoriuje saugomas įvykio fiksavimo laikas. Dažnai žymima  $T = \min(t, c)$ , kur t yra įvykio laikas, o c – stebinio *cenžūravimo* laikas.

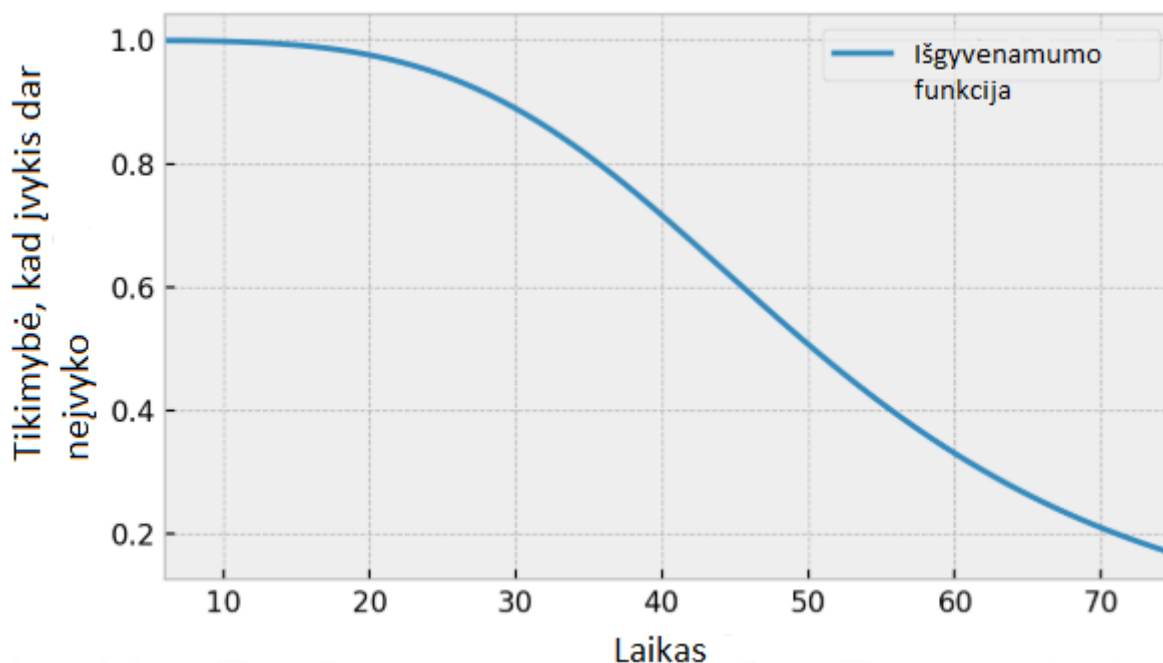
### Išlikimo funkcija

Tarkime, kad T – atsitiktinis dydis, aprašantis tiriamo populiacijos vieneto gyvenimo trukmę (dydis gali siekti ir begalybę, bet privalo būti neneigiamas). Pavyzdžiui T gali parodyti kiek laiko įmonės bendradarbiauja. Tuomet išgyvenimo funkcija  $S(t)$  yra apibrėžiama kaip:

$$S(t) = \Pr(T > t)$$

Taigi, išlikimo funkcija apibrėžia tikimybę, jog pabaigos įvykis vis dar neįvyko laiku  $t$ , arba išlikimo tikimybę praėjus laikui  $t$ . Išlikimo funkcijos (žr. 2.5 pav.) grafike matome tipinį pavyzdį – nuliniu laiko momentu išlikimo tikimybė lygi 1, vėliau ji visą laiką mažėja. Išlikimo funkcija turi šias savybes:

1.  $0 \leq S(t) \leq 1$
2.  $F_T(t) = 1 - S(t)$ , čia  $F_T(t)$  yra sukauptoji pasiskirstymo funkcija (CDF)
3.  $S(t)$  yra nedidėjanti  $t$  funkcija



2.5 pav. Išlikimo funkcija

(Davidson-Pilon, et al., 2020)

### Pavojingumo funkcija

Pavojingumo funkcija, žymima  $h(t)$ , aprašo sąlyginę tikimybę, kad įvykis įvyks per  $[t, t + dt]$  laiko, su sąlyga, kad jis dar neįvyko. Pavojingumo funkcija laiku  $t$  aprašoma formule:

$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr[t \leq T < t + dt | T \geq t]}{dt} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$$

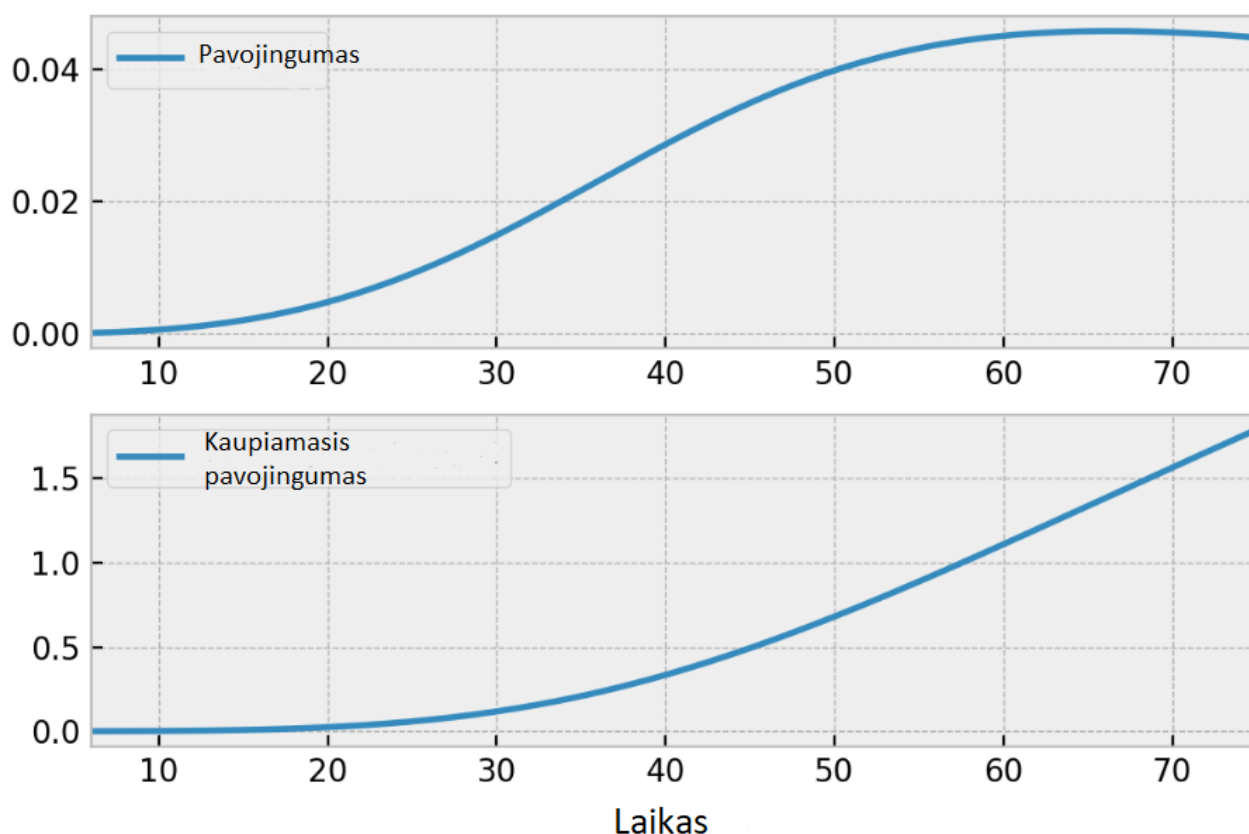
Čia  $f(t)$  – tikimybės tankio funkcija, o  $S(t)$  – išlikimo funkcija. Pavojingumo ir išlikimo funkcijos yra susietos šia lygybe:

$$S(t) = \exp\left(-\int_0^t h(u) du\right)$$

Kaupiamoji pavojingumo funkcija yra aprašoma lygybe:

$$H(t) = \int_0^t h(u)du$$

Paveikslėlyje (žr. 2.6 pav.) vaizduojami nagrinėjamos išlikimo funkcijos pavojingumo bei kaupiamąjo pavojingumo funkcijų grafikai. Matome, jog pavojingumo funkcijos įvertis didėja iki 60 laiko vieneto, o vėliau išlieka pastovus. Todėl, kadangi pavojingumas priklauso ir nuo ankstesnių įvykių, galima daryti išvadą, jog analizuojamas įvykis dažniausiai įvyksta iki 60 laiko vieneto. Tuo tarpu kaupiamąjo pavojingumo funkcija yra nuolat didėjanti (joje sumuojamas visais laiko momentais fiksuotas pavojingumas).



2.6 pav. Pavojingumo ir kaupiamąjo pavojingumo funkcijos

(Davidson-Pilon, et al., 2020)

### 2.3.1. Žvalgomoji išlikimo analizė

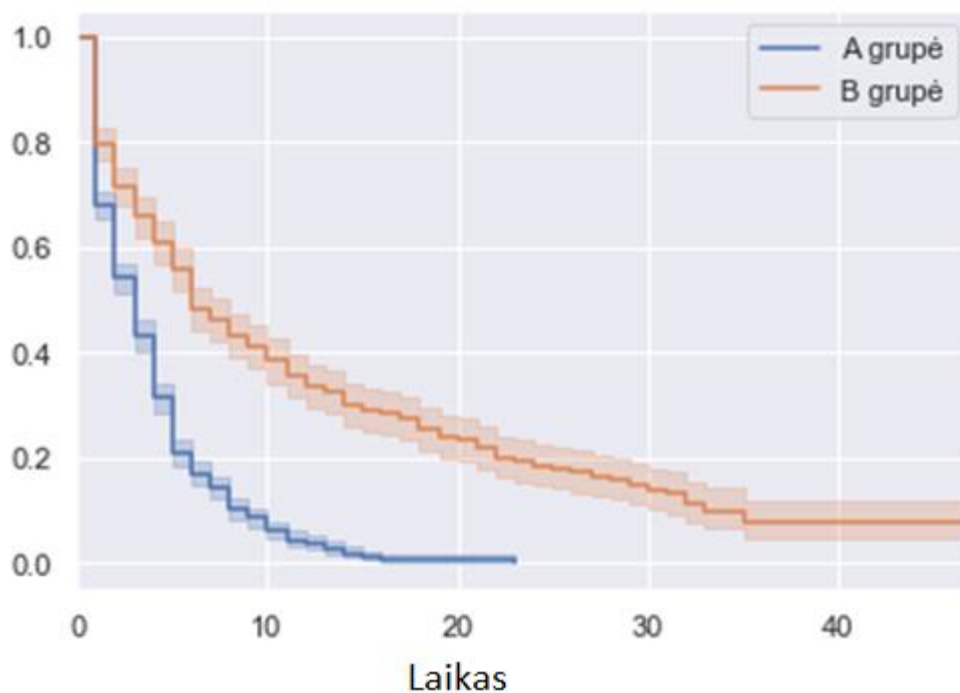
Žvalgomosios išlikimo analizės tikslas – įvertinti išlikimo funkciją. Tam pasitelkiamas neparametrinis *Kaplan-Meier* statistinis vertinimas (dar žinomas kaip produkto ribos vertinimas), kuris yra laikomas paprasčiausiu būdu išmatuoti išlikimą laike. Kaplan-Meier statistinis vertinimas pristatytas 1958 metais. Šį metodą suformavo 2 mokslininkai: Edvardas L. Kaplanas ir Paulas Mėjeris ir publikavo mokslo žurnale: „*Journal of the American Statistical Association*“. *Kaplan-Meier* statistinis vertinimas gali būti atliekamas patvirtinus 3 prielaidas:

- *Cenzūruojami* stebiniai turi tokias pačias tiriamo įvykio įvykimo tikimybes, kaip ir visi kiti stebiniai.
- Skirtingu laiku pradėti analizuoti subjektai turi vienodas išlikimo galimybes.

- Tiriamas įvykis įvyksta tiksliai aprašytu laiku.

Statistinis vertinimas atliekamas tiriant *Kaplan-Meier* kreivę, kuri parodo tikimybę išgyventi tam tikrą laiką. Šis laikas yra suskirstytas į daugelį mažų intervalų, kurių įverčiai pažingsniui mažėja. Turint pakankamai didelę duomenų imtį, kreivės įverčiai artėja link tiriamos populiacijos išlikimo funkcijos. *Kaplan-Meier* statistinis vertinimas dažnai naudojamas paskirstant stebinius į grupes ir lyginant stebinių grupių išlikimą (Goel, Khanna, & Kishore, 2010).

*Kaplan-Meier* statistiniam vertinimui atlikti reikalinga kiekvieno stebinio 2 tipų informacija: įvykio buvimas / nebuvimas ir laikas iki įvykio atsitikimo arba *cenžūravimo*. Jeigu atliekamas 2 grupių išlikimo lyginimas, reikalingi ir kiekvieno stebinio priklausomybės vienai ar kitai grupei duomenys. Pavyzdyje (žr. 2.7 pav.) matome, jog 0 laiko momentu abiejų grupių išlikimas yra 1 (neužfiksuota nei vieno tiriamo įvykio), laikui bėgant jis mažėja. A grupės išlikimas mažėja daug greičiau, nei B grupės (16 laiko momentu A grupės išlikimas pasiekia 0, tuo metu maždaug trečdalis B grupės stebinių vis dar nėra susidūrę su tiriamu įvykiu).



2.7 pav. 2 skirtingų grupių išlikimas

(Davidson-Pilon, et al., 2020)

### 2.3.2. Regresiniai išlikimo modeliai

Dažnai analizuojami duomenys susideda ne tik iš išlikimo laiko, bet ir iš kitų galimai analizei naudingų parametrų, pavyzdžiui asmens amžiaus, šalies ar lyties ir kitų parametrų. Kaip ir *Kaplan-Meier* atveju, negalime naudoti tradicinių prognozavimo metodų, tokių kaip linijinė regresija, nes prarandame *cenžūravimo* informaciją. Tokiais atvejais dažnai naudojami Cox proporcingo pavojaus bei išlikimo miškų modeliai. Modeliai taikomi analizuojant pavojaus lygį  $h(t|x)$ , funkcijoje  $t$  rodo laiką, ir  $x$  žymimi kiti tyrime naudojami rodikliai.

#### Cox proporcingo pavojaus modelis (angl. Cox's Proportional Hazard Model)

Cox proporcingo pavojaus modelis yra parentas išlikimo duomenų analizės modeliavimu. Ši statistinė metodika skirta tirti stebinio išlikimo ryšį su įvairiais stebinio parametrais bei analizuoti kaip sąveika kinta bėgant laikui. Modelis yra plačiai taikomas su sveikatos sritimi susijusiose prognozėse. Svarbus modelio privalumas – galimybė atskirti išorinio įsikišimo poveikį klientų išlikimui nuo kitų naudojamų kintamųjų poveikio. Cox algoritmo modeliavimo rezultatas – pavojingumo funkcija, vertinanti kelis stebinio parametrus. Cox modelis yra interpretuojamas analizuojant apskaičiuotus kintamųjų koeficientus. Jei regresinis koeficientas yra teigiamas tam tikram kintamajam, tai reiškia, jog pavojingumas yra didesnis ir egzistuoja didesnė tikimybė įvykti analizuojamam įvykiui. Priešingu atveju, neigiamas regresijos koeficientas apibūdina mažesnę pavojingumą ir didesnes išlikimo galimybes.

Pusiau parametrinis Cox modelis prognozuoja pavojingumo funkciją  $h(t, x_i)$ , turint omenyje, jog laiko komponentė  $\lambda_0(t)$  ir kintamųjų komponentė  $n(\vec{x}_i)$  yra proporcingos:

$$h(t, \vec{x}_i) = h_0(t) * n(\vec{x}_i)$$

- $h_0(t)$  yra pradinė pavojingumo funkcija (angl. baseline function), kuri įprastai yra neaprašyta.
- $n(\vec{x}_i)$  – rizikos funkcija, įprastai vaizduojama linijiniu būdu. Modelio funkcijos koeficientai nustatomi pagal:  $n(\vec{x}_i) = \exp(\sum_{j=1}^p x_j^i w_j)$ , čia  $w_j$  – modelio koeficientai.

(Fotso, 2019)

### **Išlikimo miškai (angl. Survival Forest)**

Sprendimo medžius taikantys ansamblių modeliai (angl. ensemble) gali būti naudojami dirbti ir su cenzūrojamais duomenų rinkiniais. Tokio tipo modeliai apibendrintai vadinami išlikimo miškų modeliais. Toliau pateikiamas struktūrizuotas išlikimo miškų taikymo algoritmas:

1. Iš pradinio duomenų rinkinio  $B$  kartų atsitiktinai parenkami stebiniai su pasikartojimu. Svarbu prisiminti, kad kiekvieną kartą vidutiniškai 37% stebinių yra neparenkama, literatūroje šie neparenkami stebiniai žymimi *OOB* (angl. out of bag).
2. Kiekvienam iš  $B$  duomenų rinkinių sudaromas išlikimo medis. Kiekvienam medžio poaibiui atsitiktinai parenkama  $p$  parametru – kandidatų. Atšaka yra dalinama pagal tą parametru, kuris generuoja didžiausią skirtumą tarp žemiau esančių poaibių.
3. Medis auginamas iki visiško išskaidymo, arba tol, kol yra išpildomas stabdymo kriterijus (pavyzdžiui žemiausias poaibis turi ne mažiau kaip 1 fiksuojamą įvykį.)
4. Kiekvienam medžiui nustatoma kaupiamojo pavojingumo funkcija, literatūroje žymima *CHF*. Išlikimo miško kaupiamojo pavojingumo funkcija nustatoma apskaičiuojant visų medžių funkcijų vidurkį.
5. Išlikimo miško prognozavimo kokybė įvertinama panaudojant *OOB* duomenis, nustatoma prognozavimo paklaida.

(Ishwaran, Kogalur, Blackstone, & Lauer, 2008)

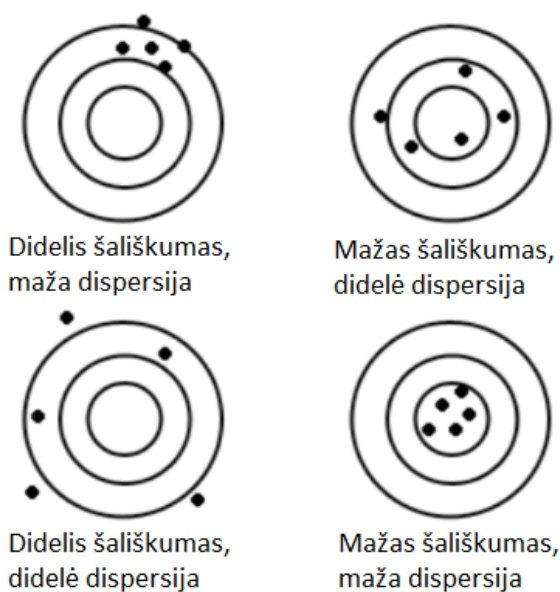


## 2.4. Geriausio modelio pasirinkimas

Modelio pasirinkimo proceso metu siekiama rasti tinkamiausią mašininio mokymo metodą tam tikrai problemai spręsti. Modelio pasirinkimo metu modeliai kandidatai yra lyginami tarpusavyje pagal skirtingus įverčius. Siekiant kokybiškai išrinkti geriausią modelį svarbu naudoti žemiau apibūdintas atrankos ir tikimybinės metodikas:

### Šališkumo ir dispersijos kompromisas

Praktikoje naudojama daug įvairių metodologijų vertinant mašininio mokymo modelio prognozavimo kokybę, tačiau visada svarbu išlaikyti žemą šališkumą ir dispersiją. Šališkumas, tai sisteminis skirtumas tarp modelio prognozuojamos ir tikrosios vertės. Dispersija – atsitiktinė paklaida. Paveikslėlyje (žr. 2.8 pav.) matome visas 4 galimas šališkumo ir dispersijos metriku kombinacijas. Esant dideliam šališkumui ir mažai dispersijai (pavyzdys kairėje pusėje viršuje), prognozių vidurkis yra toli nuo optimalaus, bet tarpusavyje prognozės skiriasi nedaug. Esant mažam šališkumui ir didelei dispersijai, (pavyzdys dešinėje pusėje viršuje) prognozių vidurkis yra optimalus, bet jos smarkiai skiriasi viena nuo kitos. Esant dideliam šališkumui ir didelei dispersijai (pavyzdys kairėje pusėje apačioje) vidutinė prognozė yra toli nuo optimalios ir atskiros prognozės skiriasi smarkiai. Paskutiniu atveju (pavyzdys dešinėje pusėje apačioje), esant mažam šališkumui ir mažai dispersijai vidutinė prognozė yra arti optimalios ir atskirų prognozių reikšmės skiriasi nežymiai.



2.8 pav. Šališkumo ir dispersijos kompromisas

Jei mašininio mokymo modelis yra permokomas, arba tiriamiems duomenims sudaromas per daug sudėtingas modelis, jis įsimesna duomenų struktūrą ir triukšmą. Tokiu atveju, modelio prognozėse fiksuojama didelė dispersija ir tiriant naujus duomenis jo prognozės bus nepatikimos. Priešingu atveju, jei modelis nėra pasiekęs optimalaus apsimokymo lygio, jis atliks netikslias prognozes tiek su apmokymo, tiek ir su testavimo duomenimis. Šiuo atveju fiksuojamas didelis modelio šališkumas.

Mašininio mokymo modeliai pasiekia tiksliausius rezultatus esant nedideliam šališkumui ir mažai dispersijai. Tačiau bandant sumažinti šališkumą didėja dispersija ir bandant sumažinti dispersiją,

didėja šališkumas. Siekiant geriausios prognozavimo kokybės reikia rasti balansą – šališkumo ir dispersijos kompromisą.

### **Stratifikuotas k-dalių kryžminis patikrinimas**

Mašininio mokymo modelio pajėgumai negali būti vertinami panaudojant tuos pačius duomenis, su kuriais modelis buvo sudarytas. Taip yra todėl, nes svarbu patikrinti kaip modelis gebės dirbti su naujais duomenimis, prie kurių gali būti neprisitaikęs. Todėl svarbu pradinius duomenis padalinti į 3 dalis: apsimokymo, patikrinimo bei testavimo imtis. Apsimokymo imtis naudojama modeliui sudaryti bei apmokyti. Patikrinimo imtis naudojama modelio parametrų derinimui ir parinkimui. Testavimo imtis naudojama galutiniam modelio įvertinimui.

Šis sprendimas gali būti efektyviai patobulinamas pritaikant stratifikuotą kryžminį patikrinimą. Duomenys yra išmaišomi ir padalinami į k dalių (kuriose išlaikomas klasių santykis). Išrenkama viena dalis, kuri paliekama testavimui, likę duomenys naudojami modelio sudarymui. Apskaičiuojamas modelio rezultatas. Procesas kartojamas k kartų ir gaunama vidutinė modelio rezultato reikšmė.

Toks sprendimo būdas tinka testuojant konkretų pasirinktą modelį. Norint palyginti kelis modelius ir išrinkti geriausią naudojamas 2 lygių kryžminis patikrinimas. Tokiu atveju, atskyrus testavimo duomenų imtį, apmokymo duomenys yra papildomai padalinami į vidinę apmokymo ir patikrinimo imtis. Kiekvienoje iteracijoje atliekamas papildomas k-1 kryžminis patikrinimas parenkant geriausius parametrus ir tik tada testuojami modelio rezultatai.

Pagrindinis stratifikuoto k - dalių kryžminio patikrinimo naudojimo trūkumas – sumažėjusi tyrimo sparta. Reikia k kartų daugiau laiko įvertinti modeliui, o naudojant 2 lygių kryžminį patikrinimą:  $k*(k-1)$  karto daugiau laiko. Kitas trūkumas – vertinamas tik modelio prognozavimo rezultatas, bet neįvertinamas algoritmo sudėtingumas, nenustatoma kaip sunku bus sudarytą modelį interpretuoti.

### **Sumaišymo matrica**

2.9 pav. matoma sumaišymo matrica. Ji skirta nustatyti mašininio mokymosi klasifikavimo modelių tinkamumą. Matrica vaizduojama lentele, turinčia 4 prognozuojamų ir esamų reikšmių kombinacijas.

		Esamos reikšmės	
		Teigiama (1)	Neigiama (0)
Prognozuojamos reikšmės	Teigiama (1)	TP	FP
	Neigiama (0)	FN	TN

2.9 pav. Sumaišymo matrica

**TP – Tikras teigiamas (angl. True Positive)** - parodo kaip dažnai buvo prognozuojama teigiama baigtis ir gaunama teigiama baigtis.

**TN – Tikras neigiamas (angl. True Negative)** - parodo kaip dažnai prognozuojama neigiama baigtis ir gaunama neigiama baigtis.

**FP – Neteisingas teigiamas (False Positive)** - kaip dažnai prognozuojama teigiama baigtis, bet gaunama neigiama baigtis (1 rūšies klaida).

**FN – Neteisingas neigiamas (False Negative)** – kaip dažnai prognozuojama neigiama baigtis, bet gaunama teigiama baigtis (2 rūšies klaida).

Sumaišymo matricos įverčiai naudojami ne tik nustatant 1 ir 2 rūšies klaidas, bet ir apskaičiuojant kitus svarbius įverčius: *prisiminimą*, *preciziškumą*, *specifiškumą*, *tikslumą*. Naudojant sumaišymo matricos įverčius yra braižoma modelių lyginime naudojama *AUC-ROC* kreivė.

**Prisiminimas (angl. Recall)** – parodo kokia dalis teigiamų baigčių buvo klasifikuota teisingai. Literatūroje dažniausiai žymimas „Recall“:

$$\text{Prisiminimas} = \frac{TP}{TP + FN}$$

**Preciziškumas (angl. Precision)** – parodo kiek teigiamai prognozuotų baigčių iš tiesų buvo teigiamos. Preciziškumo metrika labai naudinga, kai svarbu nepadaryti 1 rūšies klaidų. Pavyzdžiui, modeliui atliekant nepageidaujamų elektroninių laiškų atpažinimą, svarbu gero laiško nepalaikyti šlamštu. Jei modelio preciziškumas nėra pakankamai aukštas, vartotojas gali netekti svarbių elektroninių laiškų. Literatūroje preciziškumas dažniausiai žymimas „Precision“:

$$\text{Preciziškumas} = \frac{TP}{TP + FP}$$

**Specifiškumas (angl. Specificity)** – parodo kokia dalis neigiamų baigčių buvo klasifikuojama teisingai. Literatūroje dažniausiai žymimas „Specificity“:

$$\text{Specifiškumas} = \frac{TN}{TN + FP}$$

**Tikslumas (angl. Accuracy)** – įvertis, kuris parodo kokia visų prognozių dalis buvo teisinga. Literatūroje dažniausiai žymimas „Accuracy“:

$$\text{Tikslumas} = \frac{TP + TN}{TP + TN + FP + FN}$$

**Subalansuotas tikslumas (angl. Balanced accuracy score)** – įvertis, naudojamas nustatant nesubalansuotų duomenų rinkinių tikslumą. Apskaičiuojamas sudarant vidutinį kiekvienos klasifikavimo klasės prisiminimą (angl. recall).

**F1 įvertis (angl. F1 score)** – įvertis naudojamas, kai tyrime norima atsižvelgti tiek į prisiminimo matą, tiek į preciziškumą. Literatūroje dažniausiai žymimas „F1“:

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

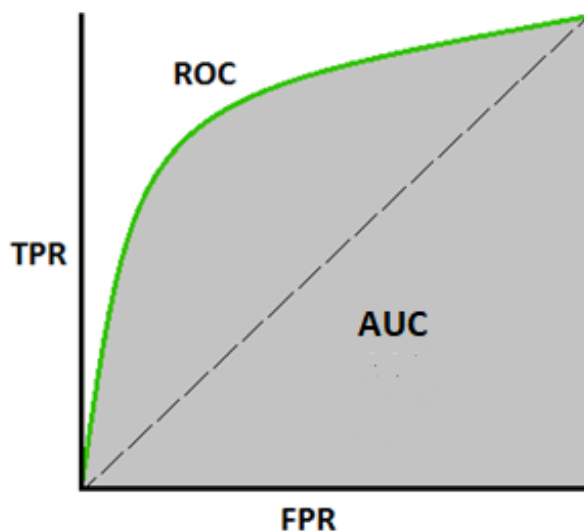
**Kappa įvertis (angl. Cohen`s Kappa)** – įvertis, vertinantis dviejų klasifikatorių sutarimą, klasifikuojant  $N$  elementų į  $C$  klasių. Taikant *kappa* įvertį nustatomas skirtingų klasifikatorių patikimumas, atsižvelgiant į tai, jog kartais jų rezultatai gali sutapti ir per visišką atsitiktinumą. *Kappa* įvertis gali svyruoti tarp 0 ir 1, bet kartais jis gali būti ir mažesnis už 0. Jei *kappa* įvertis yra lygus 0, tarp klasifikatorių yra tik atsitiktinis sutapimas. Jei įvertis lygus 1 – sutarimas yra visiškas. Esant mažesniai nei 0 *kappa* vertinimui, sutarimas yra net mažesnis už atsitiktinį. *Kappa* apskaičiuojamas pagal formulę:

$$Kappa = \frac{P_0 - P_e}{1 - P_e}$$

Formulėje  $P_0$  žymi tikimybę, jog klasifikatorių prognozės sutampa visiškai,  $P_e$  – prognozės sutampa per atsitiktinumą.

**Logloss įvertis** – klasifikavimo uždaviniuose naudojama nuostolių funkcija, apskaičiuojama taikant tikimybes. *Logloss* įvertis kokybiškai įvertina klasifikatoriaus tikslumą, bausdamas už kiekvieną neteisingai klasifikuojamą stebinį. Mažesnė *logloss* reikšmė rodo mažesnę modelio neužtikrintumą, todėl vertinamas modelis atliks tikslesnes prognozes. *Logloss* įvertis laikomas pranašesniu už kitus modelių vertinimo matus, nes jis ne tik nustato kaip teisingai ar neteisingai veikia modelis, bet įvertina ir kaip smarkiai tikimybinė modelio prognozavimo reikšmė skyrėsi nuo realios.

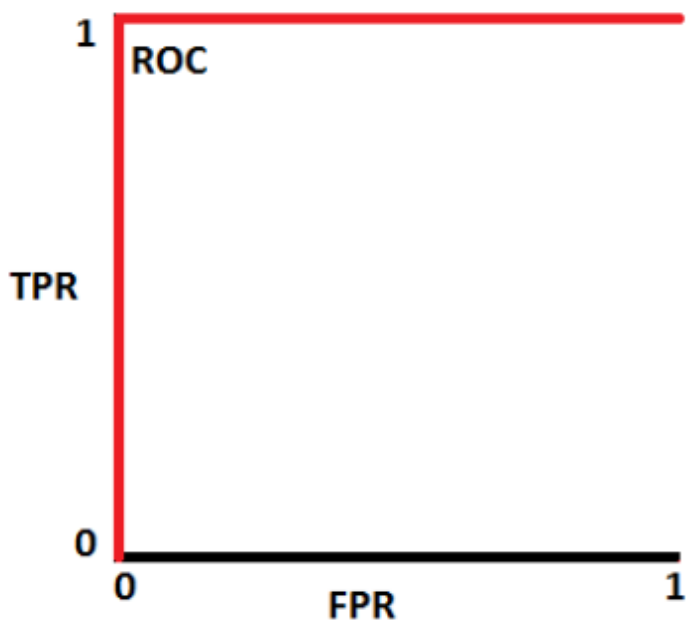
**AUC (angl. Area Under The Curve) – ROC (angl. Receiver Operating Characteristics)** – viena iš svarbiausių klasifikavimo modelių vertinimo metrikų, matuojama plotu, esančiu po ROC kreive. Pavyzdžiui (žr. 2.10 pav.) žalia ROC kreivė parodo tikimybinį modelio rezultatų pasiskirstymą. Pilkai nuspalvintas plotas AUC vaizduoja atskyrimo lygį, kuris parodo kaip tiksliai modelis geba atskirti tarp kelių skirtingų klasių. Kuo didesnė ROC AUC reikšmė (plotas), tuo tiksliau modelis atlieka klasifikavimą. TPR (angl. True Positive Rate) sutampa su prisiminimo (angl. Recall) įverčiu, o FPR (angl. False Positive Rate) įvertis yra lygus 1 - specifiškumas.



**2.10 pav.** ROC kreivė

(Narkhede, 2018)

Idealaus modelio ROC (vaizduojama raudonai 2.11 pav.) kreivė fiksuojama, kai  $TPR = 1$  ir  $FPR = 0$ , o paties blogiausio modelio atvirkščiai:  $TPR = 0$  ir  $FPR = 1$  (toks modelis atlieka atvirkštines prognozes). Jei modelio ROC kreivė brėžiama įstrižai ( $AUC = 0.5$ ), tai reiškia, kad modelis visiškai nesugeba atskirti skirtingų klasių reikšmių ir spėlioja. ROC kreivės gali būti naudojamos ir daugiau nei dviejų klasių modeliams. Pavyzdžiui, jei klasifikavimas atliekamas į 3 klases: X, Y, Z, braižomos 3 ROC kreivės. Pirmą rodytų kaip tiksliai X klasė atskiriama nuo Y ir Z. Antra – kaip tiksliai Y klasė atskiriama nuo X ir Z. Trečia – kaip tiksliai Z klasė atskiriama nuo Y ir Z. (Narkhede, 2018)



2.11 pav. Idealaus modelio ROC kreivė

(Narkhede, 2018)

**Preciziškumo – Prisiminimo (angl. Precision – Recall) kreivė** – kreivė, kurios x – ašyje vaizduojamas prisiminimas (angl. recall) ir y – ašyje vaizduojamas preciziškumas (angl. precision). Kreivėje visiškai spėjimas gali būti atvaizduojamas skirtingai, priklausomai nuo to, kaip yra pasiskirsčiusios teigiama bei neigiama klasės. Spėjimas atvaizduojamas horizontalia linija, parodančia klasių santykį. Pavyzdžiui subalansuoto duomenų rinkinio atsitiktinis spėjimas būtų vaizduojamas kaip horizontali linija, kurios preciziškumas lygus 0.5. Svarbus preciziškumo – prisiminimo kreivės akcentas yra tas, jog ji visiškai ignoruoja tikrų neigiamų reikšmių kiekį prognozėje. Todėl dažnai kreivė laikoma efektyvia vertinant duomenų rinkinius su dideliu klasių disbalansu.

**DET (angl. detection error trade-off) kreivė** – naudojama detekcijos uždavinių rezultatų vertinimui. Kreivės x – ašyje vaizduojamas neteisingai teigiamai klasei priskirtų prognozių santykis (angl. false positive rate), o y – ašyje vaizduojamas neteisingai neigiamai klasei priskirtų prognozių santykis (angl. false negative rate). Šioje kreivėje x ir y ašims taikoma logaritminė transformacija. Mažesnis kreivės apimamas plotas rodo geresnį klasifikatoriaus efektyvumą.

**Brier balas** – naudojamas įvertinti prognozuojamą išlikimo funkcijos tikslumą laiko momentu  $t$ . Brier balas gali būti naudojamas tik su tais įvykiais, kurie turi 2 galimas baigtis. Pavyzdžiui kreditingumo reitingas per ateinantį mėnesį mažes arba ne. Įvertis gali būti naudojamas ir su kategoriniais duomenimis, jei jie gali būti perstruktūrizuojami į binarines baigtis.

- Geriausia įmanoma Brier balo reikšmė yra 0. Tokiu atveju prognozės vykdomos be klaidų.
- Prasčiausia Brier balo reikšmė yra 1, kai prognozė vykdoma visiškai neteisingai.
- Vidutiniškus (0.3-0.7) Brier balo rezultatus sudėtinga interpretuoti. Todėl tokiu atveju naudojamas *Brier įgūdžių balas*.

Brier balas skaičiuojamas naudojant formulę:

$$BB = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

Čia:

- N - elementų, kuriems skaičiuojamas Brier balas kiekis.
- $f_t$  - prognozuojama tikimybė (pvz. 1 iš 3)
- $o_t$  – įvykio baigtis (1 jei įvykis įvyko, 0 jeigu ne)

(Hernandez-Orallo, Flach, & Ferri, 2012)

## 2.5. Mašininio mokymo modelių parametrų derinimas

Tas pats mašininio mokymo modelis gali pasiekti skirtingus prognozavimo rezultatus, parinkus skirtingus modelio parametrus. Sprendžiant vis kitą mašininio mokymo uždavinį ar turint kitokius duomenis, reikia naujai parinkti optimalius modelio parametrus. Kiekvienam mašininio mokymo modeliui derinami vis kiti parametrai: pavyzdžiui atsitiktiniams miškams dažnai derinamas medžių miške skaičius (paprastai didesnis medžių skaičius lemia ir geresnį algoritmo darbą, bet pridėjus per daug medžių algoritmas labai sulėtėja, todėl reikia rasti aukso vidurį). Taip pat gali būti derinami medžių gyliai, stebinių skaičius, kurį reikia pasiekti prieš skeliant atšaką bei minimalus stebinių skaičius medžio lapuose.

Mašininio mokymo modelių parametrai gali būti derinami 4 pagrindinėmis metodologijomis, kurios buvo apžvelgtos tolimesniuose skyreliuose.

### 2.5.1. Rankinis parametrų derinimas

Iš pradžių modeliui priskiriami pradiniai parametrai ir apskaičiuojami prognozavimo rezultatai. Pagal modelio gaunamas reikšmes pakeičiami naudojami parametrai, modelis permokomas. Stebima, ar modelio rezultatai pagerėjo. Procesas kartojamas be automatizavimo tol, kol pasiekiami tyrėjus tenkinantys rezultatai.

Vienintelis rankinio parametrų derinimo privalumas – tyrimą atliekantis asmuo susipažįsta su modelyje naudojamais parametrais, gali lengviau interpretuoti tam tikrus parametrų pasirinkimą. Rankinio parametrų derinimo trūkumas: procesą reikia atlikti be jokio automatizavimo. Tai yra nepatogu ir per tą patį laiką išbandoma daug mažiau parametrų kombinacijų, nei naudojant kitus, automatizuotus parametrų derinimo metodus.

### 2.5.2. Tinklelio paieška (angl. Grid search)

Tinklelio paieška pradedama paruošiant modelio parametrų – kandidatų rinkinius. Kiekvienai parametrų kombinacijai yra apmokomas tas pats mašininio mokymo modelis, gaunami rezultatai. Geriausiais modelio parametrais laikomi tie, su kuriais buvo pasiektas tiksliausias rezultatas.

Tinklelio paieškos metodikos privalumas – galimybė išbandyti visus paruoštus parametrus – kandidatus. Trūkumai: išbandomi tik tyrėjų paruošti parametrų rinkiniai, tačiau jei rinkiniai dideli visų parametrų išbandymas užima labai daug laiko. Net jei su panašiais parametrais modelis rodė prastus rezultatus, kiti, mažai besiskiriantys parametrai vis tiek naudojami.

### **2.5.3. Atsitiktinė paieška (angl. Random search)**

Atsitiktinei paieškai, taip pat kaip ir tinklelio paieškos atveju, paruošiami modelio parametrų – kandidatų rinkiniai. Tačiau nesistengiama išbandyti visų įmanomų kombinacijų, kaip tai buvo daryta tinklelio paieškoje, bet kombinacijos yra parenkamos atsitiktinai. Atsitiktinis parametrų parinkimas kartojamas tyrėjų nustatytą kartų skaičių. Tada parenkami geriausiai pasirodę parametrai. Atsitiktinumas gali būti reguliuojamas. Tokiu atveju vietoje specifinių parametrų įverčių naudojamos tankio funkcijos (pavyzdžiui normalusis skirstinys).

Atsitiktinės paieškos privalumas – valdomas paieškos laikas. Trūkumai: nebūtinai parenkami geriausio modelio parametrai, kai kurie parametrų rinkinių poaibiai gali būti per retai išbandomi.

### **2.5.4. Bajesinis optimizavimas**

Bajesinio optimizavimo metodai, priešingai nei atsitiktinė paieška ir tinklelio paieška, įsimena ankstesnių iteracijų rezultatus ir juos panaudoja nustatant geriausius parametrus. Bajesinio optimizavimo metodai remiasi prielaida, kad ieškant geriausių modelio parametrų atsitiktiniu būdu dalis parametrų kombinacijų erdvių yra pranašesnės už kitas. Tai reiškia, kad jei toliau nagrinėsime geriausių parametrų kaimynus, galbūt rasime dar geresnį parametrų rinkinį. Todėl Bajesinio optimizavimo algoritmai pradeda veikti lygiai taip pat kaip atsitiktinė paieška, bet vėliau susiaurina analizuojamą erdvę.

Didžiausias Bajesinio optimizavimo metodų privalumas – potencialiai (bet nebūtinai) geresnis parametrų derinimo efektyvumas. Dažnas trūkumas – algoritmas „sustoja“ lokaliame minimume ir nebetobulėja. Darbe pateikiamos 2 populiarios Bajesinio optimizavimo implementacijos: simuliuojamas atkaitinimas ir Parzeno medžių vertinimas.

### **Simuliuojamas atkaitinimas (angl. Simulated annealing)**

Simuliuojamas atkaitinimas gali būti naudojamas randant optimalų arba beveik optimalų sprendimą diskrečioje paieškos aibėje. Metodas naudojamas kai tiriamas didžiulis kiekis potencialių sprendimų (šiuo atveju modelio parametrų kombinacijų). Simuliuojamas atkaitinimas yra svarbus sprendimo būdas sprendžiant kombinatorinius optimizavimo uždavinius, aprašytus sudėtingomis tikslo funkcijomis. Tai tikimybinis metodas, randantis globaliai optimalų tam tikros funkcijos sprendimą.

Simuliuojamo atkaitinimo metodo pavadinimas kilo iš metalurgijoje naudojamos metodikos, kai metalai yra pakaitinami ir vėliau valdomas jų vėsinimo procesas (atkaitinimas). Šis valdomas vėsinimas taip pat yra naudojamas ir simuliuojamo atkaitinimo algoritme. Taigi, taikant algoritmą ir didėjant sprendimų erdvei, mažėja tikimybė priimti prastesnius sprendimus (ieškoti naujos sprendimų erdvės) ir didėja tikimybė toliau tirti jau rasto sprendimo erdvę.

Kiekviename simuliuojamo atkaitinimo algoritmo žingsnyje, metodas vertina dabartinės būsenos kaimynus ir tikimybiškai sprendžia: ar ieškoti naujos sprendimų erdvės, ar likti esamoje.

Parametrų derinimas simuliuojamo atkaitinimo būdu susideda iš 6 žingsnių:

1. Atsitiktiniu būdu parenkamos pradinės analizuojamo modelio parametrų reikšmės – dabartinė būsena. Įvertinamas modelio veikimas.



2. Pakeičiama dabartinė būseną – atsitiktiniu būdu pakeičiama 1 parametro vertė. Randama naujos sprendimų erdvės kaimynystė.
3. Jei parametrų kombinacija jau yra išbandyta, 2 žingsnis kartojamas tol, kol nauja, netirta parametrų kombinacija yra sutinkama.
4. Įvertinamas kaimyninės būsenos veikimas.
5. Palyginami kaimyninės ir dabartinės būsenos modelių rezultatai ir nusprendžiama ar pakeisti dabartinę būseną kaimynine. Tai daroma remiantis iš anksto sudarytais priėmimo kriterijais.
6. Priklausomai nuo 5 žingsnio rezultatų gali būti kartojami 2 - 5 žingsniai.

Priėmimo kriterijai:

- Jei kaimyninė būseną pasiekia geresnius rezultatus už dabartinę, dabartinė būseną yra pakeičiama kaimynine.
- Jei kaimyninė būseną yra prastesnė už dabartinę, dabartinė būseną pakeičiama kaimynine su tikimybe  $e^{-\beta\Delta f/T}$ , čia:
  - $\beta$  – konstanta
  - $T$  – dabartinė temperatūra (parodo kokia tikimybė pakeisti sprendimo erdvę)
  - $\Delta f$  – dabartinės ir kaimyninės modelio būsenų rezultatų skirtumas

Simuliuojamo atkaitinimo algoritmas priima 4 parametrus. Metodo efektyvumas tiesiogiai priklauso nuo to, kokie parametrai bus parinkti:

1.  $\beta$  – normalizavimo konstanta.  $\beta$  pasirinkimas priklauso nuo rezultato dispersijos paieškos erdvėje.
2.  $T_0$  – pradinė temperatūra (pradinė tikimybė pereiti nuo turimų kaimynų rinkinio prie kito; algoritmo veikimo pradžioje ši tikimybė yra didelė, bet vėliau mažėja).
3. *Atkaitinimo sparta* (parodo kaip smarkiai krenta temperatūra kiekviename algoritmo žingsnyje).
4.  $N$  – iteracijų kiekis. Pasiekus šį dydį temperatūros reikšmė yra keičiama.

(Hari, 2018)

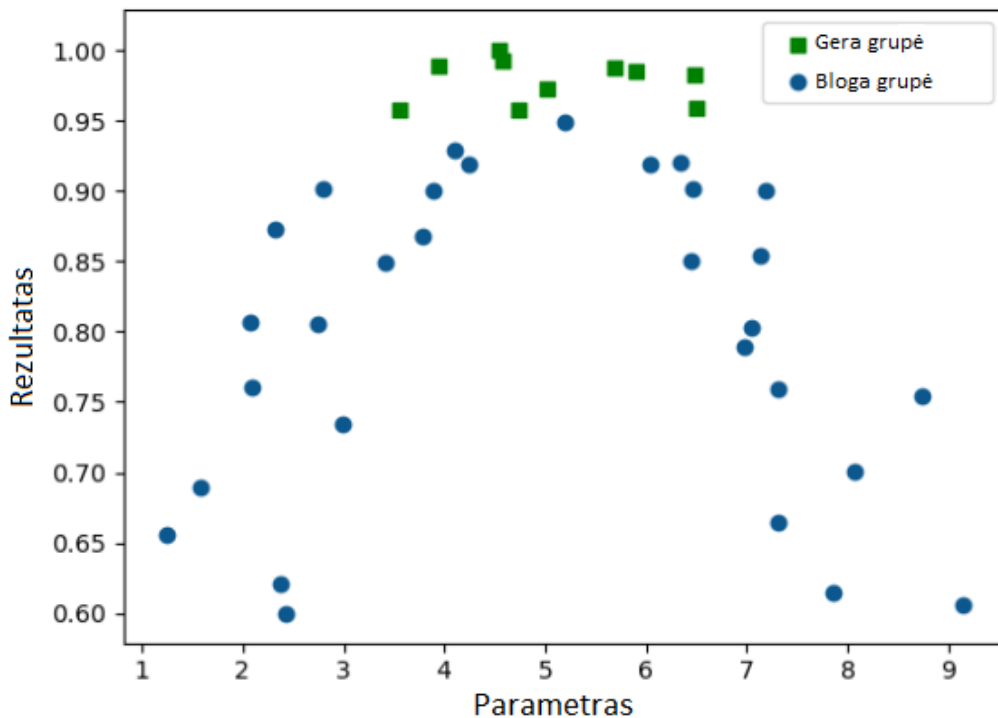
### **Parzeno medžio vertinimas (angl. Tree of Parzen estimators)**

Parzeno medžio algoritmas remiasi ankstesniais rezultatais ir sprendžia, kurį parametrų rinkinį naudoti sekančioje iteracijoje. Kaip ir simuliuojamas atkaitinimas, Parzeno medžio vertinimas yra pranašesnis už atsitiktinę ir tinklelio paiešką, nes jis išmoka pasirinkti geresnius ir atmesti prastesnius parametrų rinkinius, neieško „aklai“. Algoritmą 2011 metais straipsnyje „*Algorithms for Hyper-Parameter Optimization*“ pristatė Harvardo universiteto mokslininkas Džeimsas Bergstra.

Kaip ir anksčiau aprašytoms metodikoms, vykdant algoritmą, iš pradžių nurodoma paieškos erdvė, kurioje bus ieškoma geriausių parametrų. Tada, pirmas  $N$  iteracijų algoritmas atlieka anksčiau

aprašytą atsitiktinę paiešką ir atrenka „gerų“ ir „blogų“ parametrų rinkinius. Maždaug 20 % – 25 % parametrų derinių patenka į „gerą“ grupę, kiti - į „blogų“. Kaip matome pavyzdyje (žr. 2.12 pav.) žaliais taškais žymima „gerų“ parametrų grupė (prognozavimo tikslumas 0.95 ir daugiau), o mėlynais – „blogų“ (prognozavimo tikslumas yra mažesni nei 0.95).

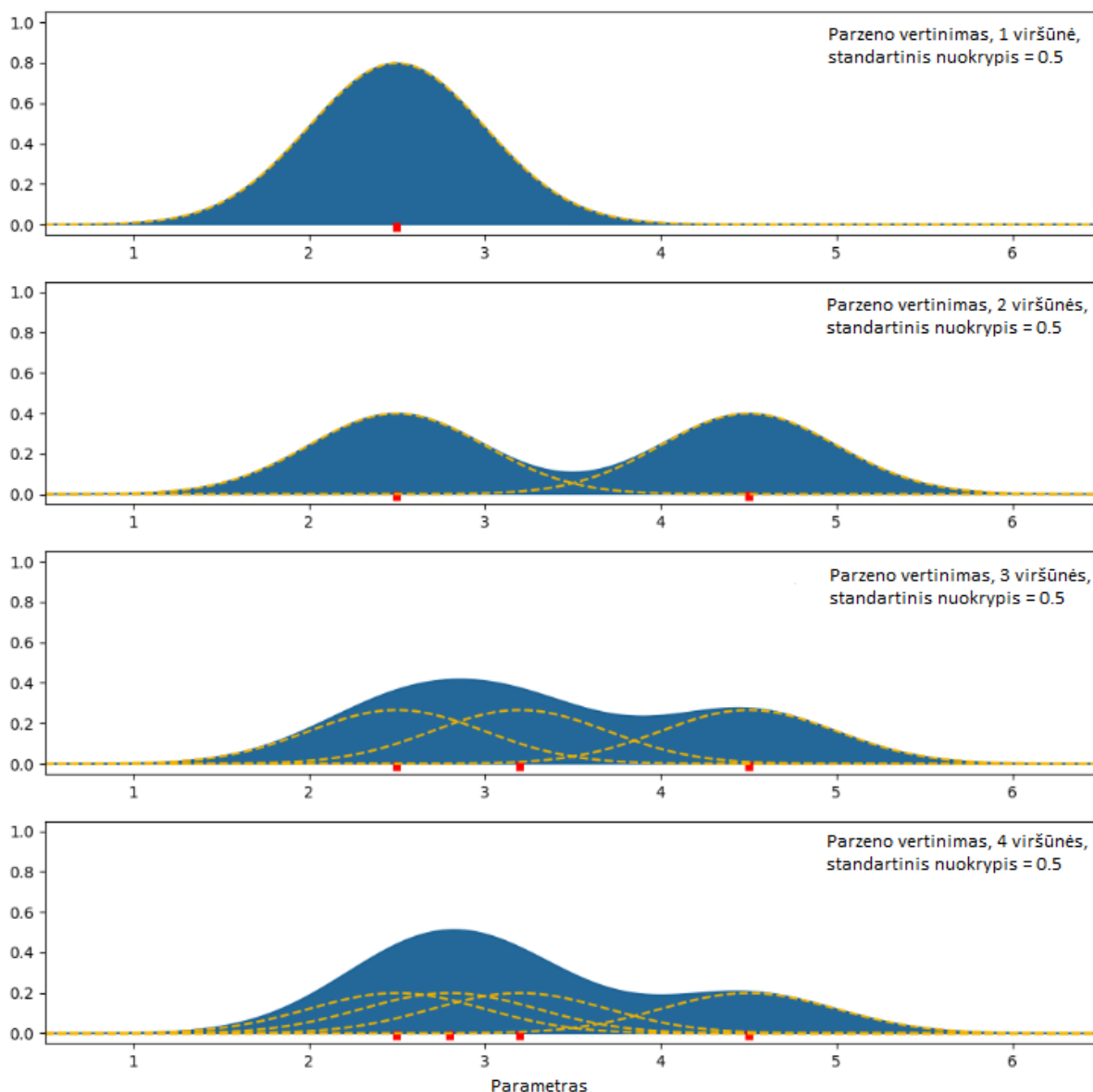
(Kopczyk, 2018)



2.12 pav. Pradinių modelio parametrų grupavimas

(Kopczyk, 2018)

Atlikus padalinimą, algoritmas sumodeliuoja „gerų“ bei „blogų“ parametrų rinkinių pasiskirstymo funkcijas. Pasiskirstymams sumodeliuoti naudojamas Parzeno lango tankio vertinimas (dar žinomas kaip branduolio tankio vertinimas). Branduolio tankio vertinimas, tai neparametris būdas iš empirinių duomenų sukurti tikimybinę tankio funkciją. Kiekvienas stebinsys empiriniuose duomenyse gali būti išreiškiamas normaliojo skirstinio pasiskirstymo funkcija, kurios vidurkis yra lygus parametro įverčiui. Normaliojo skirstinio funkcijos yra perkeliamos į vieną grafiką ir sujungiamos (žr. 2.13 pav.) 1 grafike matome tankio vertinimą vienai viršūnei, kai vidurkis yra maždaug 2.5, o standartinis nuokrypis – 0.5. Funkcija yra pasiskirsčiusi pagal normaliojo pasiskirstymo dėsnį. Iš eilės pridėdamos dar 3 viršūnės, kurių vidurkiai 4.5 (antrasis paveikslėlio grafikas), 3.2 (trečiasis paveikslėlio grafikas) ir 2.8 (ketvirtasis paveikslėlio grafikas), standartiniai nuokrypiai yra 0.5. Pridėjus naujas viršūnes, ketvirtajame paveikslėlio grafike matome apjungtą, galutinę tikimybinę pasiskirstymą.



2.13 pav. Parzeno medžio vertinimas 4 stebiniams

(Kopczyk, 2018)

„Gerų“ ir „blogų“ rinkinių pasiskirstymai yra modeliuojami atskirai. Šie pasiskirstymai naudojami skaičiuojant tikėtino pagerėjimo įvertį. Tikėtino pagerėjimo įvertis yra žymimas  $TP$  ir apskaičiuojamas:

$$TP(h) = \frac{Pr^{good}(h)}{Pr^{bad}(h)}$$

Čia  $h$  – analizuojamas parametras,  $Pr^{good}$  – tikimybė patekti į „gerą“ grupę,  $Pr^{bad}$  – tikimybė patekti į „blogą“ grupę.

Bandoma rasti tokią parametų kombinaciją, kuri turėtų didesnę tikimybę patekti į „gerą“ grupę ir mažesnę - į „blogą“. Todėl bandomi visi „geros“ grupės parametų rinkiniai – kandidatai, ieškoma geriausio tikėtino pagerėjimo įverčio. Tačiau tyrime išlaikoma iš anksto apibrėžta parametų erdvė, į kurią bandomi parametų rinkiniai privalo patekti.

Kitoje iteracijoje parametų rinkinys su didžiausiu tikėtino pagerėjimo įverčiu laikomas pradine prognoze. Algoritmas suskaičiuoja kandidato įvertį ir, remiantis ankstesniais įverčiais, vėl sumodeliuoja „gerą“ ir „blogą“ pasiskirstymus. Procesas kartojamas programuotojų nustatytą iteracijų kiekį. Pabaigus derinimą, gražinamas parametų rinkinys su didžiausiu tikėtiniu pagerėjimu (Kopczyk, 2018)

### 3. Tyrimo objektas, realizacija

Šioje dalyje apžvelgiamos mašininio mokymo metodų taikymo, prognozuojant įmonių kreditingumo reitingų pokyčius, galimybės. Lyginami skirtingų algoritmų rezultatai, analizuojant trijų svarbiausių pasaulyje kredito reitingo agentūrų bankų reitingus. Tiriama kaip tiksliai galima nuspėti „Standard & Poor`s”, „Fitch” ir „Moody`s” kreditingumo agentūrų pateikiamų kreditingumo reitingų pokyčių kryptis. Bandoma išsiaiškinti, kurios kredito reitingo agentūros reitingai labiausiai koreliuoja su finansiniais tiriamų bankų rodikliais.

Iš pradžių aprašomas duomenų gavimo procesas, kokie veiksmai buvo atlikti. Tada paaiškinama, kaip duomenys buvo parengti analizei, kokios transformacijos pritaikytos. Atliekama žvalgomoji duomenų rinkinio analizė, duomenys yra parengiami detekcijos, klasifikavimo ir išlikimo analizės uždaviniams. Detekcijos uždaviniui spręsti taikomi logistinės regresijos, atsitiktinių miškų bei atramos vektorių modeliai. Prognozuojama ar kreditingumo reitingas krito, ar ne. Sprendžiant klasifikavimo uždavinį duomenis skirstomi į 3 klases: kreditingumo reitingas mažėjo, išliko nepakitęs arba didėjo. Klasifikavimo uždaviniui naudojami atsitiktinių miškų bei atramos vektorių modeliai. Detekcijos ir klasifikavimo uždaviniuose modelių parametrai derinami tinklelio paieškos, atsitiktinės paieškos, Parzeno medžio ir simuliuojamo atkaitinimo būdais. Mašininio mokymo modelių parametrai derinami minimizuojant *logloss* įvertį. Išlikimui prognozuoti naudojami Cox ir išlikimo miškų modeliai, prognozuojamas kreditingumo reitingo kritimo įvykis. Išlikimo analizės uždavinyje bandoma iširti reitingo kritimo grėsmę per 1 metus po finansinių rodiklių pokyčio fiksavimo.

#### 3.1. Darbo priemonės

Tyrimas atliekamas naudojant *Python 3.7* programavimo kalbos versiją. Detekcijos ir klasifikavimo modelių sudarymui, patvirtinimui ir vertinimui naudojamas viena populiariausių mašininio mokymo bibliotekų pasaulyje - *sklearn*. Modelių parametrai derinami naudojant *sklearn* ir *hyperopt* bibliotekas: tinklelio paieška ir atsitiktinė paieška implementuota naudojant *sklearn* biblioteka, o *hyperopt* biblioteka naudojama atliekant Bajesinį parametrų derinimą Parzeno medžio ir simuliuojamo atkaitinimo metodikomis. Išlikimo analizei atlikti taikomos *lifelines* ir *pysurvival* bibliotekos.

#### 3.2. Duomenų gavimas

Duomenys buvo gaunami pasinaudojant Kauno technologijos universiteto ir ISM universiteto „Bloomberg“ terminalais. Duomenų gavimas buvo atliekamas 2 etapais. Iš pradžių gauta įmonių reitingų kitimo informacija. Prisijungus prie universiteto „Bloomberg“ terminalo buvo rasti istoriniai įmonių kreditingumo reitingų duomenys, kurie buvo išrinkti pagal fiksavimo datas: atrinkti kreditingumo reitingai laikotarpyje nuo 2007-01-01 iki 2019-12-31. Iš šių duomenų buvo atrinkta tik bankų informacija. Taip pat atliktas ir kreditingumo reitingo tipo filtravimas. Kadangi kiekviena agentūra pateikia skirtingus reitingo tipus, parinkti panašios paskirties bankų darbą vertinantys reitingų tipai. „Standard & Poor`s“ kreditingumo agentūrai analizuoti parinkti „LT Local Issuer Credit“ reitingo tipo duomenys. „Fitch“ kreditingumo agentūrai - „LT Issuer Default Rating“ reitingo tipo duomenys, o „Moody`s“ agentūrai – „Issuer Rating“. Kiekvienos analizuojamos kreditingumo agentūros duomenis (žr. 3.1 lentelę) sudaro 8 stulpeliai.

##### 3.1 lentelė. Kreditingumo reitingų pokyčių duomenys

Stulpelio pavadinimas	Apibūdinimas
-----------------------	--------------

Įmonės pavadinimas	Pateikiamas įmonės, kurios kreditingumo reitingą analizuojame pavadinimas. Pavadinimas rinkinyje kartojasi tiek kartų, kiek buvo užfiksuojama reitingo pokyčių. Duomenų rinkinyje yra 6907 unikalūs įmonių pavadinimai.
Kreditingumo reitingo pokyčio data	Data, kai buvo užfiksuotas kreditingumo reitingo pokytis. Gauti duomenys susideda iš įrašų, užfiksuotų tarp 2007 metų sausio 1 dienos ir 2019 metų gruodžio 31 dienos.
Kreditingumo reitingo tipas	Tyrimui nuspręsta naudoti tik „LT Local Issuer Credit“, „LT Issuer Default“ ir „Issuer Default“ reitingų tipus, nes jie laikomi vienais populiariausių tiriant bankų veiklą.
Kredito reitingo agentūra	Tyrimui pasirinkta naudoti „Standard and Poor`s“, „Fitch“ ir „Moody`s“ kredito reitingo agentūrų duomenis.
Dabartinis kreditingumo reitingas	Kreditingumo įverčio reikšmė. Geriausias „Standard and Poor`s“ duomenyse užfiksuotas kreditingumo reitingas AAA, prasčiausias – D.
Prieš tai fiksuotas kreditingumo reitingas	
Šalis	Šalis, kurioje registruota analizuojama įmonė. Duomenų rinkinyje turima informacijos apie įmones iš 92 šalių.
Identifikacinis įmonės akcijų pavadinimas	Simbolis, skirtas unikaliai identifikuoti įmonės akciją „Bloomberg“ terminale.

Lentelėje matome, jog turimi duomenys aprašo kreditingumo reitingo pokytį, datą, kada reitingas pasikeitė, šalį. Tačiau atlikti prasmingą mašininio mokymo tyrimą reikia gerokai daugiau informacijos. Todėl nuspręsta gauti papildomus istorinius tiriamų įmonių finansinius rodiklius. Atlikti veiksmai:

1. Prisijungus prie „Bloomberg“ terminalo buvo nustatyti 89 įmonių finansinės veiklos rodikliai, tokie kaip: pagrindinio kapitalo grąža, trumpalaikiai ir ilgalaikiai įsiskolinimai, pelno augimas pastaraisiais metais ir t.t. Šių rodiklių trumpiniai buvo išsaugoti tekstiniame faile, tokiu būdu juos patogiau pasiekti tolimesniuose žingsniuose. Visi darbe naudojami finansiniai rodikliai aprašyti 1 priede.
2. Parašytas VBA Excel kodas, naudojantis parinktus rodiklių trumpinius ir įmonės akcijų simbolius. Šis kodas formuoja Excel užklausas, kurios naudoja „Bloomberg“ aplikacijų programavimo sąsają.
3. „Bloomberg“ aplikacijų programavimo sąsajos užklausos grąžina istorinius kiekvienos tiriamos įmonės finansinių rodiklių pokyčius. Šiuos finansinių rodiklių pokyčius naudosime tolimesnėje analizėje.

Atlikę visus duomenų gavimo žingsnius turime 2 duomenų rinkinius. Pirmasis - kreditingumo reitingų rinkinys, sudarytą iš stulpelių, aprašytą 3.1 lentelėje ir įmonių finansinių rodiklių. Antrasis rinkinys yra sudarytas iš identifikacinio įmonės akcijų pavadinimo, 89 finansinės veiklos rodiklių bei datos kai šie rodikliai buvo užfiksuoti. Finansinės veiklos rodikliai aprašyti 1 priede.

### 3.3. Duomenų apdorojimas, paruošimas darbui

Atliekamas duomenų sutvarkymas ir paruošimas analizei. Darbai atlikti žemiau aprašyta tvarka:

Iš pradžių, iš 3.2 poskyryje aprašyto finansinės veiklos rodiklių duomenų rinkinio pašalinamos duomenų gavybos metu išlikę „Bloomberg“ aplikacijų programavimo sąsajos funkcijų pavadinimai.

Duomenys apie dabartinę ir prieš tai buvusią kreditingumo reitingo vertę yra sujungiami: sukuriamas reitingo pokyčio stulpelis. Jeigu dabartinio kreditingumo reitingo reikšmė yra didesnė už prieš tai buvusią kreditingumo reitingo reikšmę, pokyčio stulpelyje rašomas „1“. Jei fiksuojama priešinga situacija: dabartinis kreditingumo reitingas yra mažesnis už prieš tai užfiksuotą, tai reitingo pokyčio stulpelyje pažymima „-1“. Trečiu atveju – jei dabartinis ir prieš tai užfiksuotas kreditingumo reitingai sutampa, reitingo pokyčio stulpelyje rašomas „0“.

Atlikus aukščiau aprašytus veiksmus turimi rodiklių ir kreditingumo reitingų duomenų rinkiniai yra sujungiami. Prie kiekvienos kreditingumo reitingo agentūros („Standard & Poor`s“, „Fitch“ ir „Moody`s“) duomenų prijungiami finansinių rodiklių pokyčiai. Duomenų rinkinių sujungimas vykdomas atsižvelgus į įmonę, reitingo pokyčio datą bei finansinio rodiklio užfiksavimo datą. Toliau pateikiama rinkinių apjungimo darbo eiga:

1. Panaudojant identifikacinį įmonės akcijų pavadinimą abiejuose duomenų rinkiniuose randami tos pačios įmonės įrašai.
2. Kiekvienai finansinių rodiklių nustatymo datai buvo rastas artimiausias per ateinančius 12 mėnesių užfiksuotas kreditingumo reitingo pokytis. Jei reitingo pokytis per 12 mėnesių nebuvo fiksuojamas, laikyta, kad kreditingumo reitingas nesikeitė.
3. Prie finansinio rodiklio stebinio prijungtas šis artimiausias kreditingumo reitingo pokytis, įmonės šalis bei verslo sektorius.

Siekiant padidinti naudojamų rodiklių skaičių ir gauti daugiau potencialiai naudingos informacijos atliekama bruožų inžinerija. Šalia kiekvieno finansinės veiklos rodiklio pridedami nauji kintamieji:

1. Kintamasis, aprašantis ankstesnę finansinio rodiklio reikšmę 1 metų laikotarpyje (jei egzistuoja).
2. Kintamasis, nusakantis skirtumą tarp finansinio rodiklio reikšmės dabar ir užfiksuotos prieš tai 1 metų laikotarpyje (jei yra).
3. Kintamasis, rodantis finansinio rodiklio reikšmės santykį su prieš tai buvusiu finansinio rodiklio įverčiu 1 metų laikotarpyje (jei yra).
4. Kintamasis, kurio elementų reikšmės yra 1 arba 0, priklausomai nuo to ar analizuojamo rodiklio vertė yra pateikta, ar praleista.
5. Kintamasis, žymintis finansinio rodiklio fiksavimo ketvirtį (sukūrus šį kintamąjį finansinio rodiklio fiksavimo data buvo pašalinta)
6. Prie kiekvieno iš kategorinių kintamųjų (šalis, įmonė, ketvirtis) buvo sukurtas dažnio kintamasis. Šis kintamasis žymi kaip dažnai duomenų rinkinyje sutinkama tam tikra kintamojo vertė.

Kategoriniai kintamieji taip pat buvo pritaikyti naudoti mašininio mokymo metoduose. Todėl jiems buvo atliktas vienietinis kodavimas (kiekvienai klasei sukuriama po stulpelį, kur atitinkamas

žymimas „1“ ir neatitikimas žymimas „0“). Detekcijos ir klasifikavimo uždaviniuose kintamiesiems, nusakantiems įmonę, šalį, dabartinį kreditingumo reitingą, reitingo fiksavimo ketvirtį atliekamas vienetinis kodavimas. Išlikimo analizėje vienetinis kodavimas taikomas kategoriniams kintamiesiems, nusakantiems šalį, dabartinį kreditingumo reitingą ir reitingo fiksavimo ketvirtį..

Siekiant sumažinti tikimybę, jog modeliai persimokys bei sutrumpinti algoritmų skaičiavimo laiką naudojamas kintamųjų atrinkimas (angl. feature selection). Kintamųjų atrinkimo metodika remiasi prielaida, jog esant mažai arba artimai 0 dispersijai, kintamasis gali būti apytikriai prilyginamas konstantai. Jeigu tam tikras kintamasis yra konstanta, jis visiškai neprisidės prie modelio rezultatų gerinimo ir gali būti pašalinamas. Remiantis šiomis prielaidomis pašalinti kintamieji, kurių dispersija yra mažesnė nei 0.3. Šis žingsnis pastebimai paspartino apmokymo procesą.

### 3.4. Duomenų apžvalga

Apžvelgiami tyrime naudojami sujungti trijų („Standard & Poor`s“, „Fitch“ ir „Moody`s“) kreditingumo reitingo agentūrų bei jų finansinių rodiklių duomenų rinkiniai. Kiekvienam duomenų rinkiniui yra vykdomas tikrinimas dėl klasių disbalanso: įvertinamas stebinių skaičius, kai kreditingumo reitingas mažėja, nesikeičia ir didėja.

Lentelėje (žr. 3.2 lentelę) matomas didžiulis klasių disbalansas „Standard & Poor`s“ agentūros duomenyse: beveik 11839 stebiniuose fiksuojamas reitingo stabilumas (jis nesikeičia), 1059 stebiniuose fiksuojamas reitingo pokytis. Reitingo kilimas fiksuojamas šiek tiek dažniau, nei jo kritimas: 641 stebinyje fiksuojamas kreditingumo reitingo didėjimas ir 418 stebinių matomas kreditingumo reitingo mažėjimas.

3.2 lentelė. „Standard & Poor`s“ kreditingumo reitingo pokyčiai

Reitingo pokyčio kryptis	Stebinių kiekis
Reitingas didėjo	641
Reitingas nesikeitė	11839
Reitingas mažėjo	418

„Fitch“ kreditingumo reitingo agentūros duomenyse (žr. 3.3 lentelę) išvelgiamas dar didesnis klasių disbalansas: 526 stebiniuose fiksuojamas reitingo didėjimas, net 12076 – reitingo reikšmė nekinta ir 296 – kreditingumo reitingas mažėja.

3.3 lentelė. „Fitch“ kreditingumo reitingo pokyčiai

Reitingo pokyčio kryptis	Stebinių kiekis
Reitingas didėjo	526
Reitingas nesikeitė	12076
Reitingas mažėjo	296

„Moody`s“ kreditingumo agentūros duomenyse (žr 3.4 lentelę) matomas pats didžiausias tiriamų klasių disbalansas: reitingo didėjimas fiksuojamas 99 stebiniuose, nesikeitimas – 12593 stebiniuose ir mažėjimas 206 stebiniuose.

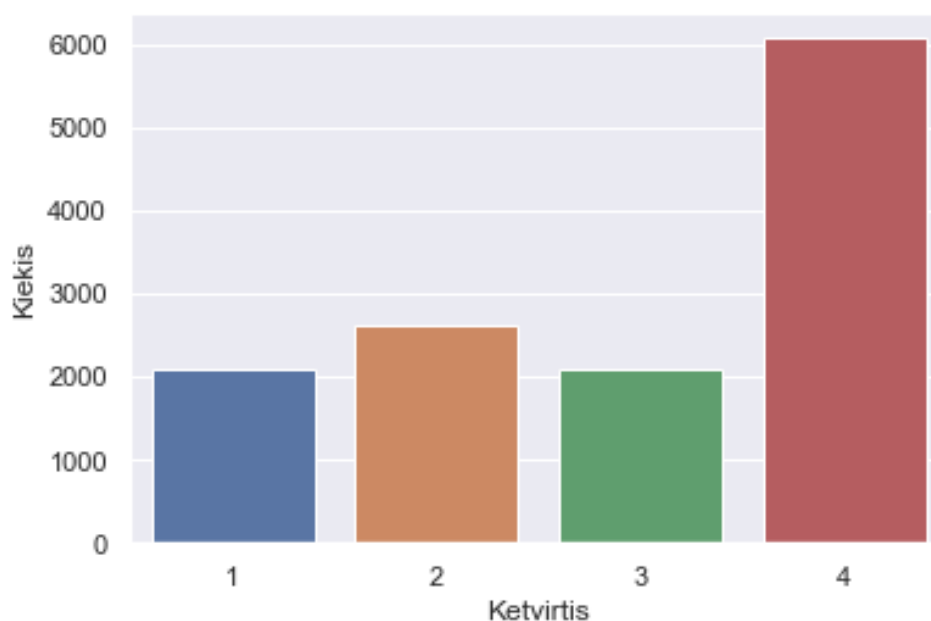
3.4 lentelė. *Moody`s* kreditingumo reitingo pokyčiai



Reitingo pokyčio kryptis	Stebinių kiekis
Reitingas didėjo	99
Reitingas nesikeitė	12593
Reitingas mažėjo	206

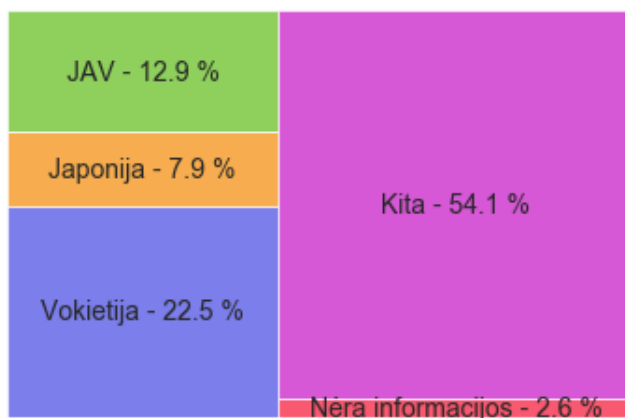
Visuose 3 tiriamuose duomenų rinkiniuose užfiksuotas labai didelis klasių disbalansas. Siekiant minimizuoti šio reiškinio įtaką rezultatams buvo nuspręsta modelių apmokymo metu stebiniams suteikti svorius (svarbą). Jei tam tikras stebiny priskirtas itin dažnai sutinkamai klasei, tai pritaikoma mažas svoris (svarba). Kitu atveju, jei stebiny priskirtas retai sutinkamai klasei, tai jam taikomas didelis svoris (svarba) modelio apmokymo procese.

Atliekamas finansinių rodiklių fiksavimo metų ketvirčio tyrimas. Apžvelgiami analizuojami 2007 metų sausio – 2019 metų gruodžio duomenys. Tikrinama, kuriuo metu dažniausiai užfiksuoti finansinių rodiklių pokyčiai. Rasti pastebėjimai fiksuojami grafike (žr. 3.1 pav.) 1, 2 ir 3 metų ketvirčiais fiksuoti panašūs stebinių kiekiai (daugiau nei 2000, bet mažiau nei 3000 reikšmių). Tačiau pastebimas didžiulis finansinių rodiklių fiksavimo augimas 4 metų ketvirtyje. Daroma prielaida, jog šis skirtumas susidaro dėl to, kad dažniausiai būtent metų pabaigoje „Bloomberg“ gauna įmonių finansinius rodiklius.



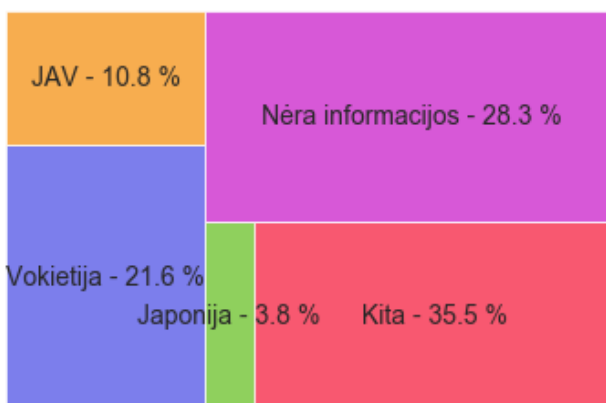
**3.1 pav.** Duomenų fiksavimo laikas, metų ketvirtis

Toliau tiriama, kuriose šalyse registruoti tiriami bankai. Iliustracijoje (žr. 3.2 pav.) vaizduojamas „Standard & Poor`s“ reitingo agentūros stebiniuose sutinkamų bankų pasiskirstymas pagal valstybę (tas pats bankas gali kartotis keliuose stebiniuose). Matome, kad beveik ketvirtadalis (22.5 %) tiriamų stebinių bankų yra registruoti Vokietijoje, 12.9 % - Jungtinėse Amerikos Valstijose, 7.9 % - Japonijoje. Kiek daugiau nei pusė (54.1 %) stebinių bankų yra registruoti kitose valstybėse. Apie 2.6 % stebinių nepateikia informacijos apie banko registravimo šalį.



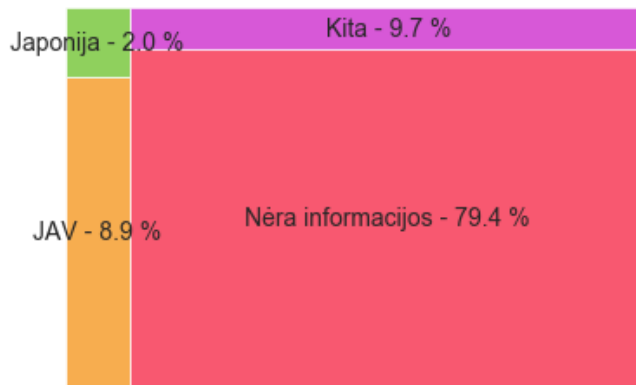
**3.2 pav.** Tiriamų „Standard & Poor`s“ įmonių registravimo šalys

„Fitch“ kreditingumo agentūros duomenų (žr. 3.3 pav.) pasiskirstymas pagal registravimo šalį yra labai panašus į „Standard & Poor`s“. 21.6 % stebinių bankų registruoti Vokietijoje, 10.8 % - Jungtinėse Amerikos Valstijose, 3.8 % - Japonijoje, 35.5 % - kitose valstybėse. Įdomu, kad „Fitch“ kreditingumo reitingo agentūros duomenyse trūksta šalies informacijos apie 28.3 % stebiniuose sutinkamų bankų.



**3.3 pav.** Tiriamų „Fitch“ įmonių registravimo šalys

„Moody`s“ kreditingumo reitingo agentūros duomenys (žr. 3.4 pav.) išsiskiria dideliu trūkstamos informacijos apie tiriamų bankų registracijos šalį kiekiu. Neturima informacijos net apie 79.4 % tiriamų stebinių bankų. Tuo metu 8.9 % - registruoti Jungtinėse Amerikos Valstijose, 2 % - Japonijoje ir 9.7 % - kitose valstybėse.



3.4 pav. Tiriamų „Moody`s“ įmonių registravimo šalys

### 3.5. Modeliavimas

Kreditingumo reitingo kryptis prognozuojama sprendžiant 3 uždavinius: detekcijos, klasifikavimo bei išlikimo analizės. Modeliavimas vykdomas atliekant 5 dalių kryžminį patikrinimą. Pateikiamas modeliavimo algoritmas, kuris buvo taikomas detekcijos ir klasifikavimo uždaviniuose. Šis algoritmas buvo taikomas visų 3 tiriamų kreditingumo reitingo agentūrų prognozėms:

1. Kategoriniams kintamiesiems atliekamas vienetinis kodavimas, šie kategoriniai stulpeliai pašalinami, paliekami tik užkoduoti.
2. Duomenyse atskiriamas prognozuojamas kintamasis – reitingo pokytis. Šiame kintamajame „1“ žymimas reitingo didėjimas, „0“ – nepasikeitimas ir „-1“ – kreditingumo reitingo sumažėjimas.
3. Jei atliekamas detekcijos uždavinys, reitingo pokyčio stulpelyje įvykdomi pakeitimai: reitingo mažėjimas žymimas „1“ (įvykis įvyko), o nesikeitimas arba reitingo didėjimas žymimi „0“ (įvykis neįvyko). Atliekant klasifikavimo uždavinį kreditingumo reitingo pokyčio krypties stulpelis nėra redaguojamas.
4. Atliekamas kintamųjų parinkimas (angl. feature selection).
5. Duomenys padalinami į 5 dalis: 4 dalys yra skirtos modelių apmokymui, 1 – rezultatų įvertinimui. Tolesni veiksmai atliekami kiekvienai iš 5 dalių atskirai.
6. Jei po padalinimo kažkuris kintamasis yra sudarytas tik iš tuščių reikšmių, jis yra pašalinamas ir einamojoje iteracijoje nėra naudojamas.
7. Likusios praleistos reikšmės užpildomos atitinkamų kintamųjų vidurkiais.
8. Visiems nekategoriniams duomenims atliekamas mastelio keitimas (angl. scaling) ir centravimas. Mastelio keitimas atliekamas siekiant, kad didelę dispersiją fiksuojantys kintamieji nenulemtų modelio rezultatų. Mastelio keitimui atlikti naudojamas *sklearn* bibliotekos *StandardScaler* metodas, kuris iš pradinės kintamojo reikšmės atima kintamojo vidurkį ir rezultatą padalina iš standartinio nuokrypio.
9. Atliekamas vidinis 5 dalių vidinis kryžminis patikrinimas, kurį naudojant surandami geriausi kiekvieno analizuojamo modelio parametrai.
10. Kiekvienam detekcijos modeliui apskaičiuojami *logloss*, *f1*, *AUC ROC*, balansuoto tikslumo įverčiai ir sudaroma sumaišymo matrica. Klasifikavimo uždavinyje sudaroma sumaišymo matrica.
11. Detekcijos uždavinyje nubrėžiamos *AUC ROC*, Preciziškumo – Prisiminimo (angl. Precision – Recall) kreivės, lyginančios skirtingų *metodų* prognozių rezultatus.

Pabaigus detekcijos uždavinio modeliavimo procesą sudarytos ROC, Preciziškumo – Prisiminimo (angl. Precision – Recall) kreivės, lyginančios kaip tiksliai pavyko nuspėti kiekvienos kreditingumo reitingo *agentūros* sudarytus reitingo pokyčius. Ištiriama, kuris mašininio mokymo metodas atliko tiksliausias prognozes.

Atlikus klasifikavimo uždavinio modeliavimą, lyginamos visų 3 kreditingumo reitingo agentūrų sumaišymo matricos. Taip pat analizuojama, kuris mašininio mokymo modelis atliko geriausias prognozes su kiekvienos kreditingumo reitingo agentūros duomenimis.

Vykdam išlikimo analizę nėra tiesiogiai tiriama ar kreditingumo reitingas sumažės, ar ne. Šiame uždavinyje tiriamos kiekvienos dienos tikimybės, jog kreditingumo reitingas sumažės. Tyrimas reitingų pokyčiai per 1 metus nuo finansinių rodiklių pasikeitimo datos. Kadangi išlikimo analizės uždavinys skiriasi savo specifika, todėl jo modeliavimo procedūra pateikiama atskirai:

1. Atliekamas kategorinių kintamųjų vienetinis kodavimas.
2. Vietoje to, kad būtų atskirtas kintamasis reitingo pokyčio kryptiniai, sudaromi 2 nauji kintamieji. Pirmasis kintamasis nusako, ar per 1 metus nuo finansinių rodiklių pokyčio datos kreditingumo reitingas krito. Jei toks įvykis fiksuojamas, kintamasis įgauna reikšmę „1“, o jeigu ne – „0“. Antrasis sukurtas kintamasis nusako laiko tarpą dienomis iki įvykio užfiksavimo. Jeigu reitingas sumažėjo po  $n$  dienų, tai šis kintamasis įgauna reikšmę  $n$ . Lygiai taip pat, jei reitingas padidėjo po  $n$  dienų, užfiksuojama ir ši reikšmė. Jeigu per 1 metus nebuvo fiksuojama jokių reitingo pokyčių, kintamajam priskiriama dienų skaičiaus 1 metuose reikšmė - 365.
3. Atliekamas kintamųjų parinkimas (angl. feature selection).
4. Kaip ir vykdant detekciją ir klasifikavimą, duomenys padalinami į 5 dalis: 4 dalys yra skirtos modelių apmokymui, 1 – rezultatų įvertinimui. Tolesni veiksmai atliekami kiekvienai iš 5 dalių atskirai.
5. Jei po padalinimo kažkuris kintamasis yra sudarytas tik iš tuščių reikšmių, jis yra pašalinamas ir einamojoje iteracijoje nebėra naudojamas.
6. Likusios praleistos reikšmės užpildomos atitinkamų kintamųjų vidurkiais.
7. Visiems nekategoriniams duomenims atliekamas mastelio keitimas (angl. scaling) ir centravimas. Mastelio keitimas atliekamas siekiant, kad didelę dispersiją fiksuojantys kintamieji nenulemtų modelio rezultatų. Mastelio keitimui atlikti naudojamas *sklearn* bibliotekos *StandardScaler* metodas, kuris iš pradinės kintamojo reikšmės atima kintamojo vidurkį ir rezultatą padalina iš standartinio nuokrypio.
8. Kiekvienam *išlikimo analizės modeliui* apskaičiuojamos charakteristikos: *C - index*, *Brier* balas, šaknis iš vidutinės kvadratinės paklaidos (*RMSE*), vidutinė išlikimo tikimybė, medianinė absoliuti paklaida bei vidutinė absoliuti paklaida.

Skirtingų *kreditingumo reitingo agentūrų* modeliai palyginami naudojant vidutinę išlikimo tikimybę.

### 3.5.1. Detekcijos uždavinio rezultatai

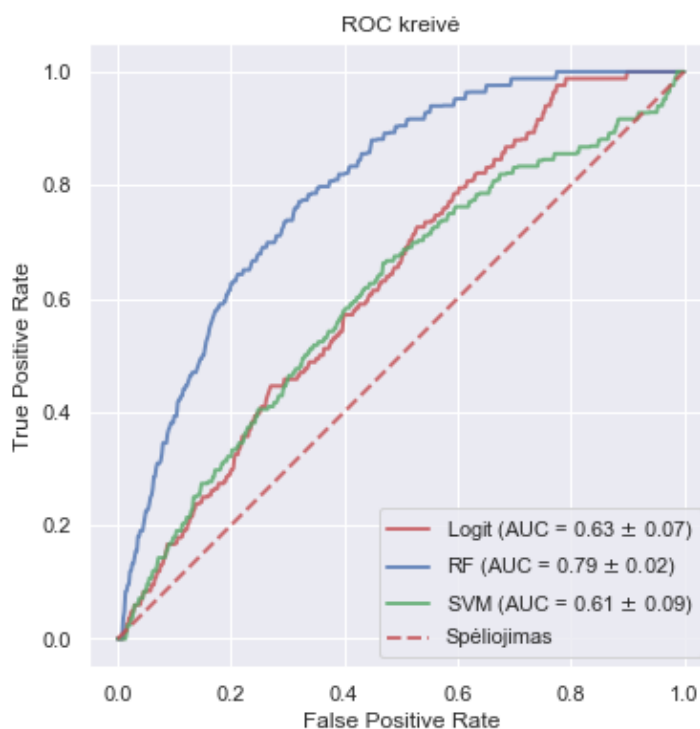
Detekcijos uždavinyje kreditingumo reitingo sumažėjimas žymimas „1“, o padidėjimas arba nepasikeitimas žymimi „0“. Detekcijos uždaviniui spręsti buvo naudoti logistinės regresijos, atsitiktinių miškų bei atramos vektorių modeliai. Visų modelių parametrų derinimui buvo išbandyti tinklo paieškos (angl. grid search), atsitiktinės paieškos (angl. random search), simuliuojamo

atkaitinimo (angl. simulated annealing) bei Parzeno medžio (angl. Parzen tree) algoritmai. Modelių parametru derinimui ir vertinimui naudotas *logloss* įvertis.

Logistinės regresijos modeliavimui naudotas *python sklearn* bibliotekos metodas „LogisticRegression“. Atsitiktinių miškų modeliavimui – „RandomForestClassifier“. Atramos vektorių modeliavimui taikytas „LinearSVC“ metodas. Klasių disbalanso efekto mažinimui pritaikytas subalansuotų svorių parametras. Tokiu būdu abiem prognozuojamoms klasėms priskirti svoriai buvo atvirksčiai proporcingi jų dažniui.

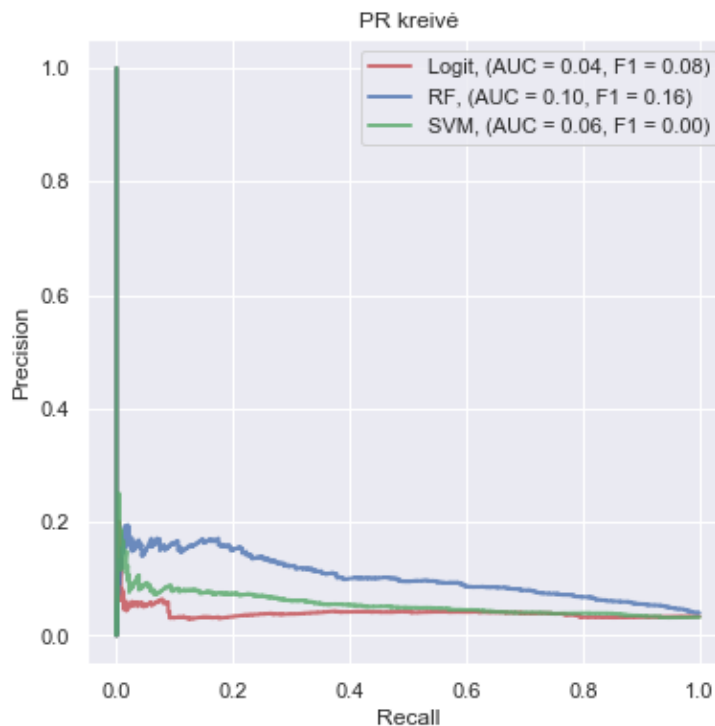
### **Standard & Poor`s agentūros duomenų rinkinio prognozių rezultatai**

Atlikę detekciją „Standard & Poor`s“ agentūros reitingo pokyčiams, apžvelgiame gautus rezultatus. Iš pradžių patikriname kaip tiksliai skirtingi mašininio mokymo modeliai atskiria reitingo mažėjimą nuo reitingo didėjimo / nesikeitimo. Vertinant klasių atskyrimo kokybę braižoma AUC - ROC kreivė (žr. 3.5 pav.) Grafike raudona spalva žymimas logistinės regresijos modelis, mėlyna spalva vaizduojamas atsitiktinių miškų modelis, žalia – atramos vektorių modelis, o punktyrine linija – atsitiktinis spėliojimas. Grafiko legendoje šalia kiekvieno modelio ploto įverčio (AUC) išskiriamas ir standartinis nuokrypis (žymimas po +/- simbolio). Grafike aiškiai matosi, jog geriausią atskyrimą tarp klasių atlieka atsitiktinių miškų modelis, fiksuojantis aukščiausią AUC įvertį: 0.79. Šis modelis pasižymi ir mažiausiu standartiniu nuokrypiu. Logistinės regresijos bei atramos vektorių modeliai rodo labai panašius detekcijos rezultatus, jų fiksuojami AUC yra: 0.63 ir 0.61 atitinkamai.



**3.5 pav.** „Standard & Poor`s“ agentūros prognozių ROC kreivė

Toliau tikrindami detekcijos rezultatus brėžiame Preciziškumo – Prisiminimo kreivę (žr. 3.6 pav.) Kadangi didžioji dalis duomenų priklauso teigiamai klasei, duomenyse fiksuojamas didelis klasių disbalansas ir rezultatuose gaunama daug I tipo klaidų, todėl Preciziškumo – Prisiminimo kreivė rodo prastus rezultatus. Visgi geriausią įvertinimą fiksuoja atsitiktinių miškų modelis (kreivėje žymimas mėlyna spalva). Modelio F1 įvertis yra lygus 0.16.



3.6 pav. „Standard & Poor`s” agentūros prognozių PR kreivė

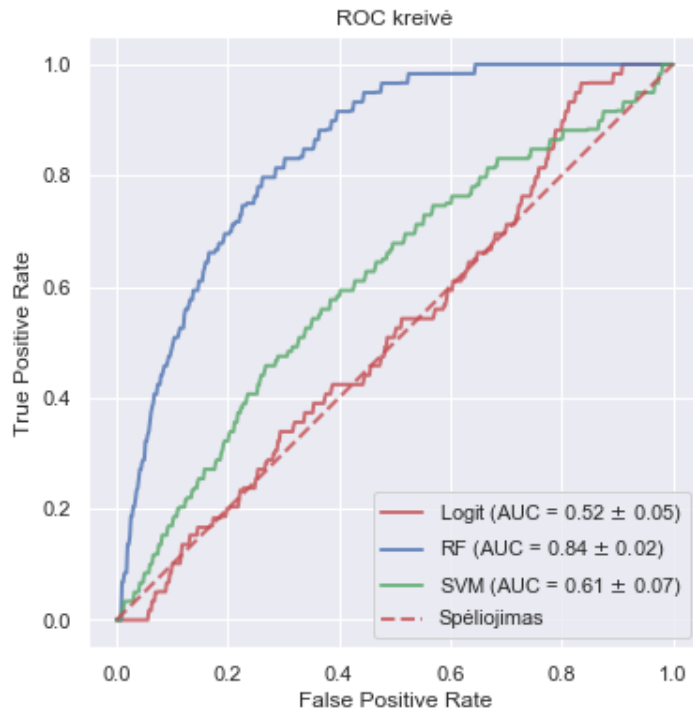
Lentelėje (žr. 3.5 lentelė) matomi „Standard & Poor`s“ agentūros detekcijos uždavinio *logloss* bei subalansuoto tikslumo įverčiai, kurie yra apskaičiuojami suvidurkinus 5 kryžminio patikrinimo dalių rezultatus. Matome, jog atramos vektorių metodo *logloss* įvertis yra mažiausias ir modelio prognozės yra arčiausiai realių verčių. Tačiau pažvelgę į subalansuotą tikslumo įvertį pastebime, jog jis įvertintas tik 0.5. Tai reiškia, jog atramos vektorių modelis geba aptikti tik vieną klasę. Tuo tarpu atsitiktinių miškų modelio subalansuoto tikslumo įvertis siekia 0.672 ir todėl jis geba atlikti klasių atskyrimą.

3.5 lentelė. „Standard & Poor`s” duomenų rinkinio prognozavimo modelių palyginimas

Modelis	<i>Logloss</i> įverčio vidurkis	Subalansuoto tikslumo vidurkis
Logistinės regresijos modelis	0.804	0.566
Atsitiktinių miškų modelis	0.460	<b>0.672</b>
Atramos vektorių modelis	<b>0.144</b>	0.5

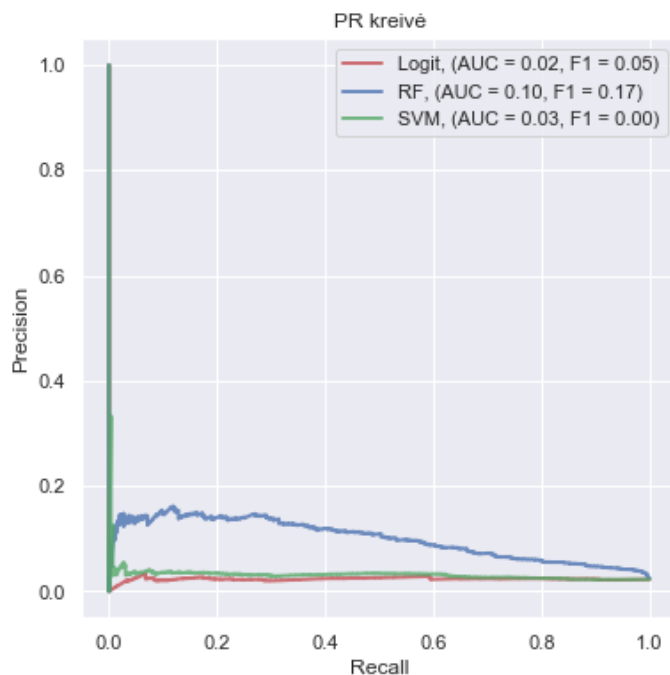
### „Fitch“ agentūros duomenų rinkinio prognozių rezultatai

„Fitch“ kreditingumo reitingo agentūros duomenims irgi nubraižyta ROC kreivė (žr. 3.7 pav.) Grafike mėlyna spalva žymimi atsitiktinių miškų metodo rezultatai gana stipriai lenkia kitų 2 metodų rezultatus (žymimų raudona ir žalia spalvomis). Šio metodo AUC įvertis: 0.84. Logistinės regresijos metodo AUC: 0.52 (rezultatas beveik lygus atsitiktiniam spėliojimui) ir atramos vektorių metodo AUC: 0.61.



3.7 pav. „Fitch“ agentūros prognozių ROC kreivė

„Fitch“ agentūros duomenų rinkinio prognozėms įvertinti nubrėžta Preciziškumo – Prisiminimo kreivė (žr. 3.8 pav.) Vėl matome, jog dėl didžiulio klasių disbalanso, visų 3 modelių rezultatai yra labai prasti. Geriausiai įvertintas mėlyna spalva žymimas atsitiktinių miškų modelis, kurio F1 įvertis: 0.17.



3.8 pav. „Fitch“ agentūros prognozių PR kreivė

Lentelėje (žr. 3.6 lentelė) matomi „Fitch“ reitingo agentūros detekcijos uždavinio *logloss* bei subalansuoto tikslumo įverčiai, kurie yra apskaičiuojami apskaičiuojant 5 kryžminio patikrinimo dalių

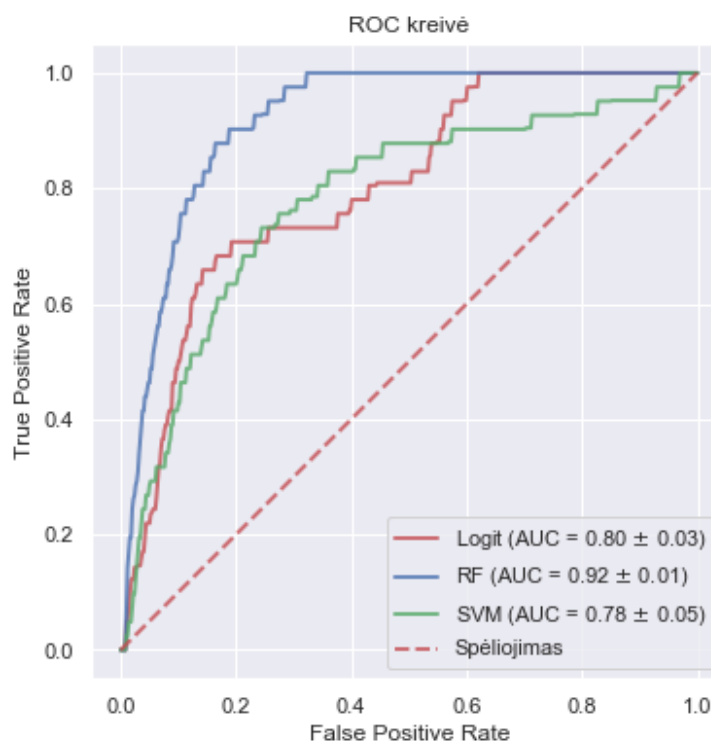
vidutinius rezultatus. Matome, jog atramos vektorių metodo *logloss* įvertis yra mažiausias (tik 0.11) ir modelio prognozės yra arčiausiai realių verčių. Tačiau pažvelgę į subalansuotą tikslumo įvertį pastebime, jog jis įvertintas tik 0.5. Tai reiškia, jog atramos vektorių modelis geba aptikti tik vieną klasę (šiuo atveju kreditingumo reitingo nepasikeitimo / didėjimo). Tuo tarpu atsitiktinių miškų modelio subalansuoto tikslumo įvertis siekia 0.687 ir jis geba aptikti tiek reitingo mažėjimo, tiek ir nesikeitimo / didėjimo įvykius.

**3.6 lentelė.** *Fitch* duomenų rinkinio prognozavimo modelių palyginimas

Modelis	<i>Logloss</i> įverčio vidurkis	Subalansuoto tikslumo vidurkis
Logistinės regresijos modelis	3.527	0.514
Atsitiktinių miškų modelis	0.368	<b>0.687</b>
Atramos vektorių modelis	<b>0.110</b>	0.5

### „Moody`s” agentūros duomenų rinkinio prognozių rezultatai

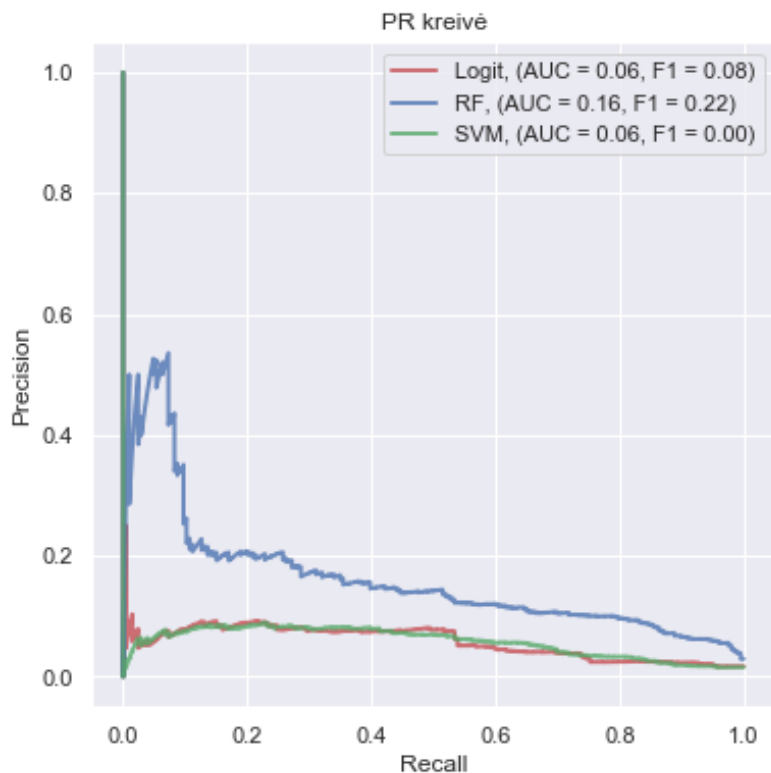
Paveikslėlyje (žr. 3.9 pav.) vaizduojama „Moody`s” agentūros duomenų rinkinio prognozių ROC kreivė. Iš karto pastebime, jog visi 3 kreivėje vaizduojami metodai geba gana tiksliai atlikti prognozes. Raudona spalva žymima logistinė regresija užfiksavo aukštą 0.8 AUC įvertį, o žalia spalva išskirtas atramos vektorių modelio AUC: 0.78. Tačiau tiksliausiai prognozes atliko atsitiktinių miškų modelis, kurio AUC įvertis siekė net 0.92.



**3.9 pav.** „Moody`s” agentūros prognozių ROC kreivė

„Moody`s” agentūros prognozėms nubrėžta Preciziškumo – Prisiminimo kreivė (žr. 3.10 pav.) Kreivės rezultatai buvo stipriai paveikti klasių disbalanso. Visgi geriausiai kreivėje vertinama atsitiktinių miškų modelio prognozė (F1 = 0.22). Prasčiau vertinama logistinės regresijos modelio prognozė (F1 = 0.08) ir prasčiausiai vertinama atramos vektorių modelio prognozė.





3.10 pav. „Moody`s“ agentūros prognozių PR kreivė

Lentelėje (žr. 3.7 lentelė) matomi „Moody`s“ agentūros detekcijos uždavinio *logloss* bei subalansuoto tikslumo įverčiai, kurie yra apskaičiuojami apskaičiuojant 5 kryžminio patikrinimo dalių vidutinius rezultatus. Matome, jog atramos vektorių metodo *logloss* įvertis yra mažiausias (vos 0.07) ir modelio prognozės yra arčiausiai realių verčių. Tačiau pažvelgę į subalansuotą tikslumo įvertį pastebime, jog jis įvertintas tik 0.5. Tai reiškia, jog atramos vektorių modelis geba aptikti tik vieną klasę (šiuo atveju kreditingumo reitingo nesikeitimo / didėjimo). Tuo tarpu atsitiktinių miškų modelio subalansuoto tikslumo įvertis siekia 0.699 ir jis geba aptikti tiek reitingo mažėjimo, tiek ir nesikeitimo / didėjimo įvykius.

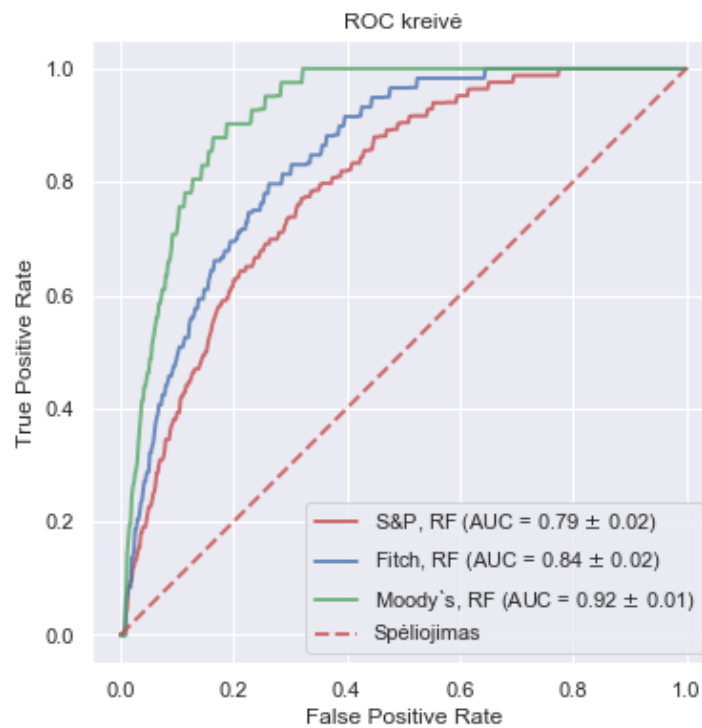
3.7 lentelė. „Moody`s“ duomenų rinkinio prognozavimo modelių palyginimas

Modelis	<i>Logloss</i> įverčio vidurkis	Subalansuoto tikslumo vidurkis
Logistinės regresijos modelis	0.581	0.662
Atsitiktinių miškų modelis	0.260	<b>0.699</b>
Atramos vektorių modelis	<b>0.07</b>	0.5

### Visų 3 kreditingumo agentūrų detekcijos prognozių palyginimas

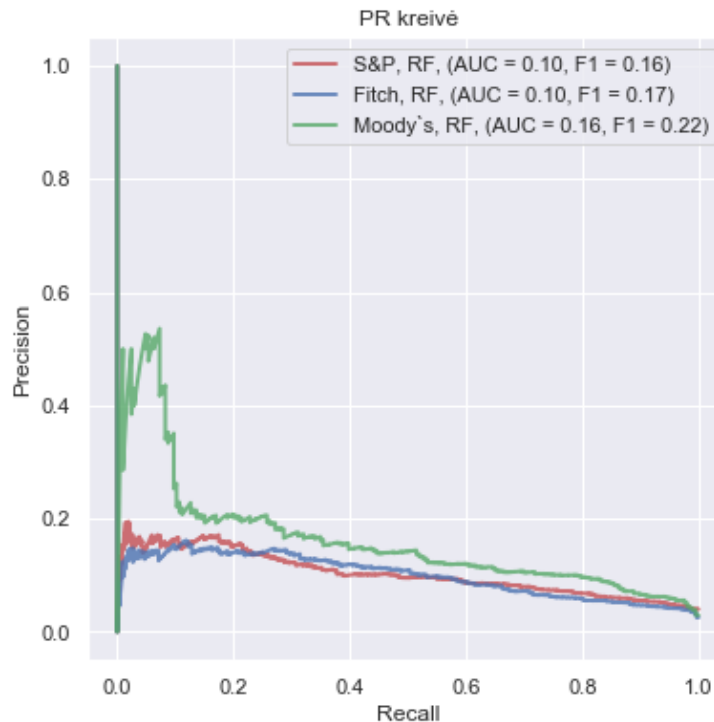
Turėdami visų agentūrų prognozių rezultatus, atliekame jų palyginimą. Palyginimui kiekvienai reitingo agentūrai buvo parinktas geriausias prognozes atlikęs mašininio mokymo modelis. Grafike (žr. 3.11 pav.) vaizduojamos reitingo agentūrų ROC kreivės. Pastebime, jog atsitiktinių miškų modelis atliko geriausias prognozes visoms tiriamoms kreditingumo reitingo agentūroms. Raudona spalva žymima „Standard & Poor`s“ kreditingumo reitingo agentūros ROC kreivė. Jos AUC įvertis yra lygus 0.79, o standartinis nuokrypis: 0.02. Mėlyna spalva žymima „Fitch“ kreditingumo reitingo agentūros ROC kreivė, AUC įvertis siekia 0.84, o standartinis nuokrypis: 0.02. Žalia spalva žymima

„Moody`s“ kreditingumo reitingo agentūros ROC kreivė. Ši kreivė fiksuoja ir patį aukščiausią ROC įvertį: 0.92. Fiksuojamas standartinis nuokrypis yra lygus 0.01.



**3.11 pav.** Kreditingumo agentūrų prognozių ROC kreivė

Lyginant visų tiriamų agentūrų tarpusavio rezultatus nubraižyta ir Preciziškumo – Prisiminimo kreivė (žr. 3.12 pav.). Joje fiksuoti labai panašūs įverčiai, tačiau geriausiai buvo įvertintas „Moody`s“ agentūros atsitiktinių miškų modelis (žymima žalia spalva): AUC = 0.16 ir F1 = 0.22.



3.12 pav. Kreditingumo agentūrų prognozių PR kreivė

### 3.5.2. Klasifikavimo uždavinio modelių rezultatai

Atliekant klasifikavimo uždavinį kreditingumo reitingo sumažėjimas žymimas „-1“, padidėjimas „1“, nepasikeitimas „0“. Klasifikavimo uždaviniui spręsti buvo naudoti atsitiktinių miškų bei atramos vektorių modeliai. Visų modelių parametrų derinimui buvo išbandyti tinklės paieškos (angl. grid search), atsitiktinės paieškos (angl. random search), simuliuojamo atkaitinimo (angl. simulated annealing) bei Parzeno medžio (angl. Parzen tree) algoritmai. Modelių parametrų derinimui ir vertinimui naudotas *logloss* įvertis.

Atsitiktinių miškų modeliavimui pritaikytas *sklearn* bibliotekos metodas - „RandomForestClassifier“. Atramos vektorių modeliavimui taikytas „LinearSVC“ metodas, kuris veikia gerokai sparčiau už tradicinį „SVC“. Klasių disbalanso efekto mažinimui kiekvienam modeliui pritaikytas subalansuotų svorių parametras. Tokiu būdu visoms 3 prognozuojamoms klasėms priskirti svoriai, kurie buvo atvirkščiai proporcingi jų dažniui.

#### „Standard & Poor`s” agentūros duomenų rinkinio prognozių rezultatai

„Standard & Poor`s” agentūros duomenų rinkiniu pritaikius atsitiktinių miškų modelį buvo sudaryta sumaišymo matrica (žr. 3.13 pav.) Matricoje 1 stulpelyje ir 1 eilutėje vaizduojamas kreditingumo reitingo mažėjimas, 2 stulpelyje ir 2 eilutėje – kreditingumo reitingo nepasikeitimas ir 3 stulpelyje ir 3 eilutėje – kreditingumo reitingo didėjimas. Analizuodami pirmąją matricą matome, kad atsitiktinių miškų klasifikatorius beveik 60 % atvejų atpažįsta reitingo mažėjimą, 54.853 % atvejų sugeba nuspėti reitingo nepasikeitimą ir beveik 49 % atvejų atpažįsta kreditingumo reitingo didėjimą. Matome, jog klasifikatorius dažnai klysta bandydamas prognozuoti reitingo mažėjimo (tik 8.278 % atvejų kai prognozuojamas mažėjimas prognozė pasitvirtina) ir didėjimo (tik 10.221 % atvejų kai prognozuojamas didėjimas prognozė pasitvirtino) įvykius. Bendras klasifikatoriaus prognozių

tikslumas siekia tik 54.706 %. Todėl metodo prognozės gali būti vertinamos tik kaip nežymiai geresnės už atsitiktinį spėliojimą.

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Preciziškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	248	2609	139	2996	8.278%
	Reitingas nesikeičia	148	6494	188	6830	95.081%
	Reitingas didėja	22	2736	314	3072	10.221%
	Viso teisingai	418	11839	641	12898	
Atsiminimas (angl. Recall)		59.33%	54.853%	48.986%		
Bendras tikslumas (OA):		54.706%				
Kappa:		0.084				

3.13 pav. „Standard & Poor`s” reitingo pokyčio klasifikavimo sumaišymo matrica; atsitiktinis miškas

Duomenų rinkiniu pritaikius atramos vektorių modelį sudaryta kita sumaišymo matrica (žr. 3.14 pav.) Tirdami šią matricą pastebime, kad atramos vektorių klasifikatorius beveik visais atvejais prognozuoja, jog kreditingumo reitingas nesikeis. Todėl modelis fiksuoja beveik idealų reitingo nesikeitimo klasės prognozavimą (prisiminimą), siekiantį 99.992 % bei didelį bendrą tikslumą – 91.789 %. Tačiau klasifikatorius visiškai nesugeba atpažinti kreditingumo reitingo didėjimo (0.156 % teisingų prognozių) ir mažėjimo (0 % teisingų prognozių).

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Preciziškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	0	0	0	0	No data
	Reitingas nesikeičia	418	11838	640	12896	91.796%
	Reitingas didėja	0	1	1	2	50%
	Viso teisingai	418	11839	641	12898	
Atsiminimas (angl. Recall)		0%	99.992%	0.156%		
Bendras tikslumas (OA):		91.789%				
Kappa:		0.002				

3.14 pav. „Standard & Poor`s” duomenų rinkinio sumaišymo matrica; atramos vektorių modelis

Palyginamos atsitiktinių miškų bei atramos vektorių modelių *logloss* ir subalansuoto tikslumo metrikos (žr. 3.8 lentelę). Akivaizdu, kad neatsižvelgus į klasių balansą atramos vektorius atlieka tikslesnį klasifikavimą, tačiau jeigu siekiama atpažinti ir reitingo didėjimo bei mažėjimo įvykius labiau verta pasitikėti atsitiktinių miškų modeliu.

3.8 lentelė. „Standard & Poor`s” duomenų rinkinio klasifikavimo modelių palyginimas

Modelis	<i>Logloss</i> įverčio vidurkis	Subalansuoto tikslumo vidurkis
---------	---------------------------------	--------------------------------

Atsitiktinių miškų modelis	0.958	<b>0.546</b>
Atramos vektorių modelis	<b>0.332</b>	0.334

### „Fitch“ agentūros duomenų rinkinio prognozių rezultatai

„Fitch“ agentūros duomenų rinkinio prognozių rezultatams pavaizduoti sudaryta sumaišymo matrica (žr. 3.15 pav.) Matrica taikoma analizuoti atsitiktinių miškų modelio rezultatus. Iš karto pastebime, jog klasifikatoriaus prisiminimo įverčiai yra gerokai aukštesni, nei užfiksuoti tiriant Standard & Poor`'s agentūros rezultatus: reitingo mažėjimui fiksuotas 60.473% prisiminimas, nepasikeitimui: 65.502 % prisiminimas bei didėjimui: 64.259 % prisiminimas. Bendras tikslumas siekia santykinai aukštą procentą: 65.336 %.

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Preciziškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	179	1715	58	1952	9.17%
	Reitingas nesikeičia	114	7910	130	8154	97.008%
	Reitingas didėja	3	2451	338	2792	12.106%
	Viso teisingai	296	12076	526	12898	
Atsiminimas (angl. Recall)		60.473%	65.502%	64.259%		
Bendras tikslumas (OA):		65.336%				
Kappa:		0.124				

3.15 pav. „Fitch“ duomenų rinkinio sumaišymo matrica; atsitiktinis miškas

Duomenų rinkiniu pritaikius atramos vektorių modelį sudaryta kita sumaišymo matrica (žr. 3.16 pav.) Tirdami šią matricą pastebime, kad atramos vektorių klasifikatorius beveik visais atvejais prognozuoja, jog kreditingumo reitingas nesikeis. Tik 2 kartus buvo prognozuojamas reitingo kritimas, kai jis nesikeitė ir 7 kartus prognozuotas didėjimas, kai nebuvo jokių pokyčių. Dėl šių prižasčių modelis fiksuoja beveik idealią reitingo nesikeitimo prisiminimo metriką, siekiančią net 99.925 % bei didelį bendrą klasifikatoriaus tikslumą – 93.557 %. Tačiau klasifikatorius visiškai nesugeba atpažinti kreditingumo reitingo didėjimo (0 % teisingų prognozių) ir mažėjimo (taip pat 0 % teisingų prognozių).

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Precižiškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	0	2	0	2	0%
	Reitingas nesikeičia	296	12067	526	12889	93.622%
	Reitingas didėja	0	7	0	7	0%
	Viso teisingai	296	12076	526	12898	
	Atsiminimas (angl. Recall)	0%	99.925%	0%		
Bendras tikslumas (OA):	93.557%					
Kappa:	-0.001					

3.16 pav. „Fitch“ duomenų rinkinio sumaišymo matrica; atramos vektorių modelis

Palyginamos „Fitch“ agentūros prognozių atsitiktinių miškų bei atramos vektorių modelių *logloss* ir subalansuoto tikslumo metrikos (žr. 3.9 lentelę). Akivaizdu, kad neatsižvelgus į klasių balansą atramos vektorių atlieka tikslesnį klasifikavimą. Atramos vektorių modelis fiksuoja daug žemesnį *logloss* įvertį, nei atsitiktinių miškų modelis (0.267 prieš 0.844). Tačiau jeigu siekiama atpažinti ir reitingo didėjimo bei mažėjimo įvykius labiau verta pasitikėti atsitiktinių miškų modeliu. Atsitiktinio miško modelio subalansuotas tikslumas – 0.634, o atramos vektorių modelio – 0.333.

3.9 lentelė. „Fitch“ duomenų rinkinio klasifikavimo modelių palyginimas

Modelis	<i>Logloss</i> įverčio vidurkis	Subalansuoto tikslumo vidurkis
Atsitiktinių miškų modelis	0.844	<b>0.634</b>
Atramos vektorių modelis	<b>0.267</b>	0.333

### „Moody`s“ agentūros duomenų rinkinio prognozių rezultatai

„Moody`s“ agentūros duomenų rinkinio prognozių rezultatams pavaizduoti sudaryta sumaišymo matrica (žr. 3.17 pav.) Matome, jog „Fitch“ duomenų rinkinyje fiksuojama labai mažai kreditingumo reitingo didėjimo įvykių, tik 99. Būtent šio įvykio prisiminimo įvertis siekia vos 30.303 %. Tačiau gana tiksliai prognozuojama reitingo nepasikeitimo klasė (prisiminimo įvertis siekia 89.605 %) bei santykinai neblogai nustatomas reitingo mažėjimas (prisiminimo įvertis siekia 58.738 %). Klasifikatoriaus rezultatai prognozuojant „Moody`s“ agentūros reitingo pokyčius nuteikia labai optimistiškai, nes užfiksuojamas ir gana aukštas bendras klasifikatoriaus tikslumas – 88.657%.

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Preciziškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	121	937	32	1090	11.101%
	Reitingas nesikeičia	66	11284	37	11387	99.095%
	Reitingas didėja	19	372	30	421	7.126%
	Viso teisingai	206	12593	99	12898	
Atsiminimas (angl. Recall)		58.738%	89.605%	30.303%		
Bendras tikslumas (OA):		88.657%				
Kappa:		0.169				

3.17 pav. „Moody`s“ duomenų rinkinio sumaišymo matrica; atsitiktinis miškas

„Moody`s“ agentūros duomenų rinkiniui pritaikius atramos vektorių modelį sudaryta kita sumaišymo matrica (žr. 3.18 pav.) Klasifikatorius vėl beveik visada fiksuoja reitingo nepasikeitimo klasę. Todėl modelis fiksuoja beveik idealią reitingo nesikeitimo prisiminimo metriką, siekiančią 99.889 %, aukštą tikslumą – 97.655 % bei didelį bendrą klasifikatoriaus tikslumą – 97.535 %. Tačiau, kaip ir nagrinėjant kitų agentūrų prognozes, klasifikatorius visiškai nesugeba atpažinti kreditingumo reitingo didėjimo (0 % teisingų prognozių) ir mažėjimo (0.485 % teisingų prognozių).

		Teisingi duomenys				
		Reitingas mažėja	Reitingas nesikeičia	Reitingas didėja	Viso suklasifikuota	Preciziškumas
Klasifikatoriaus rezultatai	Reitingas mažėja	1	13	2	16	6.25%
	Reitingas nesikeičia	205	12579	97	12881	97.655%
	Reitingas didėja	0	1	0	1	0%
	Viso teisingai	206	12593	99	12898	
Atsiminimas (angl. Recall)		0.485%	99.889%	0%		
Bendras tikslumas (OA):		97.535%				
Kappa:		0.01				

3.18 pav. „Moody`s“ duomenų rinkinio sumaišymo matrica; atramos vektorių modelis

Palyginamos „Moody`s“ agentūros prognozių atsitiktinių miškų bei atramos vektorių modelių *logloss* ir subalansuoto tikslumo metrikos (žr. 3.10 lentelę). Akivaizdu, kad neatsižvelgus į klasių balansą atramos vektorius atlieka tikslesnį klasifikavimą. Atramos vektorių modelis fiksuoja žemesnį *logloss* įvertį, nei atsitiktinių miškų modelis (0.123 prieš 0.475). Tačiau jeigu siekiama atpažinti ir reitingo didėjimo bei mažėjimo įvykius labiau verta pasitikėti atsitiktinių miškų modeliu. Atsitiktinio miško modelio subalansuotas tikslumas – 0.604, o atramos vektorių modelio – 0.335.

3.10 lentelė. „Moody`s“ duomenų rinkinio klasifikavimo modelių palyginimas

Modelis	<i>Logloss</i> įvertio vidurkis	Subalansuoto tikslumo vidurkis
Atsitiktinių miškų modelis	0.475	<b>0.604</b>
Atramos vektorių modelis	<b>0.123</b>	0.335

### 3.5.3. Išlikimo analizės uždavinio modelių rezultatai ir palyginimas su detekcija

Rezultatuose pateikiami išlikimo miškų ir Cox modelių prognozavimo rezultatai. Kreditingumo reitingo išlikimas tyrime apibūdinamas kaip tikimybė, jog tiriamos agentūros reitingas nemažės per 1 metus nuo analizuojamų finansinių rodiklių pokyčių fiksavimo datos. Kreditingumo reitingo agentūrų išlikimas vertinamas naudojant išlikimo analizės uždavinio metrikas: *C – index* ir *Brier* balą.

Lentelėse (žr. 3.11 ir 3.12 lenteles) vaizduojami išlikimo analizės rezultatai taikant išlikimo miško bei Cox modelius. Cox modelyje fiksuojamos labai panašios *C – index* reikšmės (tarp 0.549 ir 0.58). Tuo tarpu išlikimo miškų modelio rezultatuose matoma, jog „Moody`s“ kreditingumo agentūros prognozės yra vertinamos kiek prasčiau (*C – index* lygus 0.651), nei „Standard & Poor`s“ (*C – index* lygus 0.74) ir „Fitch“ (*C – index* lygus 0.743) prognozės.

Analizuojant lentelėse pateikiamą *Brier* balą, matome, jog tiksliausios prognozės atliekamos tiriant „Standard & Poor`s“ kreditingumo reitingo agentūros reitingus. Tiek išlikimo miškų modelis, tiek ir Cox modelis agentūrai nustatė žemiausius *Brier* balo įverčius (0.024 ir 0.037). Tuo tarpu remiantis *Brier* balo įverčiu prasčiausiai vertinamos „Moody`s“ kreditingumo reitingo agentūros prognozės. Išlikimo miškų modelis fiksuoja įvertį lygų 0.058 ir Cox modelis fiksuoja įvertį lygų 0.094.

3.11 lentelė. Išlikimo miško modelio įverčiai

Kreditingumo reitingo agentūra	C – index	Brier balas
„Standard & Poor`s“	0.74	<b>0.024</b>
„Fitch“	<b>0.743</b>	<b>0.024</b>
„Moody`s“	0.651	0.058

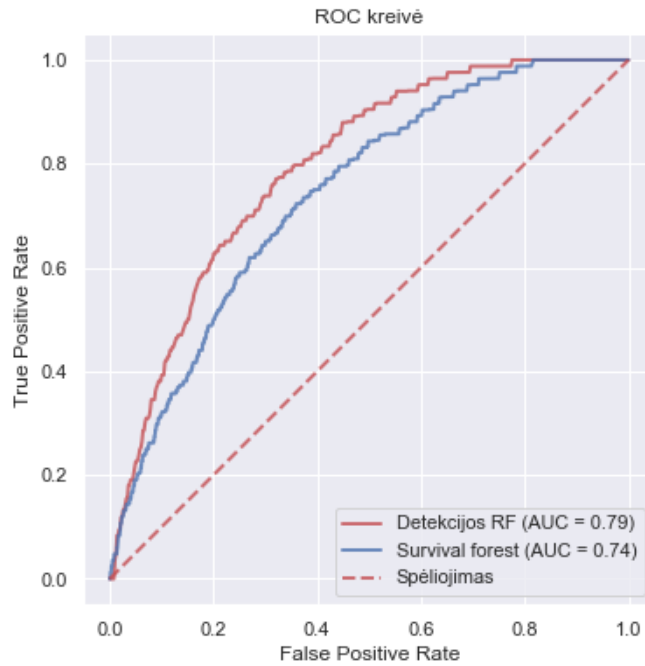
3.12 lentelė. Cox modelio įverčiai

Kreditingumo reitingo agentūra	C – index	Brier balas
„Standard & Poor`s“	0.549	<b>0.037</b>
„Fitch“	0.566	0.084
„Moody`s“	<b>0.58</b>	0.094

Taip pat atliktas išlikimo analizės ir detekcijos uždaviniuose gautų rezultatų palyginimas. Pagal apskaičiuotą vidutinę išlikimo tikimybę kiekvienai analizuojamai kreditingumo agentūrai nubrėžta ROC kreivė. Kreivė yra lyginama su atitinkamos agentūros ROC kreive. Lyginami tiksliausi modeliai: detekcijos – atsitiktiniai miškai, išlikimo analizės – išlikimo miškai.

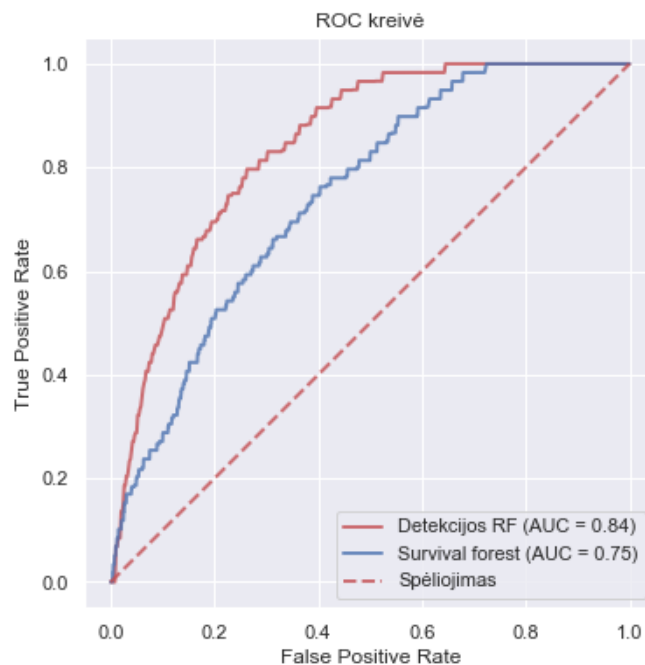
Paveikslėlyje (žr. 3.19 pav.) raudona spalva vaizduojama „Standard & Poor`s“ agentūros detekcijos uždavinio atsitiktinių miškų ROC kreivė. Mėlyna spalva vaizduojama ROC kreivė, gauta taikant išlikimo miškus ir skaičiuojant vidutines išlikimo tikimybes. Matome, jog detekcijos (ROC AUC – 0.79) įvertis yra šiek tiek aukštesnis už išlikimo analizės (ROC AUC – 0.74).





**3.19 pav.** „Standard & Poor`s” detekcijos ir išlikimo analizės palyginimas

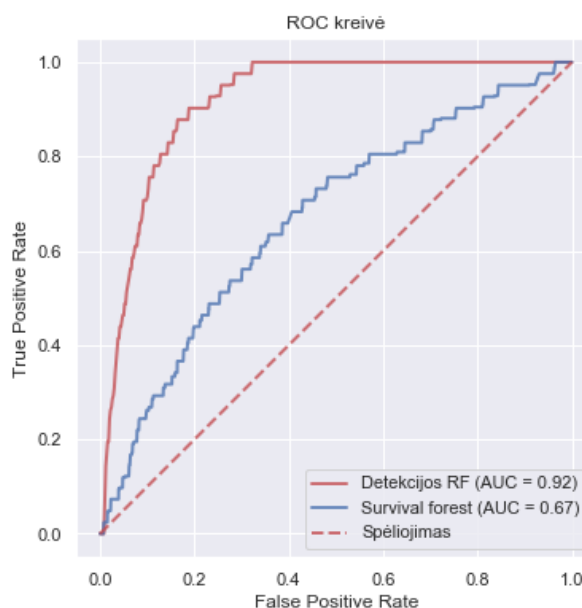
Kitame paveikslėlyje (žr. 3.20 pav.) raudona spalva vaizduojama „Fitch“ agentūros detekcijos uždavinio atsitiktinių miškų ROC kreivė. Mėlyna spalva vaizduojama ROC kreivė, gauta taikant išlikimo miškus ir skaičiuojant vidutines išlikimo tikimybes. Taip pat ir šiuo atveju detekcijos (ROC AUC – 0.84) įvertis yra šiek tiek aukštesnis už išlikimo analizės (ROC AUC – 0.75).



**3.20 pav.** „Fitch” detekcijos ir išlikimo analizės palyginimas

Paveikslėlyje (žr. 3.21 pav.) raudona spalva vaizduojama „Moody`s“ agentūros detekcijos uždavinio atsitiktinių miškų ROC kreivė. Mėlyna spalva vaizduojama ROC kreivė, gauta taikant

išlikimo miškus ir skaičiuojant vidutines išlikimo tikimybes. Detekcijos analizė pateikė daug geresnius įverčius, nei išlikimo analizė. Detekcijos ROC AUC: 0.92, išlikimo analizės: 0.67.

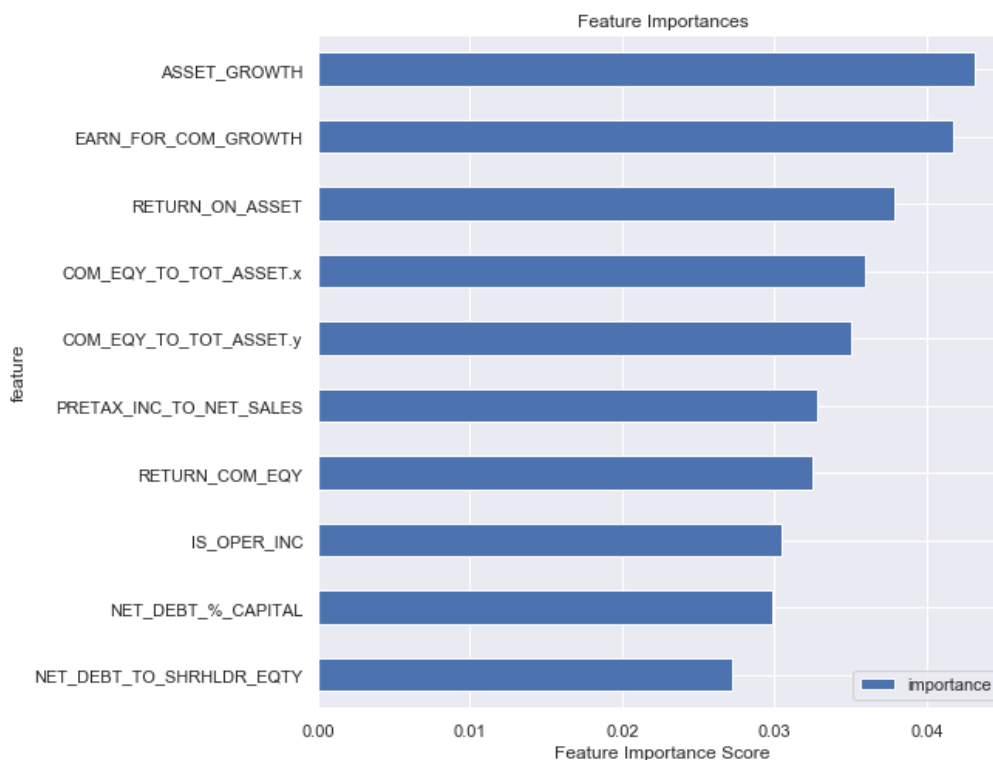


3.21 pav. „Moody`s” detekcijos ir išlikimo analizės palyginimas

### 3.6. Naudojamų finansinių rodiklių svarbos analizė

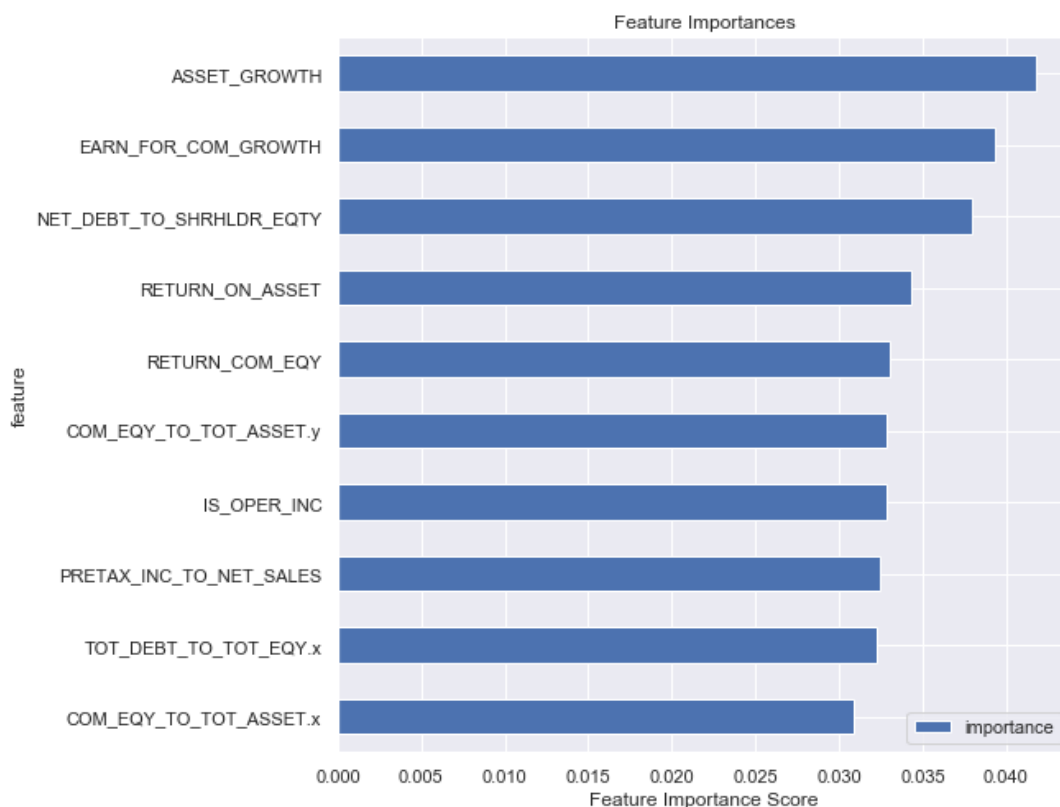
Atlikta tyrime naudojamų finansinių rodiklių svarbos analizė. Svarbos analizė pritaikyta tiksliausias prognozes tyrime pateikusiam modeliui – atsitiktiniams miškams. Finansinių rodiklių svarba buvo nustatyta analizuojant visų „Didžiojo trejeto“ kreditingumo reitingų agentūrų duomenis. Pateikiamuose grafikuose vaizduojami 10 svarbiausių finansinių rodiklių, kuriais remiantis buvo sudarytos klasifikavimo uždavinio prognozės.

Paveikslėlyje (žr. 3.22 pav.) vaizduojami 10 svarbiausių finansinių rodiklių sudarant „Standard & Poor`s“ kreditingumo reitingo agentūros prognozes. Matome, jog didžiausias koeficientas (0.043) buvo priskirtas „ASSET\_GROWTH“ kintamajam. Šis kintamasis parodo banko turto augimą per pastaruosius 1 metus. Antras svarbiausias kintamasis (svarbos koeficiento vertė 0.041) – „EARN\_FOR\_COM\_GROWTH“ (banko pelno augimas). Trečias svarbiausias kintamasis – „RETURN\_ON\_ASSET“ (svarbos koeficiento vertė 0.038).



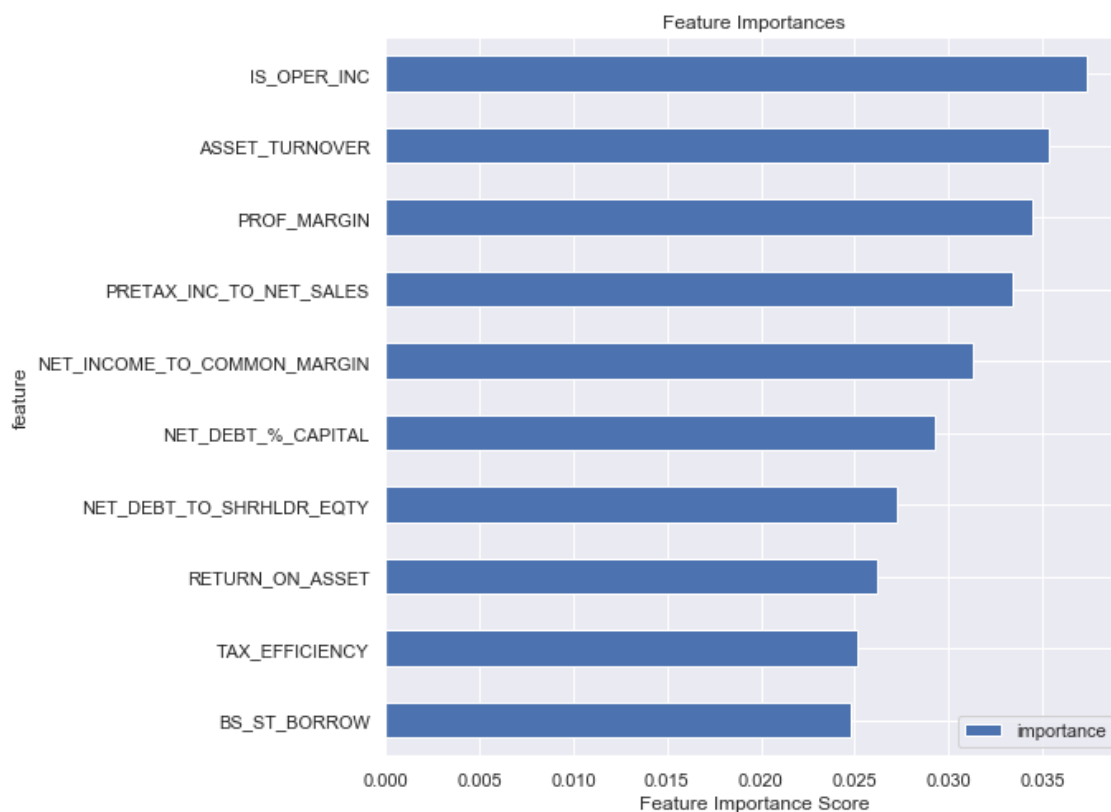
**3.22 pav.** Finansinių rodiklių svarba sudarant „Standard & Poor`s” prognozes

Paveikslėlyje (žr. 3.23 pav.) vaizduojami 10 svarbiausių finansinių rodiklių sudarant „Fitch kreditingumo reitingo agentūros prognozes. Vėl didžiausias koeficientas (vertė 0.042) buvo priskirtas „ASSET\_GROWTH“ kintamajam. Antras svarbiausias kintamasis buvo „EARN\_FOR\_COM\_GROWTH“ (svarbos koeficiento vertė 0.039). Trečias svarbiausias kintamasis – „NET\_DEBT\_TO\_SHRHLDR\_EQTY“ (banko finansinio svėro matas), koeficiento vertė: 0.038.



**3.23 pav.** Finansinių rodiklių svarba sudarant „Fitch” prognozes

Paveikslėlyje (žr. 3.24 pav.) vaizduojami 10 svarbiausių finansinių rodiklių sudarant „Moody`s“ kreditingumo reitingo agentūros prognozes. Aukščiausias koeficientas buvo priskirtas „IS\_OPER\_INC“ kintamajam, kuris aprašo bendras veiklos pajamas atėmus veiklos sąnaudas. Kintamojo svarbos koeficientas: 0.037. Antras svarbiausias kintamasis – „ASSET\_TURNOVER“ (pajamų suma gauta už turto dolerį). Kintamojo svarbos koeficientas: 0.035. Kintamasis rodo kaip efektyviai bankas disponuoja savo turtu. Trečias svarbiausias kintamasis: „PROF\_MARGIN“ (svarbos koeficientas: 0.034). Kintamasis matuoja bankų pelningumą: pelno ir pajamų santykį.



3.24 pav. Finansinių rodiklių svarba sudarant „Moody`s” prognozes

### 3.7. Tyrimo rezultatų apibendrinimas

Apibendrinat atlikto tyrimo rezultatus pateikiami tiksliausias prognozes atlikę detekcijos, klasifikavimo ir išlikimo analizės metodai. Taip pat aprašoma kaip tiksliai buvo prognozuojami kiekvienos tiriamos kreditingumo reitingo agentūros reitingų pokyčiai.

Detekcijos uždavinyje naudoti logistinės regresijos, atsitiktinių miškų bei atramos vektorių modeliai. Patikimiausiu laikomas atsitiktinių miškų modelis. Taikant šį modelį pasiekti geriausi ROC AUC ir subalansuoto tikslumo įverčiai visoms 3 kreditingumo reitingo agentūroms. Detekcijos uždavinyje tiksliausias prognozes pavyko atlikti su „Moody`s“ kreditingumo reitingo agentūros duomenimis: pasiektas 0.92 ROC AUC įvertis. Tiriant „Fitch“ kreditingumo reitingo agentūros reitingo pokyčius pasiektas šiek tiek prastesnis įvertis: 0.84 ROC AUC. Tuo tarpu tiriant „Standard & Poor`s” kreditingumo reitingo agentūros pokyčius užfiksuotas žemiausias 0.78 ROC AUC įvertis.

Klasifikavimo uždavinyje kreditingumo reitingo agentūrų reitingo pokyčiai buvo lyginti pasitelkiant tik atsitiktinių miškų ir atramos vektorių modelius. Pastebėta, jog atramos vektorių modelis beveik visą laiką prognozuoja tik teigiamą klasę ir todėl juo negalima pasitikėti. Tuo tarpu atsitiktinių miškų modelis geba nustatyti kreditingumo reitingo mažėjimą, nepasikeitimą ir didėjimą. Prasčiausiai atsitiktinių miškų modelis prognozavo „Standard & Poor`s” kreditingumo agentūros reitingo pasikeitimus. Bendras klasifikatoriaus prognozių tikslumas siekė tik 54.706 %. Pasiiektas kappa įvertis: 0.084. Tai reiškia, jog klasifikatoriaus prognozės buvo tik truputį tikslesnės už atsitiktinį spėliojimą. Klasifikatorius tiksliau prognozavo „Fitch“ kreditingumo reitingo agentūros reitingų pokyčius: užfiksuotas 65.336 % bendras tikslumas. Pasiiektas kappa įvertis: 0.124. Naudojant „Fitch“ agentūros pokyčių prognozes, visoms 3 prognozuojamoms klasėms fiksuoti aukšti

prisiminimo įverčiai. Reitingo mažėjimo klasei fiksuotas 60.473 % prisiminimo įvertis, reitingo nesikeitimo klasei – 65.502 % ir didėjimo klasei – 64.259 %. „Moody`s“ agentūros reitingo pokyčių prognozės taikant atsitiktinių miškų modelį fiksavo aukštą bendrą tikslumą – 88.657 %. Pasiektas kappa įvertis: 0.169. Taip pat nustatyta ir aukšta reitingo nesikeitimo klasės prisiminimo metrika: 89.605 %. Tačiau „Moody`s“ duomenyse modeliui sunkiai sekėsi atpažinti kreditingumo reitingo didėjimą, prisiminimo metrika siekė vos 30.303 %.

Išlikimo analizės uždavinyje agentūrų sudarytų reitingų pokyčiai buvo tiriami naudojant išlikimo miškų bei Cox modelius. Remiantis išlikimo miškų modeliu, geriausi C – index bei Brier balo įverčiai fiksuoti „Standard & Poor`s“ ir „Fitch“ kreditingumo reitingo agentūrų rezultatuose. Taikant Cox modelį geriausias C – index (0.58) įvertis buvo fiksuojamas „Moody`s“ agentūrai, o mažiausias Brier balas (0.037) – „Standard & Poor`s“ kreditingumo agentūrai. Lyginant detekcijos ir išlikimo analizės uždavinių rezultatus pastebėti aukštesni detekcijos ROC AUC įverčiai.

#### 4. Darbo rezultatai ir išvados

Atlikus baigiamąjį darbą gauti rezultatai:

1. Iširtas kreditingumo reitingo taikymo industrijoje procesas.
2. Įvertinti kreditingumo reitingo prognozavimui industrijoje naudojamų mašininio mokymo modelių privalumai ir trūkumai.
3. Apžvelgti svarbiausi mašininio mokymo parametrų derinimo būdai: tinklelio paieška, atsitiktinė paieška, Bajesinis optimizavimas. Apibūdintos jų stipriosios bei silpnosios pusės.
4. Įvykdytas kreditingumo reitingo agentūrų reitingų ir tiriamų bankų finansinių rodiklių importas iš „Bloomberg“ sistemos.
5. Atliktas kreditingumo reitingų pokyčių krypties prognozavimas „Standard & Poor`s“, „Fitch“ ir „Moody`s“ kreditingumo reitingo agentūroms. Sudarytos atskiros prognozės sprendžiant detekcijos, klasifikavimo ir išlikimo analizės uždavinius.
6. Palygintas naudojamų mašininio mokymo modelių efektyvumas prognozuojant bankų reitingų pokyčius. Nustatyta, jog tiksliausias prognozės atliekamos taikant atsitiktinių miškų modelį. Išlikimo analizės uždavinyje išlikimo miškų modelio prognozės buvo tikslesnės už Cox modelio prognozes. Skirtingų agentūrų išlikimo prognozių rezultatai buvo beveik identiški.
7. Tiksliausiai pavyko prognozuoti „Moody`s“ kreditingumo reitingo agentūros reitingo pokyčius. Taikant atsitiktinių miškų modelį pasiektas 0.92 ROC AUC įvertis. Šiek tiek prasčiau sekėsi prognozuoti „Fitch“ kreditingumo reitingo agentūros rezultatus: pasiektas 0.84 ROC AUC įvertis. Prasčiausiai buvo prognozuojami „Standard & Poor`s“ kreditingumo reitingo agentūros rezultatai: nustatytas 0.78 ROC AUC įvertis.

Atliktas darbas leidžia padaryti išvadas:

1. Sudarant įmonių kreditingumo vertinimo procesą privalo būti atsižvelgiama į dabartinę kreditingumo vertinimo veiklą, disponuojamus žmogiškuosius išteklius, agentūros informacinių sistemų išsivystymo lygį bei patirtį išduodant paskolas.
2. Atliekant tyrimą nustatyta, jog Bajesinio optimizavimo parametrų derinimo metodikos (simuliuojamas atkaitinimas bei Parzeno medis) yra daug efektyvesnės už plačiai naudojamus tinklelio paieškos bei atsitiktinės paieškos būdus. Taikant Bajesinio optimizavimo algoritmus išsamiai išanalizuojamos platesnės derinamų parametrų erdvės.
3. Duomenyse fiksuojamas klasių disbalansas daro labai didelę įtaką mašininio mokymo modelių prognozėms. Net ir tiriamiesiems stebiniams pritaikius svorius, atramos vektorių modeliai neišmoko patikimai atskirti skirtingų klasių.
4. Jei tyrime galima nustatyti ir įvykio laiką, išlikimo analizės algoritmai gali būti naudinga alternatyva klasifikavimo ir detekcijos uždaviniams. Tokiu atveju mašininio mokymo algoritmai išnaudoja papildomą fiksavimo laiko informaciją ir todėl galima tikėtis tikslesnių rezultatų.
5. Nustatyta, kad naudojant realių finansinių rodiklių pokyčių duomenis, tiksliausiai pavyko prognozuoti „Moody`s“ kreditingumo agentūros sudarytus bankų reitingų pokyčius. Prasčiausias prognozės sudarytos „Standard & Poor`s“ kreditingumo agentūros reitingo pokyčiams. Todėl galima argumentuoti, jog, lyginant visas „Didžiojo trejeto“ agentūras, „Moody`s“ labiausiai atsižvelgia į realius finansinių rodiklių pokyčius. Tuo tarpu „Standard & Poor`s“ kreditingumo reitingo agentūra, sudarydama kreditingumo reitingus, galimai remiasi ir su finansiniais rodikliais nesusijusia informacija.

## 5. Literatūros sąrašas

1. Ala`raj, M., & Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Elsevier*, 64(1), 36-38.
2. Baesens, B. (2003 m. 05 14 d.). Developing intelligent systems for credit scoring using machine learning techniques. 247(1), p. 124-128.
3. Brown, I., & Mues, C. (2012 m. 02 15 d.). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Elsevier*, 39(3), 3446-3453. doi:10.1016/j.eswa.2011.09.033
4. Brownlee, J. (2019). *Probabilistic Model Selection with AIC, BIC, and MDL*. Nuskaityta iš Machine Learning Mastery: <https://machinelearningmastery.com/probabilistic-model-selection-measures/>
5. Caire, D., & Kossman, R. (2003). *Credit Scoring: Is it Right for Your Bank?* Nuskaityta iš Microfinance: A Way to Help the Poor Build Assets: [http://www.microfinance.com/English/Papers/Scoring\\_For\\_Your\\_Bank.pdf](http://www.microfinance.com/English/Papers/Scoring_For_Your_Bank.pdf)
6. Caire, D., Barton, S., Zubiria, d. A., Alexiev, Z., Dyer, J., Bundred, F., & Brislin, N. (2006). *A HANDBOOK FOR DEVELOPING CREDIT SCORING SYSTEMS IN A MICROFINANCE CONTEXT*. Washington: DAI Washington.
7. Chen, L. (2019 m. 01 7 d.). Nuskaityta iš Towards Data Science: <https://towardsdatascience.com/support-vector-machine-simply-explained-fee28eba5496>
8. Davidson-Pilon, C., Kalderstam, J., Jacobson, N., Zivich, P., Kuhn, B., Williamson, M., . . . Anton, S. (2020). doi:10.5281/zenodo.3833188
9. Doumpos, M., & Zopounidis, C. (2014). Credit Scoring. Esantis M. Doumpos, & C. Zopounidis, *Credit Scoring. In: Multicriteria Analysis in Finance. SpringerBriefs in Operations Research* (p. 43-59). Springer, Cham. doi:10.1007/978-3-319-05864-1\_4
10. Fernandez, J. L., & Vila, S. R. (2015). *The Credit Rating Agencies, Institutional*. Barcelona: Universitat Autònoma de Barcelona.
11. Fotso, S. (2019). Nuskaityta iš {PySurvival}: Open source package for Survival Analysis modeling: <https://square.github.io/pysurvival/intro.html>
12. Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*, 1(4), 274-278. doi:10.4103/0974-7788.76794
13. Hari, S. S. (2018). *Hyperparameter Tuning using Simulated Annealing*. Nuskaityta iš Hyperparameter Tuning using Simulated Annealing: [https://santhoshhari.github.io/simulated\\_annealing/](https://santhoshhari.github.io/simulated_annealing/)
14. Hernandez-Orallo, J., Flach, P., & Ferri, C. (2012). A Unified View of Performance Metrics: Translating Threshold. *Journal of Machine Learning Research*, 13, 2813-2869.
15. <https://corporatefinanceinstitute.com/resources/knowledge/finance/credit-rating/>. (2020). Nuskaityta iš <https://corporatefinanceinstitute.com>.
16. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). RANDOM SURVIVAL FORESTS. *The Annals of Applied Statistics*, 2(3), 841-860. doi:10.1214/08-AOAS169
17. Kopczyk, D. (2018). *Hyperparameter optimization: Explanation of automatized algorithms*. Nuskaityta iš <http://dkopczyk.quantee.co.uk/hyperparameter-optimization/>



18. Lessmann, S., Baesens, B., Seowd, H.-V., & Thomasc, L. C. (2015 m. November 16 d.). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Elsevier*, 247(1), 124-136.
19. (2015). *MICROFINANCE CREDIT RISK MANAGEMENT TOOL GUIDE*. J.P.Morgan.
20. Migueis, V., Benoit, D., & Poel, D. V. (2012). Enhanced decision support in credit scoring using Bayesian binary quantile regression. *Journal of the Operational Research Society*, 64(1), 1374–1383.
21. Minelga, J., Gelžinis, A., Vaičiukynas, E., Verikas, A., Bačauskien, M., Padervinskis, E., & Uloza, V. (2014). Comparing throat and acoustic microphones for laryngeal pathology detection from human voice. *9th International Conference on Electrical and Control Technologies*, (p. 50-53).
22. Narkhede, S. (2018). *Understanding AUC - ROC Curve*. Nuskaityta iš <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
23. Saia, R., & Carta, S. (2016). *An Entropy Based Algorithm for Credit Scoring*. Università di Cagliari, Dipartimento di Matematica e Informatica. Cagliari: Springer International Publishing AG . doi:10.1007/978-3-319-49944-4\_20
24. Schreiner, M. (2003 m. 01). *SCORING: THE NEXT BREAKTHROUGH*. Nuskaityta iš CGAP: <https://www.cgap.org/research/publication/scoring-next-breakthrough-microcredit>
25. Thomas, L., Crook, J., & Edelman, D. (2002). *Credit Scoring and Its Applications, Second Edition* (T. 2). doi:10.1137/1.9781611974560
26. Tuffery, S. (2011). Data Mining and Statistics for Decision Making. Esantis S. Tuffery, *Data Mining and Statistics for Decision Making* (1st leid., p. 556-562). John Wiley & Sons, Ltd.
27. Zararsiz, G., & Cosgun, E. (2014). Introduction to Statistical Methods for MicroRNA Analysis. *Methods in molecular biology*, p. 129-55. doi:10.1007/978-1-62703-748-8\_8

## 6. Priedai

### 1 priedas. Tyrimo naudojamų finansinių rodiklių aprašas (anglų kalba)

#### Return on Common Equity

Measure of a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested, in percentage. Calculated as:

$$(\text{T12 Net Income Available for Common Shareholders} / \text{Average Total Common Equity}) * 100$$

#### Total Debt to Total Equity

Total debt divided by total shareholders' equity. Calculated as:

$$\text{Short and Long Term Debt} / \text{Shareholders' Equity} * 100$$

Where:

Short and Long Term Debt is RR251, SHORT\_AND\_LONG\_TERM\_DEBT

Shareholders' Equity is RR007, TOTAL\_EQUITY

#### Altman's Z-Score

Indicates the probability of a company filing for bankruptcy within the next two years. The higher the value, the lower the probability of bankruptcy. A score below 1.8 indicates bankruptcy is imminent. A score above 3 indicates bankruptcy is unlikely. Altman's Z-Score is only available on publicly listed companies with all the requisite fundamentals for the model. Calculated as:

$$\text{Altman's Z-Score} = 1.2 * (\text{Working Capital} / \text{Tangible Assets}) + 1.4 * (\text{Retained Earnings} / \text{Tangible Assets}) + 3.3 * (\text{EBIT} / \text{Tangible Assets}) + 0.6 * (\text{Market Value of Equity} / \text{Total Liabilities}) + (\text{Sales} / \text{Tangible Assets})$$

#### Current Ratio

Ratio to indicate the company's ability to pay back its short-term liabilities with its short-term assets. Unit: Actual.

Calculated as: Current Assets / Current Liabilities

#### Short Term Debt

Includes bank overdrafts, short-term debts and borrowings, repurchase agreements (repos) and reverse repos, short-term portion of long-term borrowings, current obligations under capital (finance) leases, current portion of hire purchase creditors, short term operating lease liabilities after

the adoption of International Financial Reporting Standards (IFRS) 16 and Accounting Standards Codification (ASC) 842, trust receipts, bills payable, bills of exchange, bankers acceptances, interest bearing loans, and short term mandatory redeemable preferred stock. Net with unamortized premium or discount on debt and may include fair value adjustments of embedded derivatives.

For banks and financials, includes call money, bills discounted, federal funds purchased, and due to other banks or financial institutions.

### **Long Term Debt**

All interest-bearing financial obligations that are not due within a year.

Includes convertible, redeemable, retractable debentures, bonds, loans, mortgage debts, sinking funds, and long-term bank overdrafts.

Excludes short-term portion of long term debt, pension obligations, deferred tax liabilities and preferred equity.

Includes subordinated capital notes.

Includes long term hire purchase and finance lease obligations.

Includes long term operating lease liabilities after the adoption of Financial Reporting Standards (IFRS) 16 and Accounting Standards Codification (ASC) 842

Includes long term bills of exchange and bankers acceptances.

May include shares issued by subsidiaries if the group has an obligation to transfer economic benefits in connection with these shares.

Includes mandatory redeemable preferred and trust preferred securities in accordance with FASB 150 effective June 2003.

Includes other debt which is interest bearing.

Net with unamortized premium or discount on debt.

May include fair value adjustments of embedded derivatives.

### **Total Debt to Earnings before Interest Taxes**

Ratio of total debt to earnings before interest and taxes (EBIT). A low number indicates that the company can service its debt from current earnings, units in actual. Calculated as:

Total Debt / Trailing 12 Month EBIT

### **Return on Assets**

Indicator of how profitable a company is relative to its total assets, in percentage. Return on assets gives an idea as to how efficient management is at using its assets to generate earnings.

Calculated as:

$(\text{Trailing 12M Net Income} / \text{Average Total Assets}) * 100$

### **Return on Capital**

Metric that measures the return that an investment generates for capital contributors, in percentage. It indicates how effective a company is turning capital into profits.

Calculated as:

$$((T12 \text{ Net Income (Losses)} + T12 \text{ Minority Interest}) / \text{Average of Total Capital}) * 100$$

### **Gross Margin**

Gross margin represents the percent of total sales revenue that the company retains after incurring the direct costs associated with producing the goods and services sold by a company. Calculated as:

$$(\text{Net Sales} - \text{Cost of Goods Sold}) * 100 / \text{Net Sales}$$

### **EBITDA Margin**

Measure, in percentage, calculates the relation of Earnings Before Interest, Taxes, Depreciation and Amortization to Revenue. Calculated as:

$$(\text{EBITDA} / \text{Revenue}) * 100$$

### **Op Inc to Int Expense**

Operating income as a percentage of interest expense. Calculated as:

$$\text{Operating Income} / \text{Interest Expense}$$

### **Pre-Tax Income to Net Sales**

Pre-tax income as a percentage of net sales. Calculated as:

$$(\text{Pre-Tax Income} / \text{Revenue}) * 100$$

### **Profit Margin**

Measuring the company's profitability, this ratio is the comparison of how much of the revenue incurred during the period was retained in income. Calculated as:

$$(\text{Net Income} / \text{Net Revenue}) * 100$$

### **Net Income to Common Margin**

Ratio of net income available to common shareholders to total revenue, expressed in percentage. This alternative calculation of profit margin makes the ratio more relevant to common shareholders.

$(\text{Net Income Available To Common Shareholders} / \text{Net Revenue}) * 100$

### **Effective Tax Rate**

A measure of total tax paid as a percentage of the company's accounting income. Calculated as:

$\text{Income Tax Expenses} * 100 / \text{Pretax Income}$

### **Revenue Growth Year over Year**

A percentage increase or decrease of sales revenue by comparing current period with same period prior year. Calculated as:

$(\text{Revenue from Current Period} - \text{Revenue from Same Period Prior Year}) * 100 / \text{Revenue from Same Period Prior Year}$

### **Net Income to Common Growth YoY**

Percentage growth in earnings to common shareholders from the prior to the current year. Calculated as:

$[(\text{current year Net Income Available to Common} / \text{prior year Net Income Available to Common}) - 1] * 100$

### **Diluted EPS before XO Items - 1 Yr Growth**

Percentage increase or decrease of Diluted EPS before XO Items by comparing current period with same period prior year. Calculated as:

$(\text{Diluted EPS before XO Items} - \text{Diluted EPS before XO Items from Same Period Prior Year}) * 100 / \text{Diluted EPS before XO Items from Same Period Prior Year}$

### **Accounts Receivable 1 Yr Growth**

Percentage growth in inventory from last year to the current year. Calculated as:

$[(\text{current year Accounts \& Notes Receivable} / \text{prior year Accounts \& Notes Receivable}) - 1] * 100$

### **Assets - 1 Year Growth**

A percentage increase or decrease of total assets by comparing current period with same period prior year. Calculated as:

$(\text{Total Assets} - \text{Total Assets Same Period Prior Year}) * 100 / \text{Total Assets from Same Period Prior Year}$

### **Modified Working Capital - 1 Yr Growth**

Percentage growth in modified working capital from last year to the current year. Calculated as:

$[(\text{Current Year Modified Working Capital} / \text{Prior Year Modified Working Capital}) - 1] * 100$

### **Employees - 1 Year Growth**

Percentage increase or decrease of employee number by comparing the current period with the same period in the prior year. Calculated as:

$(\text{Number of Employees} - \text{Number of Employees Same Period Prior Year}) * 100 / \text{Number of Employees from Same Period Prior Year}$

### **Interest Expense**

Net of Capitalized Interest Expenses.

Cost incurred by an entity for borrowed funds. Interest expense is a non-operating expense shown on the income statement. It represents interest payable on any type of borrowings - bonds, loans, convertible debt, capital leases, lines of credit, and etc.

Includes amortization of debt discount or premium, debt issuance expenses, and factoring expenses.

Includes preferred dividends on mandatory redeemable preferred and trust preferred securities in accordance with FASB 150 effective June 2003.

Excludes interest cost from defined benefit pension plan if it is disclosed separately.

May include other finance charges if it is not disclosed separately.

May include loss/gain on inflation if it is not disclosed separately.

### **Common Equity to Tot Assets**

One of many financial ratios (in Percentage) used to determine the financial health and long-term profitability of a corporation. Calculated as:

Common Equity \* 100 / Total Assets

### **Total Debt to Total Assets**

Leverage ratio in percentage that defines the total amount of debt relative to assets. This enables comparisons of leverage to be made across different companies. Calculated as:

Total Debt \*100 / Total Assets

### **Net Debt To Shareholders Equity**

Measure of a company's book leverage. Unit: Actual.  
Calculated as:

$(\text{Net Debt} / \text{Total Equity}) * 100$

### **Net Debt Percent of Total Capital**

Metric used to measure the percentage of debt a company uses to obtain funds for its operations. This measure is an important indicator of a company's ability to obtain external financing. Calculated as:  
 $(\text{Net Debt} / (\text{Total Capital} - \text{Cash \& Marketable Securities} - \text{Collaterals for Cash \& Marketable Securities})) * 100$

### **EBITDA**

Indicator of a company's financial performance which is essentially net income with interest, taxes, depreciation, and amortization added back to it, and can be used to analyze and compare profitability between companies and industries because it eliminates the effects of financing and accounting decisions. Figure is reported in millions; the Scaling Format Override (DY339, SCALING\_FORMAT) can be used to change the display units for the field.

Calculated as:

Operating Income + Depreciation & Amortization + Operating Lease Rental Expense Adjustment

### **EBITDA after CAPEX**

Measure subtracts the Capital Expenditure from Earnings before Interest, Tax, Depreciation & Amortization. Calculated as:

EBITDA - Capital Expenditure

### **CFO to Average Current Liabilities**

Cash from operations (CFO) as a multiple of average current liabilities. A high multiple indicates the company can finance its liabilities without incurring additional debt, units in actual. Calculated as:

Trailing 12 Month Cash from Operations / (Current Liabilities + one year ago Current Liabilities) / 2

### **Cashflow to Total Liabilities**

Ratio of a company's cash flow from operations to its total liabilities, in percentage. The ratio is used to determine how long it would take a company to repay its liabilities if it devoted all of its cash flow to debt repayment. Calculated as:

(T12 Cash Flow from Operations/Total Liabilities)\*100

### **Capital Expenditure Ratio**

Measures how much of the cash generated from operations will be left after payment of capital expenditures to service the company's debt. Unit: Actual.

Calculated as:

Cash From Operations / Capital Expenditures

### **Accounts Receivable Turnover**

Ratio that measures how many times a business can collect its average accounts receivable during the year. Unit: Actual. Calculated as:

Trailing 12 Month Sales / Average Account Receivable

### **Days Sales Outstanding – DSO**

Measure of the average number of days that a company takes to collect revenue after a sale has been made. Unit: Actual.

Calculated as:

Number of Days in the Period / Accounts Receivable Turnover

### **Inventory Turnover**



Ratio showing how many times a company's inventory is sold and replaced over a period. Unit: Actual. Calculated as:

Trailing 12 Month Cost of Goods Sold or Trailing 12 month Cost of Materials / Average Inventory

### **Inventory Days**

Measures how many times a company's inventory is sold and replaced over a period. Unit: Actual. Calculated as:

Number of Days in the Period / Inventory Turnover

### **Accounts Payable Turnover**

Company purchases over average accounts payable. Calculated as:

$(\text{Ending Inventory} + \text{Cost of Goods Sold} - \text{Beginning Inventory}) / \text{Average Accounts Payable}$

### **Cash Conversion Cycle**

Metric which expresses the length of time, in days, that it takes for a company to convert resource inputs into cash flows. Units: Actual. Calculated as:

Inventory Turnover Days + Account Receivable Turnover Days - Accounts Payable Turnover Days

### **Inventory to Cash - Days**

Ratio is use to measure the length of time (in days) that a company uses to sell inventory, collect receivables into cash. Calculated as:

Days Sales Outstanding - Inventory Days

### **Inventories**

Includes completed but not billed on long-term contracts for business type activities.

### **Inventory Raw Materials**

Includes items such as supplies that are consumed in the operations. It also includes purchased components that are not materially altered at the company level but are merely added to the finished product.

### **Inventory In Progress**

Includes items that are not yet completed and require additional inputs to be completed, finished products. Includes unbilled revenue, except for Russell 3000 companies in the United States.

### **Inventory Finished Goods**

Inventory Finished Goods:

Generally only applies to manufacturing companies.

Includes items that are completed and ready to be delivered/consumed to the next/end user.

### **Trailing 12M Cash From Operations**

Calculated by adding Cash From Operations from the statement of cash flows (CF015, CF\_CASH\_FROM\_OPER) for the last four quarters, two semi annuals, or annual. Figure is reported in millions; the Scaling Format Override (DY339, SCALING\_FORMAT) can be used to change the display units for the field.

### **Trailing 12M Capital Expenditures**

Calculated by adding Capital Expenditures from the statement of cash flows (RR014, CAPITAL\_EXPEND) for the last four quarters, two semi annuals, or annual. Figure is reported in millions; the Scaling Format Override (DY339, SCALING\_FORMAT) can be used to change the display units for the field.

### **Trailing 12M Free Cash Flow**

Calculated by adding Free Cash Flow (RR008, CF\_FREE\_CASH\_FLOW) for the last four quarters, two semi-annuals, or annual. Figure is reported in millions; the Scaling Format Override (DY339, SCALING\_FORMAT) can be used to change the display units for the field.

### **Historical Market Cap**

Total market value of all of a company's outstanding shares at period-end date stated in the company's fundamental currency (DS215, EQY\_FUND\_CRNCY). The period-end date is the most recent for which full fundamental data has been collected. Calculated as:

Shares Outstanding \* Last Closing Price

### **Trailing 12M Dividends Paid**

Trailing 12-month dividends paid from the Cash Flow Statement, calculated by adding up four quarters, two semi-annuals or the latest annual value of Dividends Paid. Please use the Fundamental Period Override (DS323, FUND\_PER) to calculate the trailing value from the most recent four quarters, or use Trailing 12M Dividends Paid (T0009, TRAIL\_12M\_DVD\_PAID). Units: Actual.

### **Trailing 12M Net Capital Stock**

Trailing 12-month net increase/decrease in capital stock, calculated from the most recent four quarters, two semi-annuals or the latest annual for the ratio Net Capital Stock (Cashflow). This ratio nets the amounts in Increase in Capital Stocks and Decrease in Capital Stocks. Units: Millions.

### **Trailing 12M Change in Long Term Debt**

Trailing 12-month net increase/decrease in long-term (LT) debt, calculated from the most recent four quarters, two semi-annuals or the latest annual for the ratio Net Change in LT Debt (Cashflow). This ratio nets the amounts in the Increase in LT Borrowings and Reimbursement of LT Borrowings. Unit: Millions.

### **T12 Cash Flow from Financing-Shareholder Return**

Trailing 12-month cash flow from financing activities. Unit: Millions. Calculated as:

Trailing 12M Dividends Paid + Trailing 12M Net Capital Stock + Trailing 12M Change in Long Term Debt + Trailing 12M Change in Short Term Borrowings + Trailing 12M Other Financing Activities

### **After Tax Interest Expense**

Interest expense after taxes. Scaling Format Override (DY339, SCALING\_FORMAT) can be used to change the display units for the field. Calculated as:

Trailing 12 Month Interest Expense \* (1 - Trailing 12 Month Effective Tax Rate)

### **Trailing 12M Free Cash Flow To Firm**

Measures trailing 12-month company's profitability after deducting all expenses and reinvestment for the most recent four quarters. Calculated as:

$\text{Trailing 12M Free Cash Flow} + \text{Trailing 12-month Interest Expense} * (1 - \text{Trailing 12-month Effective Tax Rate} / 100)$

### **Periodic Enterprise Value**

Measure of a company's theoretical takeover price. Calculated as:

Market Capitalization + EV Components

### **Capital Yield**

Capital yield, computed based on the cash flow from financing activities. Calculated as:

$(\text{Trailing 12 Month Cash from Financing Activities} / \text{Periodic Enterprise Value}) * 100$

### **Tax Efficiency**

One of five components for Dupont Analysis. It measures the effects of taxes impact. Unit: Actual. Calculated as:

$(\text{Trailing 12M Net Income Available To Common Shareholders} / \text{Trailing 12M Pre-Tax Income}) * 100$

### **Normalized Net Income to Net Income for Common**

Ration of normalized net income as a multiple of net income available to common shareholders. Units in actuals. Calculated as:

$\text{Trailing 12 Month Normalized Income} / \text{Trailing 12 Month Net Income Available To Common Shareholders}$

### **Interest Burden**

Measure which determines the comparative interest expense in percentage. Calculated as:

$(\text{Trailing 12M Pre-Tax Income} / \text{Trailing 12M EBIT}) * 100$

### **Trailing 12 Month EBIT to Revenue**

Earnings before interest and taxes (EBIT) by the bottom-up method as a percentage of revenues. Calculated as:

Trailing 12-Month EBIT Bottom-up / Trailing 12M Net Sales x 100

### **Asset Turnover**

Amount of sales or revenues generated per dollar of assets. The ratio is an indicator of the efficiency with which a company is deploying its assets. Unit: Actual.

Calculated as:

Trailing 12M Net Sales / ((Total Assets - Current Period + Total Assets - Prior Year Period) / 2)

### **Normalized ROE**

Returns on Common Equity based on net income excluding one-time charges. Calculated as:

[Trailing 12 Month Normalized Income / Average of Current and Prior Period (Common Equity)]  
\*100

### **5 Year Average Adjusted Return on Equity**

Simple five year average for adjusted return on equity (ROE). If the quarterly or semiannual periodicity is selected the ratio will return the average of the last five quarters or semiannuals.

### **Total Debt Excluding Operating Lease Liabilities**

Sum of short term and long term debt. This field doesn't consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842. Units: Millions Calculated as:

Total Debt - (ST Operating Lease Liabilities + LT Operating Lease Liabilities ) or Total Operating Lease Liabilities or (ST Finance Lease & Operating Lease Liabilities + LT Finance Lease & Operating Lease Liabilities ) or Total Finance Lease & Operating Lease Liabilities

### **ST Debt Excluding Operating Lease Liabilities**

Includes bank overdrafts, short-term debts and borrowings, repurchase agreements (repos) and reverse repos, short-term portion of long-term borrowings, current obligations under capital (finance) leases, current portion of hire purchases creditors, trust receipts, bills payable, bills of exchange,

bankers acceptances, interest bearing loans, and short term mandatory redeemable preferred stock. Net with unamortized premium or discount on debt and may include fair value adjustments of embedded derivatives.

For banks and financials, data includes call money, bills discounted, federal funds purchased, and due to other banks or financial institutions.

Calculated as:

Short Term Debt - Short Term Operating Lease Liabilities or Short Term Operating & Finance Lease Liabilities

### **LT Debt Excluding Operating Lease Liabilities**

All interest-bearing financial obligations that are not due within a year. Includes convertible, redeemable, retractable debentures, bonds, loans, mortgage debts, sinking funds, and long-term bank overdrafts. Excludes short-term portion of long-term debt, pension obligations, deferred tax liabilities and preferred equity.

Calculated as:

Long Term Debt - Long Term Operating Lease Liabilities or Long Term Operating & Finance Lease Liabilities

### **Total Debt to EBITDA Ex Operating Lease Activities**

Measure of a company's ability to pay off its incurred debt and gives the investor the approximate amount of time that would be needed to pay off all debt, ignoring the factors of interest, taxes, depreciation and amortization. This field doesn't consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842. Unit: Actual.

Calculated as:

Total Debt Excluding Operating Lease Liabilities / Trailing 12 Month EBITDA After Operating Lease Expenses

### **Net Debt to EBITDA Ex Operating Lease Activities**

Measure computes the company's ability to pay off its debt by utilizing the earnings before interest, taxes, depreciation and amortization (EBITDA). Data does not consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842.

Calculated as:

Net Debt Excluding Operating Lease Liabilities / Trailing 12M Earnings After Operating Lease Expenses

### **EBITDA After Operating Lease Expn To Cash Int Paid**

Ratio of earnings before interest, taxes, depreciation and amortization (EBITDA) to cash interest paid. This field doesn't consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842. Unit: Actual

Calculated as:

$$\text{EBITDA After Operating Lease Expenses} / (\text{Cash Interest Paid} - \text{Cash Paid for Interest Related to Operating Lease or Cash Paid for Interest Related Operating and Finance Lease Liabilities})$$

### **Cash Interest Paid Excluding Operating Lease**

Cash paid out for long-term and short-term interest on any interest-bearing obligations of all consolidated entities. This field doesn't consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842. Units: Millions

Calculated as:

$$\text{Cash Paid for Interest} - \text{Cash paid for Interest Related to Operating Lease or Cash paid for Interest Related to Operating \& Finance Lease}$$

### **Interest Expenses After Operating Lease Activities**

Cost incurred by an entity for borrowed funds. Interest expense is a non-operating expense shown on the income statement. This field doesn't consider operating lease liabilities as debt after a company adopts International Financial Reporting Standards (IFRS) 16 or Accounting Standards Codification (ASC) 842. Unit: Millions

Calculated as:

$$\text{Interest Expense} - \text{Interest Expenses On Operating Lease Liability or Interest Expense on Operating and Finance Lease Liabilities}$$

### **CE to Total Assets Less Operating Lease Assets**

Percentage of ratios used to determine the financial health and long term profitability of a corporation. This ratio doesn't consider operating lease assets under total assets defined by new accounting

standard International Financial Reporting Standards (IFRS) 16 and Accounting Standards Codification (ASC) 842.

Calculated as:

$\text{Common Equity} * 100 / \text{Total Assets Less Operating Lease Assets}$

### **LT Debt Excluding Operating Lease Liabs to Eqty**

Long term debt as a percentage of total equity. This ratio doesn't consider operating lease liabilities as debt after a company adopts new accounting standard International Financial Reporting Standards (IFRS) 16 and Accounting Standards Codification (ASC) 842.

Calculated as:

$\text{Long Term Debt Excluding Operating Lease Liabilities} * 100 / \text{Total Equity}$

### **EBITDA to Interest Expense**

Also known as EBITDA (earnings before interest, taxes, depreciation, and amortization) to Interest Coverage ratio. It is used to measure a company's ability to meet its interest expenses. Unit: Actual  
Calculated as:

$\text{EBITDA} / \text{Interest Expense Banks/Finance}$

### **EBITDA Minus CAPEX/Interest**

Earnings before interest, taxes, depreciation and amortization (EBITDA) minus capital expenditures (from the statement of cash flows), divided by total interest incurred, including capitalized interest.

Unit: Actual.

Calculated as:

$(\text{EBITDA} - \text{Capex}) / \text{Total Interest Incurred}$

### **EBITDA to Cash Interest Paid**

Ratio of earnings before interest, taxes, depreciation and amortization (EBITDA) to cash interest paid. A high multiple indicates the company can service its interest payments out of current earnings. Units in actuals. Calculated as:

$\text{EBITDA} / \text{Actual Cash Paid for Interest on Debt}$

### **EBITDA after CAPEX to Cash Interest Paid**

Ratio of earnings before interest, taxes, depreciation and amortization (EBITDA) after capital expenditures, to cash interest paid. A high multiple indicates that the company can service its interest payments out of current earnings. Units in actual amounts. Calculated as:



EBITDA after CAPEX / Actual Cash Paid for Interest on Debt

### **EBIT to Cash Interest Paid**

Ratio of earnings before interest and taxes (EBIT) to cash interest paid. A high multiple indicates the company can service its interest payments out of current earnings. Units in actuals. Calculated as:

Earnings Before Interest and Taxes / Actual Cash Paid for Interest on Debt

### **Actual Cash Paid for Interest on Debt**

Cash Paid for Interest:

Cash paid out for long-term and short-term interest on any interest-bearing obligations of all consolidated entities. This amount is not reduced by allowance for funds used during construction or any capitalized interest.

Interest paid only. Not netted with interest received. Displayed as a positive number.

### **LT Debt to Total Equity**

Long term debt as a percentage of total equity. Calculated as:

$(\text{Long Term Borrowings} / \text{Total Shareholder's Equity}) * 100$

### **LT Debt to Total Capital**

Measure of a company's financial leverage calculated by dividing its long term debt by total stockholders' equity, in percentage. Calculated as:

$\text{Long Term Debt} * 100 / \text{Total Capital}$

### **LT Debt to Total Assets**

Measures the percentage of long term debt to total assets. Unit: Actual.

Calculated as:  $(\text{Long Term Borrowings} / \text{Total Assets}) * 100$

### **Total Debt to Total Equity**

Total debt divided by total shareholders' equity. Calculated as:

Short and Long Term Debt / Shareholders' Equity \* 100

### **Total Debts to Total Capital**

Measure of a company's financial leverage that presents its total debt as a percentage of total capital.  
Calculated as:

Total Debt x 100 / Total Capital

### **Operating Income or Losses**

Total operating revenue minus total operating expenses.

Total operating expenses includes interest expense, provision for loan losses, commissions and fees paid and other operating (non-interest) expenses.