

**VILNIAUS UNIVERSITETAS  
KAUNO HUMANITARINIS FAKULTETAS**

**INFORMATIKOS KATEDRA**

Verslo informatikos studijų programa  
Kodas 62109P101

**PAULIUS DANĖNAS**

**MAGISTRO BAIGIAMASIS DARBAS**

**DIRBTINIO INTELEKTO METODŲ TAIKYMAS KREDITO RI-  
ZIKOS VERTINIME**

**Kaunas 2008**

**VILNIAUS UNIVERSITETAS  
KAUNO HUMANITARINIS FAKULTETAS**

**INFORMATIKOS KATEDRA**

**PAULIUS DANĖNAS**

**MAGISTRO BAIGIAMASIS DARBAS**

**DIRBTINIO INTELEKTO METODŲ TAIKYMAS KREDITO RI-  
ZIKOS VERTINIME**

Leidžiama ginti \_\_\_\_\_  
(parašas)

Magistrantas \_\_\_\_\_  
(parašas)

Darbo vadovas doc. dr. Gintautas Garšva  
(darbo vadovo mokslo laipsnis, mokslo peda-  
goginis vardas, vardas ir pavardė)

Darbo įteikimo data \_\_\_\_\_

Registracijos Nr. \_\_\_\_\_

**Kaunas 2008**

## TURINYS

Santrumpų sąrašas .....	2
Lentelės .....	3
Ilustracijos .....	3
Summary .....	4
Įvadas .....	5
1. Analitinė darbo dalis.....	7
1.1. Dirbtinio intelekto metodai ir jų ankstesni taikymai kredito rizikos vertinimui .....	7
1.1.1. Formalus mašininio mokymo pagrindimas .....	9
1.1.2. Neuroniniai tinklai .....	12
1.1.3. Neuroninių tinklų taikymai kredito rizikos vertinimui.....	14
1.1.4. Atramos vektorių mašinos ( <i>Support Vector Machines</i> ) .....	18
1.1.4.1. Pagrindiniai SVM teoriniai principai .....	18
1.1.4.2. SVM praktiniai aspektai ir modifikacijos.....	20
1.1.4.3. SVM metodo taikymas kredito rizikos vertinimui .....	22
1.1.5. <i>Fuzzy</i> logika ir <i>fuzzy</i> ekspertinės sistemos.....	23
1.1.5.1. <i>Fuzzy</i> ekspertinės sistemos.....	24
1.1.5.2. <i>Fuzzy</i> logikos taikymo rizikos vertinime galimybės ir prielaidos.....	25
1.1.5.3. Ekspertinių taisyklių taikymo finansinės rizikos vertinime pavyzdys .....	27
1.1.6. Evoliuciniai algoritmai .....	29
1.2. Kredito rizikos vertinimo problema bei pagrindinės sąvokos .....	33
1.2.1. Rizikos sąvoka, klasifikacija ir galimi vertinimo būdai .....	33
1.2.2. Kredito rizikos sąvoka ir pagrindiniai jos valdymo aspektai .....	37
1.2.3. Kredito rizikos vertinimui naudojami rodikliai ir kintamieji .....	38
1.2.4. Kredito rizikos vertinimas, naudojant reitingavimą .....	42
1.2.5. Kredito rizikos vertinimas balais.....	43
1.2.6. Modernieji kredito rizikos vertinimo modeliai.....	44
1.3. Ekspertinės sistemos ir sprendimų priėmimas.....	47
1.3.1. Ekspertinių sistemų evoliucija ir jų taikymo kredito rizikos vertinime prielaidos....	48
1.3.2. Ekspertinės sistemos ir jų vaidmuo sprendimų priėmimo procese.....	52
1.3.3. Dirbtinio intelekto modelių finansinių sprendimų modeliavimui kūrimas .....	54
1.3.4. Intelektuali SPS kredito rizikos vertinimo procese .....	57
1.4. Atskirų DI metodų taikymo kredito rizikos vertinimui privalumai ir trūkumai.....	59
1.5. Analitinės dalies išvados .....	61
2. SVM metodu paremtos eksperimentinės sistemos tyrimas .....	63
2.1.1. Kuriamos sistemos prototipo aprašymas .....	63
2.1.2. Kuriamos sistemos paskirtis ir jos panaudojimo galimybės.....	67
2.1.3. Sistemoje naudojami metodai.....	70
3. eksperimento ir sistemos realizacijos aprašymas .....	71
3.1. Tyrimo metodikos aprašymas.....	71
3.2. Eksperimentas ir jo rezultatas.....	71
3.2.1.1. Įvertinimo metodikos aprašymas.....	73
3.2.1.2. Tinkamiausio diskriminantinio modelio bei branduolio funkcijos nustatymas. 74	
3.2.1.3. Kitų parametrų įtakos tyrimas .....	76
3.2.2. Gautų rezultatų apibendrinimas.....	78
3.3. Sukurtos sistemos realizacijos aprašymas .....	79
4. Tolimesnio tyrimo galimybės ir perspektyva .....	81
Išvados .....	83

Literatūra .....	85
Priedai .....	89
1 priedas Galimi kredito rizikos vertinimo kintamieji ir kriterijai.....	90
2 priedas Darbe naudojamų diskriminantinių rizikos vertinimo modelių sudėtis ir išraiškos .....	96
3 priedas Neuroninių tinklų taikymas bankroto prognozavime bei kredito rizikos vertinime 1992-1998 m. ....	98
4 priedas Fuzzy logikos operacijos ir išvadų darymo metodai .....	99
5 priedas Ekspertinių taisyklių realizacija CLIPS/JESS ekspertinėse aplinkose .....	100
6 priedas Evoliucinių metodų strategijos ir algoritmai .....	103
7 priedas Sistemoje naudojamų duomenų aprašymas ir duomenų bazės schema.....	107
8 priedas Pagrindinių šiuolaikinių DBVS privalumai ir trūkumai.....	111
9 priedas Sukurtos sistemos realizacijos vartotojo sąsajos pavyzdžiai.....	112

## SANTRUMPŲ SĄRAŠAS

BVP – bendrasis vidaus produktas

AI/DI – angl. artificial intelligence/liet. dirbtinis intelektas

DB – duomenų bazė

IR – informacijos išgavimas (angl. Information retrieval)

DM – duomenų gavyba (angl. Data Mining)

SPS - sprendimų paramos sistema (angl. Decision Support Systems, sutr. DSS)

ISPS – intelektualiai sprendimų paramos sistema

MDDDB – daugiamatė duomenų bazė (angl. Multidimensional Database)

OLAP – realaus laiko analitinis apdorojimas (angl. Online Analytic Processing)

SOM – saviorganizuojantis žemėlapis (angl. Self-Organizing Map)

SQL – užklausų kūrimo kalba, naudojama duomenų bazėse (Structured Query Language)

ML – mašininis mokymas (angl. Machine Learning)

NN – neuroninis tinklas (angl. Neural Network)

DNN – dirbtinis neuroninis tinklas (angl. Artificial Neural Network)

GA – genetinis algoritmas (angl. Genetic Algorithm)

*Fuzzy* logika – neraiškioji (angl. fuzzy) logika

SVM – atramos vektorių mašinos (angl. Support Vector Machines)

v-SVM – nu-SVM (v-SVM) atramos vektorių mašinų metodas

v-SVC - v-SVM klasifikavimo metodas

SV – atraminis vektorius (angl. Support vector)

Branduolys (angl. Kernel) – funkcija, pagal duotus du taškus grąžinanti jų panašumą

RBF – radialinės bazės funkcija (angl. Radial Basis Function)

Jess –posistemis ekspertinių sistemų kūrimui JAVA aplinkoje

UML – unifikuota modeliavimo kalba (angl. Unified Modelling Language)

SysML – moeliavimo kalba, išplečianti UML galimybes (angl. System Modelling Language)

CSV – specialus duomenų failo formatas (angl. Comma Separated Values)

LibSVM – darbe naudojama SVM realizacija, sukurta Kinijos mokslininkų Chang ir Lin

## LENTELĖS

1 lentelė.	Neuroninių tinklų taikymo kredito rizikos vertinimui tyrimų pavyzdžiai .....	16
2 lentelė.	DNN privalumai ir trūkumai .....	17
3 lentelė.	SVM algoritmai bei jų realizacijos .....	21
4 lentelė.	Fuzzy logikos taisyklių taikymas nagrinėjant kredito rizikos galimybę.....	26
5 lentelė.	Kiekybinių finansinių rodiklių modeliavimo taisyklės FINEVA .....	28
6 lentelė.	Kokybinių finansinių rodiklių modeliavimo taisyklės FINEVA.....	28
7 lentelė.	Rizikų analizė ir galimi vertinimo metodai .....	34
8 lentelė.	Kiekybiniai ir nekiekybiniai parametrai .....	41
9 lentelė.	Vertinimo balais modelių esminiai privalumai ir trūkumai .....	43
10 lentelė.	Modernių kredito rizikos vertinimo modelių palyginimas .....	46
11 lentelė.	Dirbtinio intelekto apibrėžimų kategorijos .....	49
12 lentelė.	Pirmosios dirbtiniu intelektu paremtos programos .....	50
13 lentelė.	DI, duomenų gavybos ir statistikos mokslų vystymosi evoliucija.....	51
14 lentelė.	Dirbtinio intelekto metodų taikymo kredito rizikos vertinimui privalumai ir trūkumai.....	60
15 lentelė.	Klasikinių ir DI metodų taikymo kredito rizikos vertinimui lyginamoji analizė.....	61
16 lentelė.	Sistemą sudarančių komponentų aprašymas .....	66
17 lentelė.	v-SVM metodo parametrai ir aprašymai.....	72
18 lentelė.	v įtakos tyrimas Altman KRV modelio ir RBF bei tiesinės branduolio funkcijos atveju.....	77
19 lentelė.	$\gamma$ parametro įtakos tyrimas Altman KRV modelio ir RBF branduolio funkcijos atveju.....	78
20 lentelė.	Tyrimo metu gauti rezultatai .....	79

## ILIUSTRACIJOS

1 pav.	Duomenų gavybos modeliai ir uždaviniai.....	7
2 pav.	Pagrindiniai klasifikavimo ir klasterizavimo algoritmai .....	8
3 pav.	Mokymosi iš pavyzdžių principinė schema .....	9
4 pav.	Struktūrinis rizikos minimizavimas .....	11
5 pav.	Neuroninio tinklo struktūra ir veikimo principas .....	13
6 pav.	SOM žemėlapiu pritaikymo klasifikavimui pavyzdys .....	17
7 pav.	Tiesinė atraminių vektorių mašina .....	19
8 pav.	FINEVA naudojami finansiniai įverčiai ir rodikliai.....	27
9 pav.	Genetinio algoritmo principinė schema .....	31
10 pav.	Banko rizikos klasifikacija.....	34
11 pav.	Skolininko analizės procesas.....	39
12 pav.	Laukiamo išsipareigojimų nevykdymo dažnio apskaičiavimas pagal Moody's KMW .....	45
13 pav.	Sprendimų priėmimas ir jį apimantys procesai .....	53
14 pav.	Intelektualios SPS vieta kreditų rizikos vertinimo procese.....	58
15 pav.	Informacijos pateikimas DSS ir jos panaudojimas reitingavimo procese .....	59
16 pav.	Sistemos reikalavimų diagrama .....	63
17 pav.	Sistemos duomenų srautų diagrama .....	64
18 pav.	Sistemos komponentų diagrama.....	65
19 pav.	Sistemos diegimo diagrama .....	65
20 pav.	Modelio kūrimo kompozicijos diagrama .....	66
21 pav.	Galimos ISPS kredito rizikos vertinimui architektūra .....	67
22 pav.	Galima modelio kūrimo veiksmų sekos diagrama .....	68
23 pav.	Išplėstinės sistemos komponentų diagrama.....	69
24 pav.	v parametro įtakos Altman modelio ir RBF bei tiesine branduolio funkcijomis atveju rezultatai .....	78
25 pav.	Sukurtos sistemos vartotojo sąsajos pavyzdys .....	80
26 pav.	Galimas GA/SVM hibridinis algoritmas.....	82

DANĖNAS, Paulius. (2008) . *Application of artificial intelligence methods in credit risk evaluation*. MBA Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty of Humanities, Department of Informatics. 119 p.

## **SUMMARY**

This master work describes the most widely used artificial intelligence methods and the possibilities to apply them in credit risk evaluation which is one of the most important fields in banking in finance. The main problem here is to evaluate the risk arising when a creditor gives a credit to a particular individual or an enterprise, using various mathematical, statistical or other methods and techniques. This risk arises when the debtor isn't able to pay for the loan to the creditor in time which means additional loss. It can appear in many forms depending on the type of debtor (individual, enterprise, government of an abroad country) and type of financial instrument or action that is done with it (giving of a loan, transactions of financial derivatives, etc.), this is the reason why financial institutions and for its evaluation and management use various different methodologies which comprise a lot of methods and techniques from credit scoring (evaluating by a particular formula, usually linear) and evaluating different factors, like management and business strategies or policies, to classification by various criterions by using modern and sophisticated methods, either algebraic, either artificial intelligence and machine learning. This field is widely researched and many new techniques are being found.

The research here is concentrated mainly on Support Vector Machines (abbr. SVM) which is one of the most popular artificial intelligence and machine learning techniques and whose effectiveness has been widely proved. This research is done in order to investigate the possibilities to adapt SVM method to the problem described above, and to implement a system which uses one of SVM techniques.

## ĮVADAS

Kredito rizikos vertinimas – viena svarbiausių sričių bankininkystėje ir finansuose, kurios pagrindinis uždavinys yra rizikos, atsirandančios kreditoriui išduodant paskolą kuriam nors konkrečiam kredituojamam asmeniui, nustatymas bei įvertinimas, tam naudojant įvairius matematinius metodus bei modelius. Ši rizika atsiranda dėl skolininko nesugebėjimo laiku atsiskaityti kreditoriui, kas pastarajam reiškia papildomus nuostolius. Ši rizikos rūšis pasireiškia daugeliu formų, priklausomai nuo skolininko tipo (ar tai individualus asmuo, ar organizacija, ar užsienio vyriausybė) bei finansinio instrumento tipo bei veiksmo su juo (paskolos išdavimas, finansinių derivatyvų transakcijos ir kt.), todėl finansinės institucijos jos vertinimui bei valdymui naudoja įvairias metodikas, kurios apima daug metodų bei modelių, nuo kredito vertinimo balais (angl. *credit scoring*), apskaičiuojant įvertinimą pagal tam tikrą formulę, bei atskirų rodiklių, tokių, kaip rizikos vertė, įvertinimas iki organizacijų bei kitų subjektų, kuriems skolinami tam tikri finansiniai ištekliai, klasifikavimo pagal tam tikrus kriterijus, tam naudojant modernius ir sudėtingus metodus, prie kurių galima priskirti tiek statistinius/ekonometrinius/matematinus, tiek ir dirbtinio intelekto srities metodus.

**Tyrimo tema** aktuali tuo, kad tiriamą sritį šiuo metu gana plačiai nagrinėjama visame pasaulyje; ne vienoje mokslinių straipsnių bazėje (čia galima paminėti tokius šaltinius, kaip „default-risk.com“, „gloriamundi.org“, skirtus kredito bei finansinei rizikai, bei „CiteSeer“ duomenų bazę) galima rasti straipsnių, vienu ar kitu aspektu susijusių su nagrinėjama tema. Lietuvoje ši sritis nagrinėjama taip pat plačiai, tačiau tik teoriniu aspektu, taikant įprastinę metodiką; nagrinėjamos krypties kontekste galima išskirti Vilniaus universiteto Kauno Humanitarinio fakulteto mokslininkų pastangas. Pagrindinis tyrimo privalumas būtų tas, kad jo rezultatus būtų galima pritaikyti ir praktiškai, kuriant ekspertinę sistemą, kurią būtų galima naudoti tiek moksliniais tikslais, tiek ir praktikoje, realizuojant ją kaip konkretaus banko informacinės sistemos atskirą modulį.

**Tyrimo objektas** – dirbtinio intelekto metodai ir jų pritaikymo kredito rizikos vertinime galimybės.

**Tyrimo tikslas** – išnagrinėti dirbtinio intelekto taikymo kreditavimo rizikos valdyje galimybes bei pasiūlyti sprendimą, kurie dirbtinio intelekto metodai ir kaip galėtų būti pritaikyti, vertinant kredito riziką.

### **Tyrimo uždaviniai:**

1. Išnagrinėti kredito rizikos vertinimo balais modelių ir moderniujų modelių privalumus ir trūkumus;
2. Ištirti dirbtinio intelekto metodų panaudojimo kredito rizikos vertinime galimybes ir dirbtiniu intelektu paremtų metodų privalumus bei trūkumus bei palyginti juos su kredito rizikos



reitingavimo ir modernių modelių galimybėmis;

3. Atrinkti geriausiai tinkančius kredito rizikos vertinimui dirbtinio intelekto metodus;
4. Sumodeliuoti ir suprojektuoti bent 1 DI metodą naudojančią kredito rizikos vertinimo sistemą;
5. Realizuoti sumodeliuotą sistemą;
6. Ištirti ir įvertinti gautuosius empirinio tyrimo rezultatus.
7. Pateikti pasiūlymus, susijusius su nagrinėjama tema.

Rašant darbą, naudoti tokie *metodai*:

- Visuotinio pažinimo metodas (darbo tikslų nustatymas, uždavinių formulavimas, informacijos apie dirbtinio intelekto metodus kredito rizikos vertinimui rinkimas ir analizė; duomenims apibendrinimas; išvadų formulavimas).

- Bendrieji mokslinio tyrimo metodai:
  - indukcijos metodas (tyrimas, darbo išvadų formulavimas);
  - dedukcijos metodas darbo eigoje nuo bendro sprendimo prie atskirų dalių);
  - palyginimo metodas (metodų charakteristikų, panašumų ir skirtumų palyginimas).

- Analizės metodai:
  - duomenų analizės metodas (esamos situacijos vertinimas).

- Modeliavimo metodas (medžiagos grupavimui, apibendrinimui bei išdėstymui).

- Abstrakcijos metodas – juo remiantis, daromos kiekvienos darbo dalies ir galutinės viso darbo išvados.

Darbą sudaro trys pagrindinės dalys – analizės dalis, eksperimentinės sistemos tyrimo dalis bei eksperimentinė dalis. Kiekvienoje iš šių dalių pateikiama atitinkama informacija apie atliekamą tyrimą ir kuriamą informacinę sistemą:

- *Analizės dalyje* aprašyta nagrinėjama problema, apžvelgtos pagrindinės su ja susijusios sąvokos, klausimai bei ankstesni tyrimai, pasirinktas metodas, pagal kurį bus atliktas tyrimas.

- *Eksperimentinės sistemos tyrimo dalyje* iškelti techniniai reikalavimai būsimai sistemai, apibrėžtos funkcijos, kurias ji turi atlikti, pateikiama jos koncepcinė schema ir aptariamos jos išplėtimo ateityje galimybės.

- *Eksperimento dalyje* aprašomas su sukurta informacine sistema atliktas tyrimas, pateikiami jo rezultatai bei pasiūlymai tolimesniam tyrimui.

Darbe naudotos knygos apie kredito riziką, finansus bei intelektinius metodus, Interneto medžiaga, moksliniai straipsniai, paskaitų konspektai. Darbo apimtis – 88 puslapiai be priedų, 119 puslapių su priedais. Darbe yra 20 lentelių, 26 diagramos bei 9 priedai.

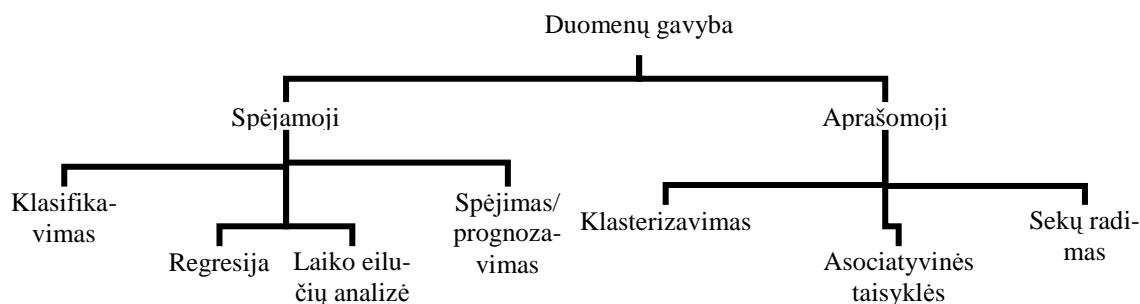
# 1. ANALITINĖ DARBO DALIS

Šioje darbo dalyje apžvelgiama nagrinėjama problema, jos pagrindinės sąvokos bei anksčiau taikyti metodai. Nagrinėjamu atveju problema yra kredito rizikos vertinimo dirbtinio intelekto pagrindų, todėl apžvelgiami tokie klausimai, kaip:

- Kredito rizikos vertinimo problema, įvairios rizikos rūšys, sąvokos, vertinimas bei jų sąveika su kredito rizika;
- Trumpai apžvelgiami klasikiniai rizikos vertinimo būdai, diskriminantiniai vertinimo balais modeliai bei modernūs modeliai, trumpai aptariami jų privalumai ir trūkumai;
- Apžvelgiama dirbtinio intelekto ir ekspertinių sistemų sąvokos, raida, vystymasis bei prielaidos tai taikyti kredito rizikos vertinimui;
- Aptariamas ekspertinių sistemų vaidmuo sprendimų priėmimo procese bei jų vieta kredito rizikos vertinime;
- Pateikiamas formalus mašininio mokymo pagrindimas, apžvelgiami pagrindiniai metodai (neuroniniai tinklai, atraminių vektorių mašinos, fuzzy logika, evoliuciniai skaičiavimai), jų pritaikymas kredito rizikos vertinime, ankstesni tyrimai bei įvertinamos jų pritaikymo galimybės.

## 1.1. Dirbtinio intelekto metodai ir jų ankstesni taikymai kredito rizikos vertinimui

Kredito rizikos vertinimas – viena svarbiausių sričių bankininkystėje ir finansuose, kurios pagrindinis uždavinys yra rizikos, atsirandančios kreditoriui išduodant paskolą kuriam nors konkrečiam kredituojamam asmeniui, nustatymas bei įvertinimas, tam naudojant įvairius matematinius metodus bei modelius. Išsivysčius dirbtinio intelekto sričiai, atsirado galimybė šioje srityje naudoti ir dirbtinio intelekto bei hibridinius (apimančius du ar daugiau dirbtinio intelekto ar matematinius-statistinius metodus) modelius. Šiuo atveju nagrinėjamą problemą galima aprašyti ir kaip problemą, kurios sprendimui gali būti naudojami tokie duomenų gavybos metodai, kaip klasifikavimas, klasterizavimas ar regresija.



Šaltinis: Margaret Dunham. Data Mining: Introductory and Advanced Topics

1 pav. Duomenų gavybos modeliai ir uždaviniai

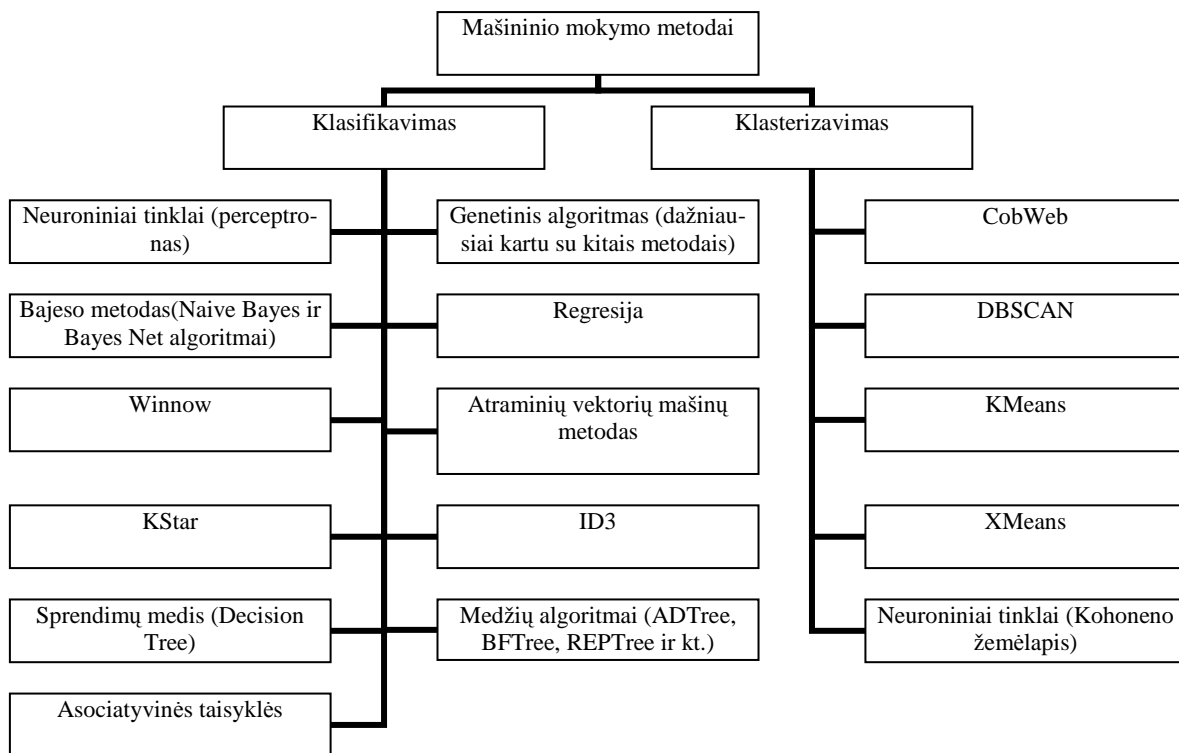
Duomenų klasifikavimo atveju sprendžiama problema gali būti suprantama kaip duomenų suskirstymas į tam tikras klases, pagal kurias nustatoma kreditavimo galimybė, arba jų išskaidymas į tam tikrus klasterius, pagal kurių skaičių ir tam tikro duomenų vieneto padėtį viename ar kitame klasteryje galima priimti vieną ar kitą sprendimą.

Pagrindinės duomenų gavybos modelių grupės pagal jų atliekamas funkcijas bei sprendžiamus uždavinius pateikiamos 1 pav.

Šiuo metu galima išskirti tokius pagrindinius dirbtinio intelekto metodus ir kryptis:

- dirbtiniai neuroniniai tinklai ir *neuro-fuzzy* sistemos;
- neraiškioji logika (*fuzzy logic*; dėl aiškumo toliau darbe tiesiog *fuzzy* logika);
- evoliuciniai skaičiavimai bei genetiniai algoritmai;
- duomenų gavyba;
- atramos vektorių mašinos (*support vector machines*, sutr. SVM).

2 pav. pateikiami šiuo metu populiariausi klasifikavimo ir klasterizavimo metodai.



Šaltinis: sudaryta autoriaus.

## 2 pav. Pagrindiniai klasifikavimo ir klasterizavimo algoritmai

Šiame skyrelyje trumpai apžvelgiami šiuo metu plačiai nagrinėjami bei taikomi pagrindiniai dirbtinio intelekto metodai ir jų taikymai kredito rizikos vertinime. Prieš nagrinėjant dirbtinio intelekto metodus ir jų pritaikymo galimybes, svarbu apžvelgti pagrindinius mašininio mokymo šakos principus, naudojamus kuriant šiais metodais paremtas sistemas. Šia šaka paremti tokie metodai, kaip:

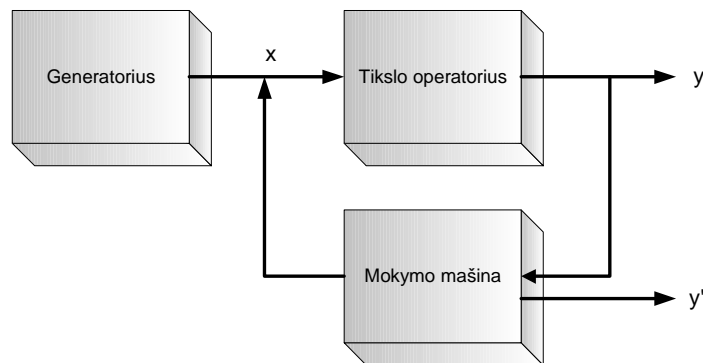
- Kai kurie neuroninių tinklų tipai (perceptronas, daugiasluoksnis perceptronas);
- Atramos vektorių mašinos.

Šios dvi sritys bei jų pritaikymo galimybės kredito rizikos vertinimui išsamiau apžvelgiamos 4.2 ir 4.5 skyreliuose.

### 1.1.1. Formalus mašininio mokymo pagrindimas

Ši dirbtinio intelekto atšaka yra viena pagrindinių mokslo šakų, kurios principai ir metodai taikomi tiek modelių ir struktūrų atpažinimui, tiek ir statistiniam klasifikavimui. Šis metodas pagrįstas mokymosi iš pavyzdžių principu, kurį struktūriškai galima aprašyti kaip generatoriaus, tikslo operatoriaus (mokytojo) ir mokymo mašinos sąveiką. Šiuo atveju išskiriami tokie komponentai:

- Generatorius, aprašantis mokymo mašinos ir mokymo operatoriaus veikimo aplinką. paprastai tai būna įėjimo duomenų vektorius ir tikimybės pasiskirstymo funkcija  $f(x)$ ;
- Tikslo (mokymo) operatorius, transformuojantis įėjimo duomenų vektorius  $x$  į reikšmes  $y$  pagal pasiskirstymo funkciją  $F(y/x)$ . Tai paprastai būna fiksuota, bet nežinoma reikšmė;
- Mokymo mašina – pagal duotus nepriklausomus ir identiškai paskirstytus pagal jungtinę pasiskirstymo funkciją  $F(x,y) = F(y/x)F(x)$  mokymosi duomenis  $(x_1,y_1), (x_2,y_2), \dots, (x_n,y_n)$  sukuria funkcionalų aibę  $g(x, \alpha), \alpha \in \Lambda$  bei nuspėja mokytojo atsakymą  $y_i$  bet kuriam vektoriui  $x_i$ , t.y., apsimuoja pagal nežinomą operatorių arba pasirenka optimalią funkciją iš minėtos funkcionalų aibės.



**3 pav. Mokymosi iš pavyzdžių principinė schema**

Taigi mokymosi procesas yra tinkamos funkcijos iš duotos funkcijų aibės pasirinkimo procesas, kurio pagrindiniai du tikslai yra:

- Mokymo operatoriaus imitavimas, duoto generatoriaus išėjimui sukuriant geriausių spėjimų rezultatus;
- Mokymo operatoriaus radimas, apmokymo metu siekiant rasti būti arčiausiai mokymo operatoriaus.

Formaliai mašininio mokymo proceso uždavinius galima suformuluoti taip:

- Vektorinės erdvės  $R^n$  poaibyje  $Z$  duota aibė priimtinių funkcijų  $\{g(z)\}$ ,  $z \in Z$  ir apibrėžtas funkcionalas

$$R = R(g(z)) \tag{1}$$

kaip pasirinktos funkcijos kokybės kriterijus. Uždavinio tikslas – rasti funkciją  $g^*(z)$  iš aibės  $\{g(z)\}$ , kuri minimizuoja (1) funkciją. Šiuo atveju akcentuojama tinkamos funkcijos paieška ir sprendžiamas klausimas, kaip rasti funkcionalo minimumą duotoje funkcijų aibėje.

- Vektorinės erdvės  $R^n$  poaibyje  $Z$  apibrėžta tikimybės pasiskirstymo funkcija  $F(z)$  ir apibrėžtas funkcionalas

$$R(g(z)) = \int L(z, g(z)) dF(z) \tag{2}$$

kai funkcija  $L(z, g(z))$  yra integruojama su kiekvienu  $g(z) \in \{g(z)\}$ . Uždavinio tikslas – minimizuoti (2) funkciją, t.y., suformuluoti konstruktyvų kriterijų funkcijos pasirinkimui, o ne jos ieškoti duotoje funkcijų aibėje. Šiuo atveju akcentuojama tai, kas turi būti minimizuota tam, kad būtų galima pasirinkti funkciją  $g(z) \in \{g(z)\}$  t.y., kas garantuoja, kad funkcionalo (2) reikšmė maža. Ši minimizavimo problema gali būti suformuluota taip: funkcijų aibė  $g(z)$  pateikiama parametrinėje formoje  $\{g(z, \alpha)\}$ ,  $\alpha \in \Lambda$ . Aibė  $\Lambda$  šiuo atveju gali būti tiek skaliariniai duomenys, tiek vektoriai, tiek abstrakcijos. Šiuo atveju reikalingos funkcijos radimas reiškia tinkamo parametro  $\alpha \in \Lambda$  nustatymą. Tada duotą funkcionalą galima perrašyti taip:

$$R(\alpha) = \int Q(z, \alpha) dF(z), \alpha \in \Lambda, \text{ kur } Q(z, \alpha) = L(z, g(z, \alpha)), \tag{3}$$

Funkcija  $Q(z, \alpha)$  vadinama *praradimo funkcija*. Iš (3) funkcijos galima nustatyti ir numatomą tikslumo praradimą arba riziką:

$$R(\alpha^*) = \int Q^*(z, \alpha) dF(z), \alpha \in \Lambda \tag{4}$$

(4) formulė vadinama *rizikos funkcionalu* arba tiesiog *rizika*. Tolimesniuose skaičiavimuose sprendžiamas būtent šio funkcionalo minimizavimo uždavinys, kurio sprendimui paprastai naudojami du rizikos skaičiavimo būdai:

- *empirinis* – vietoje rizikos funkcionalo minimizuojamas empirinis rizikos funkcionalas

$$R_{emp}(\alpha) = \frac{1}{l} \sum_{i=1}^l Q(z, \alpha), \alpha \in \Lambda \tag{5}$$

Šiuo atveju funkcija, minimizuojanti šį funkcionalą, yra  $Q(z, \alpha_l) = Q(z, \alpha(z_1, z_2, \dots, z_l))$ . Ši rizika dar vadinama mokymosi paklaida. Empiriniam rizikos minimizavimui taip pat yra taikoma keletas metodų:

- didžiausio galimumo metodas (*maximum likelihood method*) - šiuo atveju minimi-

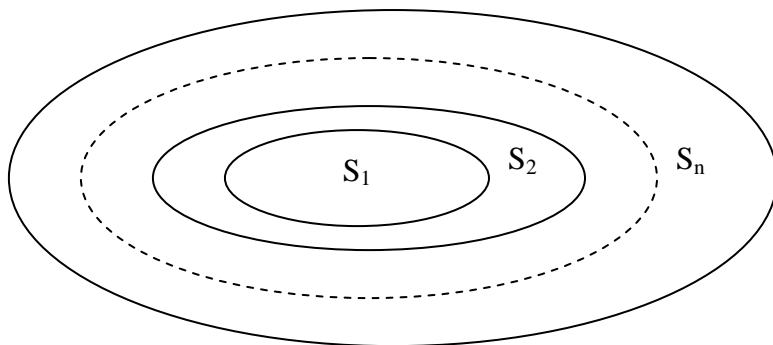
zuojamas funkcionalas  $R_{emp}(\alpha) = -\frac{1}{l} \sum_{i=1}^l \log p(z_i, \alpha)$ ;

- *struktūrinis* – indukcinis principas, leidžiantis mokytis iš nedidelio dydžio pavyzdžių. Šis metodas leidžia kontroliuoti apibendrinimo savybę besimokančioms mašinoms, naudojančioms apmokymo duomenų mažo dydžio pavyzdžius. Struktūrinio rizikos minimizavimo esmė yra rizikos minimizavimas atskiriems duomenų poaibiams, siekiant gauti poaibį su optimalia (mažiausia) rizikos riba (4 pav.). Kitaip tariant, struktūrinio rizikos minimizavimo principas aprašo *kompromisą tarp duotų duomenų aproksimacijos kokybės ir aproksimuojamos funkcijos sudėtingumo* (Vapnik, 2000). Šis principas leidžia rasti funkciją, kuri fiksuotam duomenų kiekiui minimizuoja dešinę rizikos intervalo ribą, taip pat padeda nustatyti optimalų ryšį tarp empirinių duomenų kiekio, duomenų aproksimacijos kokybės bei vertės, charakterizuojančios funkcijų aibės apimtį. Šiuo atveju lygia-grečiai visoms funkcijoms iš aibės visiškai apribotų funkcijų  $0 \leq Q(z, \alpha) \leq B$ ,  $\alpha \in \Lambda$  su tikimybe 1- $\eta$  teisinga nelygybė

$$R(\alpha) \leq R_{emp}(\alpha) + \frac{B\varepsilon(l)}{2} \left(1 + \sqrt{1 + \frac{4R_{emp}(\alpha)}{B\varepsilon(l)}}\right), \text{ kur } \varepsilon(l) = 4 \frac{h(\ln \frac{2l}{h} + 1) - \ln \frac{\eta}{4}}{l} \quad (6)$$

Struktūrinis rizikos minimizavimas iliustruojamas 4 pav. Iš čia esančios funkcijų aibės hierarchiškai sudaroma poaibių aibė  $S_1 \subset S_2 \subset \dots \subset S_n$ , kurioje  $S_k = \{Q(z, \alpha), \alpha \in \Lambda_k\}$  ir  $S^* = \bigcup_k S_k$ .

Didėjant indeksui n, mažėja visų empirinių rizikų dešinioji riba (Vapnik, 2000).



VC dimensijos h aibės  $S_k$  elementams:  
 $h_1 \leq h_2 \leq \dots \leq h_n \leq \dots$

Rizikos ribos aibės  $S_k$  elementams:  
 $B_1 \leq B_2 \leq \dots \leq B_n \leq \dots$

$\tau_k$  parametrų ribos aibės  $S_k$  elementams:  
 $\tau_1 \leq \tau_2 \leq \dots \leq \tau_n \leq \dots$

Šaltinis: sudaryta autoriaus pagal (Vapnik, 2000).

#### 4 pav. Struktūrinis rizikos minimizavimas

Galima išskirti ir tris pagrindinius uždavinių tipus, kurie sprendžiami, panaudojant mašininį mokymą:

- *Modelių (struktūrų) atpažinimas* – kai stebimos atsirandančios situacijos ir nustatoma,

kuriai iš  $k$  klasių kiekviena jų priklauso. Formaliai tai aprašoma kaip  $F(\omega|x), \omega \in \{0,1,\dots,k-1\}$ , kur  $F(\omega|x)$  yra sprendimo priėmimo taisyklė. Šiuo atveju uždavinio sprendimo tikslas yra funkcionalo  $R(\alpha) = \int L(\omega, \phi(x, \alpha)) dF(\omega, x)$  minimizavimas funkcijų aibėje  $\phi(x, \alpha), \alpha \in \Lambda$ , kur žinomi atsitiktinės nepriklausomos pavyzdžių poros  $(\omega_1, x_1), (\omega_2, x_2), \dots, (\omega_l, x_l)$ , bet nežinoma pasiskirstymo funkcija  $F(\omega, x)$ ; čia  $\phi$  aprašo sprendimų taisykles,  $l$  yra įėjimos duomenų (pavyzdžių) skaičius. Vektorius  $z$  šiuo atveju susideda iš  $n+1$  koordinačių: koordinatės  $\omega$ , įgyjančios prasmę tik su baigtiniu reikšmių skaičiumi, bei vektoriaus  $x$  koordinačių  $x^1, x^2, \dots, x^n$ .

- *Regresijos uždavinys* – sprendžiamas, kai dvi elementų aibės  $X$  ir  $Y$  susijusios funkciniu ryšiu taip, kad kiekvienam aibės  $X$  elementui  $x$  atitinka unikalus aibės  $Y$  elementas  $y$ , t.y., žinomos pavyzdžių poros  $(y_1, x_1), (y_2, x_2), \dots, (y_l, x_l), y \in Y$ . Šiuo atveju yra ieškomas stochastinis ryšys empiriniuose duomenyse, t.y., įvertinama sąlyginė pasiskirstymo funkcija  $F(y|x)$ , kas yra sudėtingas uždavinys. Čia pagrindinis uždavinys yra regresijos funkcijos  $r(x) = \int y dF(y|x)$  radimas funkcijų aibėje  $f(x, \alpha), \alpha \in \Lambda$  pagal sąlygas  $\int y^2 dF(y, x) < \infty, \int r^2 dF(y, x) < \infty$ . Šiuo atveju regresijos įvertinimo uždavinys apribojamas iki empirinių duomenų rizikos funkcionalo  $R(\alpha) = \int (y - f(x, \alpha))^2 dF(y, x)$  minimizavimo.

- *Tankio įvertinimas* – šiuo atveju turime tikimybių tankių aibę  $p(x, \alpha) = \frac{dF(x)}{dx}$ , o uždavinio sprendimas apribojamas iki funkcionalo  $R(\alpha) = -\int \ln p(x, \alpha) dF(x)$  minimizavimo.

### 1.1.2. Neuroniniai tinklai

Neuroniniai tinklai - pats populiariausias ir daugiausiai tirtas bei šiuo metu taikomas dirbtinio intelekto metodas, kuris bando pakartoti žmogaus smegenų lankstumą sprendžiant problemas. Kaip dirbtinio intelekto sritis jis išsamiausiai ir tiksliausiai apibrėžiamas *Britannica Concise* žodyne: „tai lygiagretaus skaičiavimo tipas, kuriame skaičiuojamieji elementai modeliuojami neuronų tinkle, atitinkančiame gyvūnų nervines sistemas. Šis modelis, kurio tikslas yra simuliuoti būdą, kuriuo smegenys apdoroja informaciją, leidžia kompiuteriui iki tam tikro laipsnio „mokyti“. Neuronų tinklas tipiškai susideda iš tam tikro susijusių procesorių (elementų) skaičiaus, kurių kiekvienas valdo tam tikrą žinių sritį, bei turi keletą įėjimų ir vieną išėjimą į tinklą. Pagal gaunamus įėjimos duomenis elementas gali mokyti apie ryšius tarp duomenų rinkinių, kartais naudojant *fuzzy* logikos principus“.

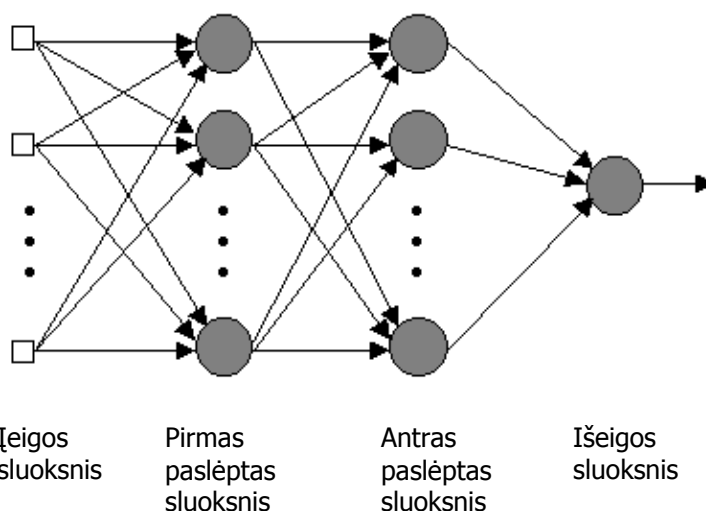
Taigi neuroniniai tinklai gali būti apibūdinti kaip modeliavimo technika, paremta biologiniais neuronų stebėjimais ir gebanti pamėgdžioti kitos sistemos veikimą. Tai kompiuterinė architektūra, sumodeliuota pagal žmogaus smegenyse susietą neuronų sistemą, kuri imituoja smegenų gebė-

jimą rūšiuoti ir mokytis iš klaidų, taip pat nustatyti pagrindinius ryšius tarp duomenų. Formaliai ji yra aprašoma kaip orientuotasis grafas su daug elementų (apdorojantys elementai) ir lankų (ryšių) tarp jų; kiekvienas iš šių elementų funkcionuoja nepriklausomai nuo kitų ir naudoja tik lokalius įėjigos ir išėjigos duomenis savo veiksenos valdymui, todėl ši galimybė leidžia neuroninius tinklus naudoti ir paskirstytoje bei (arba) lygiagrečioje aplinkoje (Dunham, 2003). Ši autorė pateikia ir du jo apibrėžimus, kaip orientuotojo grafo ir kaip skaičiuojamojo modelio. Kaip duomenų struktūra neuroninis tinklas – tai orientuotasis grafas  $F = \langle V, A \rangle$  su viršūnėmis  $V = \{1, 2, \dots, n\}$  ir lankais  $A = \{ \langle i, j \rangle \mid 1 \leq i, j \leq n \}$  su šiais apribojimais:

1.  $V$  yra suskirstytas į įėjigos elementų aibę  $V_I$ , paslėptų elementų aibę  $V_H$  ir išėjigos elementų aibę  $V_O$ ;
2. Viršūnės taip pat yra suskirstytos į sluoksnius  $\{1, \dots, k\}$  su visais įėjigos mazgais sluoksnyje 1 ir išėjigos mazgais sluoksnyje  $k$ . Visi paslėpti mazgai yra sluoksniuose nuo 2 iki  $k - 1$ , kurie vadinami paslėptais sluoksniais;
3. Bet kuris lankas  $\langle i, j \rangle$  privalo turėti mazgą  $i$  sluoksnyje  $h - 1$  ir mazgą  $j$  sluoksnyje  $h$ ;
4. Lankas  $\langle i, j \rangle$  pažymėtas skaitine verte  $w_{ij}$ ;
5. Mazgas  $i$  pažymėtas funkcija  $f_i$ .

Kaip algoritmas neuroninis tinklas apibrėžiamas taip – tai skaičiavimo modelis, susidedantis iš trijų dalių:

1. Neuroninio tinklo grafo, apibrėžiančio neuroninio tinklo duomenų struktūrą;
2. Mokymosi algoritmo, nusakančio, kaip vyksta tinklo apsimokymas;
3. Atsiminimo metodų, apibrėžiančių, kaip informacija gaunama iš tinklo.



Šaltinis: sudaryta autoriaus

**5 pav. Neuroninio tinklo struktūra ir veikimo principas**



Neuroninių tinklų apmokymo procese gali skirtis tik naudojami algoritmai, neuronų skaičius, sluoksnių skaičius. Neuroninių tinklų sudėčiai nusakyti naudojami 2 terminai: neurodinamika ir architektūra. Neurodinamika apibūdina atskiro neurono savybes, tokias kaip perdavimo funkcija ir kaip apjungiami įvedami duomenys. Neuroninių tinklų architektūra nusako jo struktūrą, tame tarpe neuronų skaičių kiekviename sluoksnyje ir tarpusavio ryšių skaičių.

*Paslėptų sluoksnių skaičius.* Paslėpti sluoksniai leidžia neuroniniam tinklui apibendrinti duomenis. Tinklas, turintis vieną paslėptą neuronų sluoksnį gana gerai sugeba apskaičiuoti bet kokią tolydžią funkciją; norint pasiekti kuo geresnių rezultatų, geriausia naudoti vieno arba dviejų paslėptų neuronų sluoksnių tinklą.

*Paslėptų neuronų skaičius.* Nepaisant to, kad neuronų skaičiaus parinkimas yra svarbus faktorius kuriant neuroninį tinklą prognozavimui, nėra nustatyta, koks turėtų būti neuronų skaičius. Optimalus variantas gali būti pasiekiamas tik eksperimentiniu būdu su skirtingais neuronų skaičiais sistemose. Parinkus per mažą paslėptų elementų skaičių, tikslo funkcija gali būti neaprosimuota, tuo tarpu didelis elementų skaičius gali sukelti persimokymą (*overfitting*).

*Išvedimo neuronų skaičius.* Išvedimo neuronų skaičiaus nustatymas yra šiek tiek lengvesnė užduotis paslėptų neuronų skaičiaus nustatymas. Išvedimui dažniausiai naudojamas vienas neuronas; klasifikavimo atveju paprastai laikoma, kad išvedimo neuronų skaičius yra toks pats, kaip ir klasių skaičius, nors kai kuriais atvejais tai ir nėra teisinga (Dunham, 2003).

*Perdavimo funkcijos.* Perdavimo funkcijos yra matematinės formulės, kurios apibrėžia apdorojamo neurono išvedimą. Jos taip pat gali būti vadinamos transformacija, aktyvizacija; tai gali būti pradinės funkcijos. Dauguma dabartinių neuroninių tinklų naudoja sigmoidinę (S-formos) arba hiperbolinio tangento (*tanh*) funkciją, tačiau tam gali būti naudojama bet kokia tiesinė arba netiesinė funkcija. Perdavimo funkcijos tikslas yra apriboti išvedamų duomenų dydžius, kurie gali paralyžuoti neuroninius tinklus, ir tokiu būdu nutraukti neuroninių tinklų mokymosi procesą.

*Pradiniai svoriai.* Jie paprastai pasirenkami nedideli ir teigiami ir priskiriami atsitiktinai kiekvienam elementui.

Neuroninių tinklų apmokymo esmė yra surasti tokius svorius neuronams, kad galima būtų minimizuoti paklaidos funkciją. Globalaus minimumo radimas ne visada pavyksta, kadangi paklaidos paviršius gali turėti daugybę vietinių minimumų; tada, algoritmas gali tiesiog „pakibti“.

### **1.1.3. Neuroninių tinklų taikymai kredito rizikos vertinimui**

Ši sritis kredito rizikos vertinimui pradėta naudoti seniausiai iš čia apžvelgiamų metodų – remiantis A. Vellido, P.J.G. Lisboa, J. Vaughan 1999 m. atlikta studija, galima pastebėti, kad pirmą kartą neuroniniai tinklai kredito rizikos vertinimui panaudoti dar 1992 m. Iš jos galima pastebėti,

kad dauguma tyrimų paremti daugiasluoksnių perceptronų tinklų, kuriuose išeigos iš visų neuronų eina tik į sekančius, bet ne į prieš tai buvusius neuronus (*feedforward multilayer perceptron*, sutr. *MLP*), naudojimu; šių tinklų apmokymui naudojamas *back propagation* metodas, kurio esmė – gradientinės informacijos naudojimas tinklo svorių modifikavimui, leidžiantis sumažinti klaidos funkcijos vertę (Vellido, Lisboa, Vaughan, 1999). Reikia paminėti ir tai, kad šis metodas laikomas neoptimaliu ir netgi neefektyviu (Piramuthu, Shaw, & Gentry, 1994; Lenard, Alam, Madey, 1995 teigimu). Kai kurie autoriai naudojo ir saviorganizuojančių tinklų metodą (*self-organizing maps*) bei šio metodo įvairius variantus. Kaip galima pastebėti iš lentelės, šiais metodais sprendžiami keli klausimai, susiję su kredito rizika. Du pagrindiniai iš jų yra bankroto prognozavimas, bandantis įvertinti įmonės finansinę būklę bei nustatyti, kaip greitai galima pastebėti silpnąją vietą veikloje, bei kredito įvertinimas (*credit evaluation*, taip pat *credit scoring*), kurio esmė yra bandymas suklasifikuoti organizacijas bei asmenis, imančius kreditą, į „blogus“ bei gerus. Priėjimas prie duomenų šios srities tyrimui yra ribotas, todėl tik keletas autorių (Richeson, Zimmerman, & Barnett, 1994; Williamson, 1995; Jagielska & Jaworski, 1996; Desay & Crook, 1996; Torsun, 1996; Glorfeld & Hardgrave, 1996) dirbo su realiais duomenimis, kurie apėmė nuo 40 iki 310000 paskolų duomenis. Šiuose darbuose naudojama nuo 6 iki 27 kintamųjų, tačiau tik keli autoriai naudojo kintamųjų pasirinkimo metodiką (Williamson, 1995; Glorfeld, 1996; Glorfeld & Hardgrave, 1996); taip pat tik Glorfeld (1996), Glorfeld ir Hardgrave (1996) bei Desay ir Crook (1996) naudojo kryžminio tikrinimo metodus. Kaip dar vieną iš trūkumų studijos autoriai A. Vellido, P.J.G. Lisboa, J. Vaughan išskiria tą faktą, kad buvo atlikta nedaug tyrimų atskirų kredito rizikos sričių, tokių, kaip pardavimo kredito rizika, vertinimui.

Šiuose darbuose tiriami įvairūs statistiniai kredito įvertinimo metodai, jų efektyvumą lyginant su neuroninių tinklų metodo taikymo rezultatais, tokie, kaip diskriminantinė analizė, tiesinė ir logistinė regresijos, matematiniai programavimo metodai, rekursinis skirstymas (sprendimų medžiai), neparametrinio išlyginimo metodai (pavyzdžiui, artimiausio kaimyno), laiko kitimo ir kt. metodai. 1 lentelė iliustruoja neuroninių tinklų pritaikymo ir jų kryžminimo su kitais statistiniais ir dirbtinio intelekto metodais, naudojant įvairias paradigmas, galimybes.

Iš šioje lentelėje pateikiamų darbų dar galima paminėti Besens, Van Gestel, Stepanova, Van den Poel darbą, kuriame galima rasti ir prieš tai tirtų neuroninių tinklų išlikimo analizės metodų palyginimą, kuriuo remiantis, autoriai pateikė išvadą, kad geriausiai tinkami yra Faragi, Mani ir Brown požiūriai. Jų gauti rezultatai parodė, kad geriausiai modeliavimui tinkamas yra daugiasluoksnių perceptrono modelis, o pagreitinimo filtruojant modelio našumas didesnis nei rinkinio vidurkio modelis.

1 lentelė. Neuroninių tinklų taikymo kredito rizikos vertinimui tyrimų pavyzdžiai

Autoriai	Metai	Aprašymas	Nuoroda
<b>Glorfeld, Hardgrave</b>	1996	Naudoti NN su Gauso maksimalaus tinkamumo klasifikatoriais	(Vellido, Lisboa, Vaughan, 1999)
<b>Deng</b>	1993	Sinergetinis naujų faktų gavimo iš esamų atmintyje ( <i>Memory-Based Reasoning</i> ) bei taisyklėmis paremto induktyvaus mokymosi ( <i>Rule-Based Inductive Learning</i> ) metodų integravimas su neuroniniais tinklais į ekspertinę sistemą	(Vellido, Lisboa, Vaughan, 1999)
<b>Williamson</b>	1995	Genetinio algoritmo, automatiškai parenkančio optimalią konfigūraciją bei rezultatus, panaudojimas	(Vellido, Lisboa, Vaughan, 1999)
<b>Arminger, Enache, Bonne</b>	1997	Klasifikavimo metodų apjungimas	(Vellido, Lisboa, Vaughan, 1999)
<b>Handzic, Tjandrawibawa, Yeo</b>	2003	Naudotas daugiasluoksnio perceptrono neuroninio tinklo tipas kartu su vidurkinimu ( <i>Ensemble Averaging</i> ) bei pagreitinimu filtruojant ( <i>Boosting by Filtering</i> ), kas leido gauti tikslius rezultatus (paskutiniu atveju procentinė klaida siekė 1.32 %)	(Handzic, Tjandrawibawa, Yeo, 2003)
<b>di Tollo</b>	2006	Naudota mokymo su mokytoju paradigma	(di Tollo, 2006)
<b>Besens, Van Gestel, Stepanova, Van den Poel</b>	2004	Naudota "išlikimo analizės" paradigma, paremta tam tikros populiacijos išlikimo rodikliu	(Besens, Van Gestel, Stepanova, Van den Poel, 2004)
<b>Keung Lai, Yu, Wang, Zhou</b>	2006	Naudojamas trijų fazių neuroninių tinklų metodas blogų ir gerų kreditorių atskirumui. Jame naudojama daug skirtingų neuroninių tinklų modelių, kuriuos sukūrus, taikomas iškoreliavimo maksimizavimo algoritmas tinkamų aibės narių išrinkimui; čia klasifikavimui naudojamas ir patikimumu paremtas metodas. Kaip teigia patys autoriai, atlikę eksperimentą, jie gavo gerus klasifikavimo rezultatus	(Keung Lai, Yu, Wang, Zhou, 2006)
<b>Keung Lai, Yu, Wang, Zhou</b>	2006	Šis darbas traktuoja neuroninius tinklus kaip metamokymo metodą, kurio esmė yra skirtingų neuroninių tinklų pagal skirtingus apsimokymo duomenų rinkinius su skirtingomis pradinėmis sąlygomis bei apsimokymo algoritmais apmokymas skirtingų kredito vertinimo modelių (bazinių modelių) formulavimui. Iš šių bazinių modelių gali būti generuojamas metamodelis, kas leidžia pagerinti patikimumą, t.y., tikslumą prognozuojant praradimus	(Keung Lai, Yu, Wang, Zhou, 2006)

Šaltinis: sudaryta autoriaus.

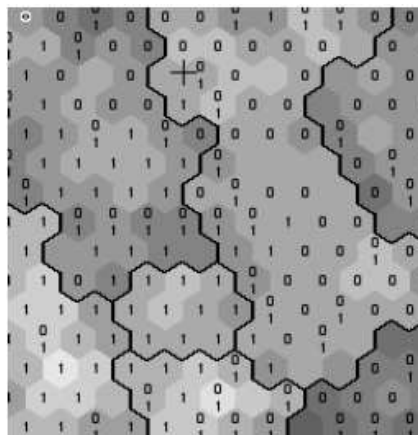
Savo atliktoje neuroninių tinklų panaudojimo versle bei finansuose studijoje A. Vellido, P.J.G. Lisboa, J. Vaughan taip pat išskyrė ir apibendrino kai kuriuos dirbtinių neuroninių tinklų privalumus bei trūkumus, kurie su M.Dunham pateikiamais privalumais ir trūkumais pateikiami 2 lentelėje.

Privalumai	Trūkumai
tinkami darbui su nepilnais, trūkstamais ar neaptriukšmintais duomenimis	sudėtingumas bei neaiškumas (“juoda dėžė”) – galimybės įvertinti tarpusavyje nepriklausomų kintamųjų svarbą bei sugeneruoti taisyklių rinkinį modelio veikimo aprašymui nebuvimas
tai neparamestrinis metodas, nereikalaujantis jokių pirminių skirstinio ar duomenų susiejimo įvertinimų	“taisyklių ištraukimo” problema
galimybė susieti bet kokią sudėtingą netiesiškumą bei aproksimuoti bet kurią sudėtingą funkciją	Modelio architektūros parinkimas bei per didelis prisitaikymas ( <i>overfitting</i> )
našumo gerinimas besimokant; tai gali tęstis netgi po to, kai pateikti apmokymo duomenys	Įvairūs modelio pasirinkimo aspektai
Galimybė naudoti lygiagrečiai ir taip pasiekti geresnį našumą	Sudėtingas testavimas ir verifikavimas
Nedidelis klaidos rodiklis ir kartu aukštas tikslumas, tinkamai atlikus apmokymą	

Šaltinis: sudaryta autoriaus pagal (Vellido, Lisboa, Vaughan, 1999; Dunham, 2003)

Vienas iš naujausių dirbtinių neuroninių tinklų tipų, neseniai pradėtas tirti ir taikyti finansuose, yra saviorganizuojantys tinklai (angl. *self-organizing maps*, sutr. SOM). Šio metodo autorius yra Suomijos profesorius Tim Kohonen, todėl šis neuroninių tinklų tipas dar žinomas kaip Kohoneno žemėlapis. Šio metodo esmė yra apmokymas, naudojant nestebimą mokymąsi, sukuriantis duomenų nedidelio dimensijų skaičiaus atvaizdą (žemėlapi), tuo pat metu išsaugant įeigos erdvės topologines savybes. Tai leidžia patogiai realizuoti nedidelio dimensijų skaičiaus atvaizdavimus su didelio dimensijų skaičiaus duomenimis. Šioje srityje taip pat buvo atlikti tyrimai – tiek Roterdamo Erasmus universitete (Van der Berg, 2002), tiek ir kituose.

Svarbu paminėti, kad šie tyrimai buvo atlikti ir Vilniaus universitete (Merkevičius, Garšva, Simutis, 2004). 6 pav. pateikiamas jų tyrimo rezultatų pavyzdys, kuriame pavaizduoti kredito klasių, žymimų kaip „gera“ ir „bloga“, klasteriai. Priklausomai nuo „gerų“ ir „blogų“ klasių skaičiaus klasteryje, jis traktuojamas kaip „geras“ ir atvirkščiai (Merkevičius, Garšva, Simutis, 2004).



Šaltinis: E. Merkevičius, G. Garšva, R. Simutis. Forecasting Of Credit Classes With The Self-Organizing Maps

6 pav. SOM žemėlapiu pritaikymo klasifikavimui pavyzdys

Reikia atsižvelgti ir į tai, kad SOM metodas gali būti taikomas tiek finansinių ataskaitų tyrimui bei bankroto prognozavimui (Kaski, 2001), tiek kredito klasių sudarymui bei prognozavimui (Merkevičius, Garšva, Simutis, 2004). Taip pat galima paminėti ir daugiau finansų ir ekonomikos sričių, kuriuose šis metodas gali būti pritaikytas - finansinė analizė bei prognozavimas, nesėkmių nuspėjimas, finansinių instrumentų reitingavimas, investavimo galimybių analizė, komercinė bei šalies kredito rizikos analizė, finansinės veiklos stebėjimas, ekonomikos kryptių analizė, vartotojų profiliavimas bei reitingavimas, marketingas ir kt. (Deboeck, 1998).

#### **1.1.4. Atramos vektorių mašinos (*Support Vector Machines*)**

Atramos vektorių mašinos (angl. *Support Vector Machines*, sutr. SVM) metodas, kaip ir neuroniniai tinklai, yra metodas, pagrįstas mašininio mokymusi iš pavyzdžių. Šis algoritmas sukurtas 7-ame dešimtmetyje, jo autorius rusų mokslininkas Vladimir Vapnik. SVM metodas taip pat yra ir statistinio mokymo teorijos, aprašančios besimokančių mašinų savybes, leidžiančias gerai atlikti apibendrinimus pagal dar nematytus joms duomenis, dalis, pastaruosius tris dešimtmečius kurtos ir tobulintos Vapnik ir Chervonenkis. SVM algoritmas buvo sukurtas AT&T Bell kompanijos laboratorijose ir buvo orientuotas pramoniniams sprendimams. Šis metodas leido greitai pasiekti puikius rezultatus tokiose srityse, kaip regresija, laiko eilučių prognozė, atpažinimo sistemos ir kt. Nepaisant to, kad šis klasifikavimo metodas pradėtas naudoti palyginti neseniai, jau galima rasti jo pritaikymų įvairiose srityse, nuo bioinformatikos iki teksto klasifikavimo bei vaizdų atpažinimo.

##### **1.1.4.1. Pagrindiniai SVM teoriniai principai**

Atramos vektorių mašinos – tai besimokančios mašinos (modeliai), galinčios atlikti binarinę klasifikaciją (modelių atpažinimą) bei realios vertės funkcijų aproksimaciją (regresijos įvertinimą). SVM netiesiogiai susieja jų  $n$ -matavimų įėjimo duomenų erdvę su daugelio matavimų galimybių erdve, kurioje sukuriamas tiesinis klasifikatorius. Jų speciali savybė yra ta, kad jos tuo pačiu metu minimizuoja empirinę klasifikavimo klaidą ir maksimizuoja geometrinę ribą, dėl ko ji vadinama maksimalios ribos klasifikatoriumi.

Dažniausiai nagrinėjama modelių atpažinimo ir suskirstymo į dvi klases problema. Šiuo atveju siekiama įvertinti funkciją  $f: X \rightarrow \{\pm 1\}$ , siejančią įėjimo ir išėjimo duomenis. Tarkime, kad yra duoti empiriniai duomenys  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in \mathcal{X} \times \{\pm 1\}$ , kurioje  $\mathcal{X}$  yra netuščia aibė, iš kurios imamos  $x_i$  struktūros;  $y_i$  vadinami žymėmis ir tikslais (*targets*). Tai reiškia, kad, turėdami kažkokią naują struktūrą  $x \in \mathcal{X}$ , turime užduotį rasti atitinkamą  $y \in \{\pm 1\}$ ; kitaip tariant, turime pasirinkti tokį  $y$ , kad  $(x, y)$  būtų kaip galima panašesnė į mokymui naudotus pavyzdžius. Taigi šiuo atveju reikalingi  $\mathcal{X}$  ir  $\{\pm 1\}$  panašumo matavimai. Formaliai jie gali būti aprašyti kaip funkcija  $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}, (x, x') \rightarrow k(x, x')$ , kuri pagal duotus pavyzdžius  $x$  ir  $x'$  grąžina realų skaičių, charakte-

rizuojantį jų panašumą. Ši funkcija vadinama *branduoliu* arba *kerneliu* (angl. *kernel*).

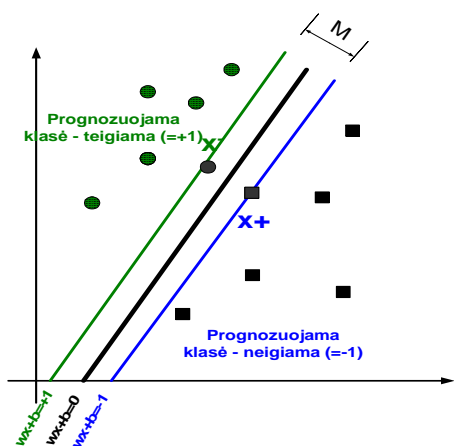
Kaip panašumo matavimo tipas gali būti naudojamos taškinės sandaugos (angl. *dot products*); pavyzdžiui, dviejų vektorių  $x$  ir  $x'$  taškinė sandauga aprašoma kaip  $(x \cdot x') := \sum_{i=1}^N (x_i)(x'_i)$ , kur  $x_i$  ir  $x'_i$  žymi atitinkamai vektorių  $x$  ir  $x'$  i-tuosius elementus. Geometriškai šią sandaugą galima išvaizduoti kaip kampo tarp vektorių  $x$  ir  $x'$ , kurių ilgis normalizuotas iki 1, kosinusą. Tai leidžia apskaičiuoti atstumą tarp dviejų vektorių kaip skirtumo vektoriaus ilgį, be to, tokių sandaugų skaičiavimas naudoti visas geometrines konstrukcijas, kurios gali būti aprašytos kampo, ilgio ir atstumo terminais. Norint naudoti duomenų sandaugą su taškiniais vektoriais, juos reikia atvaizduoti tam tikroje taškinės sandaugos erdvėje  $F$ , turinčioje mažiau matavimų nei vektorinė erdvė  $\mathbb{R}^N$ . Formaliai tai aprašoma kaip funkcija  $\Phi: \mathcal{X} \rightarrow F, x \rightarrow \bar{x}$ . Tokiu atveju sakoma, kad duomenų aibė  $\mathcal{X}$  atvaizduota savybių erdvėje (angl. *feature space*)  $F$ . Toks duomenų atvaizdavimas turi tris privalumus (Scholkopf, 2000):

1. Taškinės sandauga erdvėje  $F$  leidžia apibrėžti panašumo matavimą

$$k(x, x') := (\bar{x} \cdot \bar{x}') = (\phi(x) \cdot \phi(x')) \quad (6)$$

2. Atsiranda galimybė su duomenimis dirbti geometriniu atžvilgiu, t.y., naudojant tiek tiesinę algebrą, tiek analitinę geometriją.

3. Galimybė pasirinkti susiejimą  $\Phi$  leidžia sukurti daugybę mokymosi algoritmų, t.y., pritaikius netiesinį susiejimą, galima pakeisti duomenų atvaizdavimą į tokį, kuris labiau tinka sprendžiamai problemai ar taikomam algoritmui.



„Teigiama“ plokštuma:  $\{x: wx+b = +1\}$   
 „Neigiama“ plokštuma:  $\{x: wx+b = -1\}$   
 Klasifikatoriaus riba:  $\{x: wx+b = 0\}$

$x^-$  - „neigiamos“ plokštumos taškas  
 $x^+$  - „teigiamos“ plokštumos taškas, artimiausias  
 $x^-$   
 $M$  - ribos plotis ( $|x^+ - x^-| = M$ )

Galimos taškų (klasių) reikšmės:  
 $+1$ , jei  $wx+b \geq 1$   
 $-1$ , jei  $wx+b \leq -1$

Šaltinis: sudaryta autoriaus.

### 7 pav. Tiesinė atraminių vektorių mašina

Tarkime, turime mašiną, kurios užduotis yra išmokti susiejimą  $x_i \rightarrow y_i$ . Pati mašina šiuo atveju apibrėžiama kaip aibė galimų susiejimų  $x \rightarrow f(x, \alpha)$ ; čia funkcijos  $f(x, \alpha)$  yra pažymėtos tam

tikrais parametrais  $\alpha$ . Tokia mašina determinuota - kiekvienam įvestam duomenų vienetui  $x$  ir tam tikram parametrai  $\alpha$  visada pateikia tą patį rezultatą  $f(x, \alpha)$ .

7 pav. pateikiamas paprasčiausias tiesinės SVM pavyzdys ir grafinė metodo iliustracija. Šis pavyzdys atspindi klasifikavimą į teigiamus ir neigiamus elementus pagal tam tikrus požymius.

Laukiama testavimo paklaida (rizikos dydis) apmokytai mašinai apibrėžiama taip:

$$R(\alpha) = \int \frac{1}{2} |y - f(x, \alpha)| dP(x, y) \quad (7)$$

Empirinės rizikos (treniravimo paklaidos) dydis SVM mašinai nustatomas pagal formulę

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x_i, \alpha)| \quad (8)$$

#### 1.1.4.2. SVM praktiniai aspektai ir modifikacijos

Kaip jau minėta, pagrindinė SVM paskirtis yra klasifikavimo uždaviniai, nors, kaip ir kiti mašininio mokymo metodai, jis gali būti taikomas ir regresijos uždavinių sprendimui. Šiuo atveju apsiribojama klasifikavimo uždavinio sprendimu. Pats klasifikatorius apibūdinamas kaip skiriančioji hiperplokštuma (angl. hyperplane), kurios abejose pusėse yra binariniai sprendiniai (t.y., sprendiniai, atitinkantys tik 0 arba 1). Tai iliustruoja ir 13 pav. SVM tikslas rasti tokią hiperplokštumą, kuri minimizuotų klaidos ribą. Ši hiperplokštuma aprašoma atramos vektorių aibe, kuriems (ir tik kuriems) Lagrandžo daugikliai nelygūs nuliui. Šių vektorių radimas iš apmokymo duomenų apima ir kvadratinių optimizavimo uždavinių sprendimą.

SVM formuluoatės naudoja tiesinį atskyrimą (kaip matyti 7 pav.), kuris ne visada galimas. Tokiu atveju turi būti naudojamas netiesinis SVM apmokymas, kuris tiesinės SVM mašinos apmokymą papildo dar vienu žingsniu – įvedamų duomenų erdvės transformavimą į kitą vektorinę erdvę (paprastai daug didesnio matavimų skaičiaus) tokiu būdu, kad transformuota duomenų erdvė gali būti atskiriama tiesiškai, t.y., kad egzistuoja tiesinė sprendimo riba (hiperplokštuma), kuri gali atskirti teigiamas ir neigiamas reikšmes transformuojamoje erdvėje. Šiam transformavimui naudojamos įvairios *branduolio* funkcijos; pačios populiariausios iš jų:

- tiesinė:  $K(x, z) = x \cdot z$ ;
- polinominė:  $K(x, z) = (\langle x \cdot z \rangle + \theta)^d$ ;
- radialinių bazių funkcija (Radial Basis Function arba RBF):  $K(x, z) = e^{-\frac{|x-z|^2}{2\sigma}}$ ; kur  $\sigma > 0$ ;
- sigmoidinė:  $K(x, z) = \tanh(\beta \langle x, z \rangle - \lambda) = \frac{1}{1 + e^{-\beta \langle x, z \rangle - \lambda}}$ ; kur  $\beta, \lambda \in R$

Taip pat išskiriami keli SVM modeliai; detalesnė jų klasifikacija pateikiama lentelėje. Čia plačiau nagrinėjamas vienas iš populiariesnių SVM variantų –  $\nu$ -SVM, turintis nu parametras su  $[0, 1]$  ribomis. Šis parametras susijęs su atraminių vektorių skaičiaus ir apmokymo klaidos rodikliais

(LibSVM FAQ).

Formaliai metodus aprašomas taip (Chang, Lin): duoti apmokymo vektoriai  $X_i \in R, i = 1, \dots, l$ , dvi klasės, bei vektorius  $y \in R^l$  toks, kad  $y_i \in [-1; 1]$ ; pirminė forma aprašoma kaip

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \rho} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} - \nu \rho + \frac{1}{l} \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq \rho - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l, \rho \geq 0. \end{aligned} \quad (9)$$

Atitinkamai kita formuluotė aprašoma taip (Chang, Lin):

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha \\ \text{subject to} \quad & 0 \leq \alpha_i \leq 1/l, \quad i = 1, \dots, l, \\ & \mathbf{e}^T \alpha \geq \nu, \quad \mathbf{y}^T \alpha = 0. \end{aligned} \quad (10)$$

kur  $\mathbf{e}$  yra visų vienetų vektorius,  $Q$  yra  $l \times l$  teigiama pusiau apibrėžta matrica,  $Q_{ij} \equiv Y_i Y_j K(x_i, x_j)$  ir  $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$  yra branduolio funkcija.

Sprendimo funkcija yra  $\text{sgn}(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b)$ . Crisp ir Burges, 2000 bei Chang, Lin, 2001

įrodė, kad  $\mathbf{e}^T \alpha \geq \nu$  gali būti pakeista į  $\mathbf{e}^T \alpha = \nu$ . Tokiu atveju sprendžiama problema susiveda į toki uždavinį:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha \\ \text{subject to} \quad & 0 \leq \alpha_i \leq 1, \quad i = 1, \dots, l, \\ & \mathbf{e}^T \alpha = \nu l, \\ & \mathbf{y}^T \alpha = 0. \end{aligned} \quad (11)$$

Apskaičiavus  $\alpha / \rho$ , turime tokią sprendimo funkciją:  $\text{sgn}(\sum_{i=1}^l y_i \alpha_i / \rho) K(x_i, x) + b$ . Tada gauname  $y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) = \pm 1$ , kuri yra tokia pati, kaip klasikinio C-SVC klasifikatoriaus atveju.

3 lentelė. SVM algoritmai bei jų realizacijos

Pavadinimas	Autoriai	Aprašymas
<b>C-SVC</b>	Boser et al., 1992 Cortes ir Vapnik, 1995	Klasifikavimo metodas, naudojantis $C > 0$ kaip viršutinę ribą
<b>v-SVC</b>	Scholkopf et al., 2000	Klasifikavimo metodas, naudojanti $\nu$ parametą atraminių vektorių skaičiaus ir apmokymo klaidos kontrolei
<b>Vienos klasės SVM</b>	Scholkopf et al., 2001	Metodas, skirtas multidimensinio skirstinio įvertinimui
<b><math>\epsilon</math>-Support Vector Regression (<math>\epsilon</math>-SVR)</b>	Vapnik, 1998	Skirtas regresinės lygties parametų skaičiavimui
<b>v-Support Vector Regression (<math>\nu</math>-SVR)</b>	Scholkopf et al., 2000	Skirtas regresinės lygties parametų skaičiavimui; naudoja $\nu$ parametą atraminių vektorių skaičiui.
<b>Least Squares Su-</b>	J.A.K. Suykens, T. Van	SVM metodas, smarkiai susijęs su reguliarizacijos tink-



Pavadinimas	Autoriai	Aprašymas
<b>Support Vector Machines (LS-SVM)</b>	Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, 2002	lais, Gauso procesais bei branduolio Fišerio diskriminantine analize. Kainos funkcija yra reguliarizuota mažiausių kvadratų funkcija, kurios naudojimas veda prie tiesinių Karush-Kuhn-Tucker sistemų.
<b>Lagrangian Support Vector Machine (LSVM)</b>	O.L. Mangasarian and D. R. Musicant, 2000	Paremtas tiesinės SVM standartinio kvadratinio programavimo Lagranžo formuluotės performulavimu, kas leidžia diferencijuojamos neapribotos iškilosios funkcijos $m$ matavimų erdvėje minimizavimu, čia $m$ – taškų skaičius, kuriuos reikia suklasifikuoti duotoje $n$ matavimų erdvėje.
<b>Active Support Vector Machines (ASVM)</b>	O.L. Mangasarian and D. R. Musicant, 2000	Paremtas aktyvių aibių strategija kvadratinio programavimo su ribomis problemos sprendimui. Šiuo atveju maksimizuojamas atstumas tarp hiperplokštumų bei ribos paklaida minimizuojama naudojant 2-norm atstumo funkcijos kvadratą
<b>Smooth Support Vector Machine (SSVM)</b>	Yuh-Jye Lee, O. L. Mangasarian, 2001	Naudoja tradicinio kvadratinio programavimo vienodo neapriboto optimizavimo performulavimą, kuris sprendžiamas naudojant labai greitą Newton-Armijo algoritmą ir išplėstas į netiesinių atskyrimo paviršių naudojimą, tam naudojant netiesinius branduolio funkcijos metodus.
<b>Newton Method for LP Support Vector Machine (LPSVM)</b>	G. Fung, O. L. Mangasarian, 2001	LPSVM naudoja greitą Niutono metodą, nuslepiančią įėjimo duomenų erdvės savybes. Šis metodas apėsti klasifikavimo problemas labai didelių matavimų erdvėse ir generuoja klasifikatorių, priklausantį nuo labai mažai įėjigos savybių.
<b>Proximal Support Vector Machine (PSVM)</b>	G. Fung, O. L. Mangasarian, 2001	Taškai klasifikuojami, juos priskiriant prie vienos iš dviejų artimiausių lygiagrečių plokštumų.
<b>SVM Light</b>	Thorsten Joachims	Klasikinio SVM metodo realizacija C kalba, naudojanti greitą optimizavimo algoritmą, sprendžianti klasifikavimo, regresijos bei reitingavimo (ranking) problemas, galinti efektyviai dirbti su daugybe atraminių vektorių ir šimtais tūkstančių duomenų.

Šaltinis: sudaryta autoriaus.

Šiame darbe bus remiamasi tik  $v$  –SVC klasifikavimo metodo taikymu nagrinėjamai problemai.

#### 1.1.4.3. SVM metodo taikymas kredito rizikos vertinimui

Sukurti ir keli paramos vektorių mašinos metodo pritaikymai kredito rizikos vertinimui; čia jis naudojamas kartu su neuroniniais tinklais (Zan Huang, Hsinchun Chena, Chia-Jung Hsua, Wun-Hwa Chenb, Soushan Wu, 2003) arba kartu naudojant ir fuzzy logiką (Yongqiao Wang, Shouyang Wang, K. K. Lai, 2005). Pirmuoju atveju naudojami klaidos sklidimo atgal neuroniniai tinklai (*backpropagation networks*); čia sukuriamas standartinis trijų sluoksnių tinklas, kuriame įėjigos sluoksnio elementai yra finansiniai kintamieji (iš viso naudojamas 21 kintamasis), išėjigos elementai yra vertybinių popierių vertinimo klasės, o paslėptų sluoksnių elementų skaičius yra (*įėjigos elemen-*

*tų skaičius – išeigos elementų skaičius*)/2. Tokia architektūra kartu su SVM leidžia pasiekti optimalius rezultatus; kaip teigia autoriai, SVM pagal anksčiau atliktus tyrimus našumu aplenkė kitus klasifikatorius, tokius, kaip neuroniniai tinklai bei tiesinis diskriminantinis klasifikatorius. Antruoju atveju sukuriama *fuzzy* paramos vektorių mašina, kuri palyginti su standartine SVM bei mažiausių kvadratų SVM, leidžia sumažinti mažiau svarbių duomenų taškų jautrumą. Dar vieni autoriai kredito rizikos vertinimui naudojo ir mažiausių kvadratų metodu paremtą SVM (Keung Lai, Yu, Zhou, Wang, 2005). Lyginant su standartine (Vapnik, pagal šio metodo autoriaus pavardę) SVM, mažiausių kvadratų SVM gali transformuoti kvadratinio programavimo problemą į tiesinio programavimo problemą, taip sumažindama skaičiavimų sudėtingumą. Šio tyrimo autoriai naudojo 14 kintamųjų, tokių, kaip individo pajamos bei išlaidos, lytis, amžius ir kt. Tyrimui buvo naudoti 1225 individų duomenys, iš kurių 323 buvo „blogi“ kreditoriai; šiuo atveju pasiektas bendras 89% tikslumas, kurį tyrimo autoriai įvertino kaip gerą. Paramos vektorių technologija panaudota ir bankroto prognozavimui (Kyung-Shik Shin, Taik Soo Lee, Hyun-jung Kim, 2004; Jae H. Min, Young-Chan Lee, 2005); pirmuoju atveju autoriai pateikia išvadas, kad SVM našumu aplenkia backpropagation tipo neuroninius tinklus. Antruoju atveju SVM branduolio funkcijos optimalių parametrų reikšmių radimui naudojamas paieškos tinkle (angl. *grid-search*) metodas. Kaip ir prieš tai minėtame, taip ir šiame darbe autoriai įrodė SVM metodo našumą, taigi, remiantis šių autorių darbais, galima daryti išvadą, kad paramos vektorių metodo taikymas kredito rizikos vertinime turi didelę perspektyvą.

#### 1.1.5. **Fuzzy logika ir fuzzy ekspertinės sistemos**

*Fuzzy* logika, dar kitaip vadinama neraiškioji logika - tai tokia logika, kurios reikšmės nėra vien „teisinga“ arba „klaidinga“. Šioje logikoje teiginiai gali būti nevisiškai teisingi arba iš dalies klaidingi. Bulio (iš ang. k. *boolean*) logika yra paremta tik „teisingas-klaidingas“ teiginiais, tačiau ši logika neatitinka žmogaus samprotavimų. Pavyzdžiui, teiginys „šiandien yra saulėta diena“ gali būti 100% teisingas jei nėra nei vieno debesėlio, 80% teisingas jei danguje yra keletas debesų, 50% jei dangus nėra visiškai giedras – miglotas ir 0% jei šiandien lyja. Tuo tarpu Bulio logikoje šis teiginys būtų vien tik klaidingas arba vien tik teisingas. Taigi *fuzzy* logika labiau atspindi žmogaus mąstymo logiką.

Kaip ir įprastinėje logikoje, taip ir šioje galioja neigimo, konjunkcijos ir disjunkcijos dėsniai, tačiau pačios taisyklės skiriasi (jų pavyzdys pateikiamas 2 priede). Jei paprastojoje logikoje  $p \in \{0;1\}$ , tai neiškiojoje logikoje  $p$  gali įgyti reikšmes nuo 0 iki 1, t.y.,  $p \in [0;1]$ .

Veiksmai su *fuzzy* logikos kintamaisiais aprašomi taip:

$$\text{truth}(\text{not } x) = 1.0 - \text{truth}(x)$$

$$\text{truth}(x \text{ and } y) = \text{minimum}(\text{truth}(x), \text{truth}(y))$$

$$\text{truth}(x \text{ or } y) = \text{maximum}(\text{truth}(x), \text{truth}(y))$$

Panašios operacijos atliekamos ir su *fuzzy* reikšmių aibėmis (*fuzzy sets*). 4 priedas iliustruoja Bulio logikos ir *fuzzy* logikos pagrindines operacijas bei jų skirtumus.

#### 1.1.5.1. Fuzzy ekspertinės sistemos

Pradėjus taikyti fuzzy logiką realių situacijų modeliavimui bei konkrečių sričių uždavinių sprendimui, atsirado ir naujas ekspertinių sistemų tipas, naudojantis tiek funkcijas, tiek ir taisykles. Taisyklės jose paprastai apibrėžiamos kaip *if... then* tipo sąlyginiai sakiniai, nusakantys tam tikras kitimo ar rezultatų išvedimo sąlygas. Kaip tokio sakinio pavyzdį galima pateikti sąlygą: jei *x* yra mažas ir *y* yra didelis, tai *z* – vidutinis; čia *x* ir *y* yra įvedami kintamieji, o *z* – išvedami duomenys (if *x* is MAZAS and *y* is DIDELIS then *z* is VIDUTINIS). Šiuo atveju „MAZAS“ yra funkcija (*fuzzy* logikos poaibis)  $f(x)$ , „DIDELIS“ atitinka funkciją  $f(y)$ , o „VIDUTINIS“ –  $f(z)$ . Sakinio dalis nuo „jei“ (*If*) iki „tada“ (*Then*) vadinama taisyklės prielaida, o sakinio dalis, prasidedanti nuo „tada“ (*Then*) - taisyklės išvada. Dauguma *fuzzy* ekspertinių sistemų leidžia daryti daugiau nei vieną išvadą vienai taisyklei. Paprastai *fuzzy* ekspertinė sistema turi daugiau nei vieną taisyklę. Taisyklių grupė vadinama *žinių baze*.

Turint taisyklės apibrėžimą ir funkcijas reikia sužinoti, kaip jas pritaikyti atitinkamiems įvedimo kintamiesiems, kad galima būtų gauti išvedimo kintamuosius. Šis procesas vadinamas *išvadų darymu*. *Fuzzy* ekspertinėse sistemose išvadų darymas susideda iš 4 pagrindinių etapų: fuzifikacija, išvadų darymas, komponavimas ir defuzifikacija. Pastarasis procesas yra nebūtinasis.

1. *Fuzifikacija*. Šiame procese kintamiesiems, kurių funkcijos buvo apibrėžtos, suteikiamos atitinkamos reikšmės, tam, kad būtų galima apibrėžti kiekvienos taisyklės prielaidos tiesos laipsnį. Jei tiesos laipsnis nėra nulinė reikšmė, tada taisyklė vykdo atitinkamus skaičiavimus.

2. *Išvadų darymas*<sup>1</sup>. Šiame žingsnyje tiesos laipsnis kiekvienos taisyklės prielaidoje yra apskaičiuojamas ir naudojamas kiekvienos taisyklės išvadų darymo dalyje. Rezultatai – neaiškus (*fuzzy*) poaibis, priskirtas kiekvienos taisyklės išvedamiems kintamiesiems. Yra du išvadų darymo metodai: MIN ir PRODUCT. MIN išvadų darymo metodo atveju išvedimo funkcija yra „iškerpama“ (*clipped off*) aukščiausiam tiesos laipsnio lygyje, kas atitinka *fuzzy* logikos AND operatorių. PRODUCT metode išvedimo funkcija yra matuojama (*scaled*) pagal prielaidos tiesos laipsnį.

3. *Struktūrizavimas*. Šiame procese visi *fuzzy* poaibiai priskiriami išvedimo kintamiesiems ir yra apjungiami kartu tam, kad sudarytų vieną poaibį kiekvienam išvedimo kintamajam. Yra dvi komponavimo taisyklės: MAX ir SUM komponavimai. MAX komponavime apjungtas išvedimo poaibis sudaromas, imant maksimumus iš visų *fuzzy* poaibių. SUM komponavime apjungtas išvedimo poaibis sudaromas imant tikslią visų *fuzzy* poaibių sumą. Šių dviejų metodų rezultatai gali būti

---

<sup>1</sup> Išvadų darymo ir struktūrizavimo pavyzdys pateiktas 4 priede

didesni nei 1, todėl SUM naudojamas tik tada, kada yra naudojamas defuzifikavimo žingsnis.

4. *Defuzifikavimas*. Po komponavimo proceso kartais tiesiog naudinga patikrinti *fuzzy* poaibio rezultatus, bet tam dažniausiai reikia paversti neaiškią reikšmę į aiškią. Tai atlieka defuzifikavimo procesas, kuriam yra daugybė defuzifikavimo metodų. Dažniausiai naudojami centroido (kai aiški reikšmė yra suskaičiuojama randant tokio kintamojo reikšmę, kuri sudarytų funkcijos svorio centrą) ir maksimumų (kai randama tokia reikšmė, kuri labiausiai atitinka tiesos laipsnį) metodai. Kartais šis procesas apjungiamas su komponavimo procesu, tam pasinaudojus matematinėmis prielaidomis, kurios supaprastina skaičiavimą.

*Fuzzy* ekspertinės sistemos naudojamos tiesinei ir netiesinei kontrolei, struktūrų atpažinimui, sprendimų priėmimui, proceso imitavimui bei finansinėse sistemose. *Fuzzy* logika naudoja lingvistinius kintamuosius, o tai į analizę leidžia įtraukti tokius faktorius, kaip pvz., politikos įtaka, kurie paprastai nenaudojami prognozavime.

Pati *fuzzy* logika viena retai kur taikoma. Dažniausiai ji taikoma kartu su neuroniniais tinklais, genetiniais algoritmais, nes apjungus keletą metodų, galima gauti geresnius rezultatus. Dažnai yra naudojamos *neuro-fuzzy* sistemos – tai *fuzzy* sistemos, kurios naudoja mokymosi algoritmus, nustatant *fuzzy* taisykles ir *fuzzy* aibes bei taip apdorojant duomenis. Modernios *neuro-fuzzy* sistemos dažniausiai yra daugiasluoksniai neuroniniai tinklai. Reikia atkreipti dėmesį, kad *neuro-fuzzy* tinkluose ryšių svoriai ir aktyvizacijos funkcijos skiriasi nuo įprastinių neuroninių tinklų. Tyrimai įrodė, kad *fuzzy* logikos ir neuroninių tinklų apjungimas duoda geriausią rezultatą. Svarbu paminėti ir tai, kad šiuo metu plačiai tiriamas ir šio metodo taikymas atramos vektorių mašinų metodikoje.

*Neuro-fuzzy* modelis gali būti naudojamas įvairiems finansiniams uždaviniams spręsti: prognozuoti banko akcijų būsimas vertes, nuspėti investavimo galimybes, esant ekonominei krizei, nekilnojamojo turto vertinimui, vertybinių popierių įvertinimui. Tokios sistemos gali būti naudojamos ir kredito rizikos vertinime (S. Piramuthu, 1999; R. Mahotra, D.K. Mahotra, 2002; Khan, 2002).

#### **1.1.5.2. Fuzzy logikos taikymo rizikos vertinime galimybės ir prielaidos**

*Fuzzy* logiką galima naudoti kaip pagalbinę priemonę nustatant tiek kredito, tiek ir kito tipo, pavyzdžiui, operacinę, riziką. Rizikos valdyme už istorinius duomenis daug patikimesni yra ekspertiniai duomenys. Šiuo atveju gali būti nustatomi rizikingi kriterijai, naudojant kokybinius metodus, pavyzdžiui, sudarant rizikingiausių kriterijų, pagal kuriuos vertinamos kredituojamo asmens ar organizacijos kreditavimo galimybės, sąrašą. Paprastai gali būti sudėtinga surinkti visą informaciją apie atskirus rizikos faktorius ar sritis, todėl svarbiausių kriterijų nustatymas tampa sudėtingu uždaviniu. Gali būti ir taip, jog labiausiai įtakojantys kriterijai tiesiog bus nepastebėti; tokiu atveju kredi-

torius gali patirti daugybę nuostolių.

Kredito rizikos valdyme svarbu tiksliai apibrėžti rizikingus faktorius ir nustatyti jų ryšius su kitais kriterijais. Pritaikius neaiškiają (*fuzzy*) logiką, galima sudaryti kuo tikslesnį rizikingų faktorių sąrašą, o taip pat ir nustatyti, į kuriuos iš jų kreipti didžiausią dėmesį.

Toliau pateikiami 4 žingsniai, leidžiantys nustatyti faktorių svarbą ir įtraukti juos į rizikingiausiųjų sąrašą, naudojant *fuzzy* logikos metodą. Tai leidžia sukurti modelį, nusakantį ryšį tarp kredito rizikos proceso ir jos vertinimui naudojamų pagrindinių indikatorių.

1. *Nustatyti pagrindinius indikatorius.* Nustatomi pagrindiniai rizikos vertinimo indikatoriai. Tai gali būti tiek finansinis, tiek operacinis ar kitas kintamasis, su kuriuo galima apskaičiuoti galimos rizikos galimybę. Galimi kintamieji plačiau aprašyti 1.1 skyrelyje. Šiuo atveju nėra jokių apribojimų, nusakančių, kiek reikia imti indikatorių.

2. *Priskirti reikšmes pagrindiniams rizikos indikatoriams.* *Fuzzy* logika leidžia naudoti lingvistinius kintamuosius, nustatant kredito rizikos tikimybę. Tokie lingvistiniai kintamieji, kaip „aukštas“, „žemas“, „vidutinis“, „mažas“, „didelis“, gali būti priskirti kiekvienam indikatoriumi, kas leidžia sukurti logines vertinimo sąlygas.

3. *Nustatyti kiekvieno rizikos indikatoriaus poveikį rizikos galimybei.* Kai yra nustatyti pagrindiniai rizikos indikatoriai, toliau reikia nustatyti, kaip kinta nuostolio galimybė, priklausanti nuo atitinkamo indikatoriaus. Šiam tikslui naudojamos *fuzzy* logikos taisyklės, paremtos „*jei ... tada*“ (iš angl. *if ... then*) sakiniu, kurios nustato išpareigojimų nevykdymo galimybės dydį įvairiuose indikatorių lygmenyse ir apima visus galimus pagrindinių rizikos vertinimo indikatorių variantus. 4 lentelėje pateikiamas tokių taisyklių sudarymo pavyzdys kredito rizikos vertinimo atveju. Šiuo atveju nagrinėjama tokių indikatorių, kaip EBIT ir turto santykio bei nepaskirstyto pelno įtaka, remiantis klasikiniu Altman modeliu (1.2 skyrelis).

4 lentelė. *Fuzzy* logikos taisyklių taikymas nagrinėjant kredito rizikos galimybę

Taisyklė 1	Jei	EBIT/turtas	yra	Didelis
		Nepaskirstytas pelnas	yra	Mažas
	Tada	Rizikos tikimybė	yra	Mažas
Taisyklė 2	Jei	Nepaskirstytas pelnas	yra	Mažas
		EBIT/turtas	yra	Mažas
	Tada	Rizikos tikimybė	yra	Didelis
Taisyklė 3	Jei	Nepaskirstytas pelnas	yra	Didelis
		EBIT/turtas	yra	Mažas
	Tada	Rizikos tikimybė	yra	Vidutinis
Taisyklė 4	Jei	Nepaskirstytas pelnas	yra	Didelis
		EBIT/turtas	yra	Didelis
	Tada	Rizikos tikimybė	yra	Labai mažas

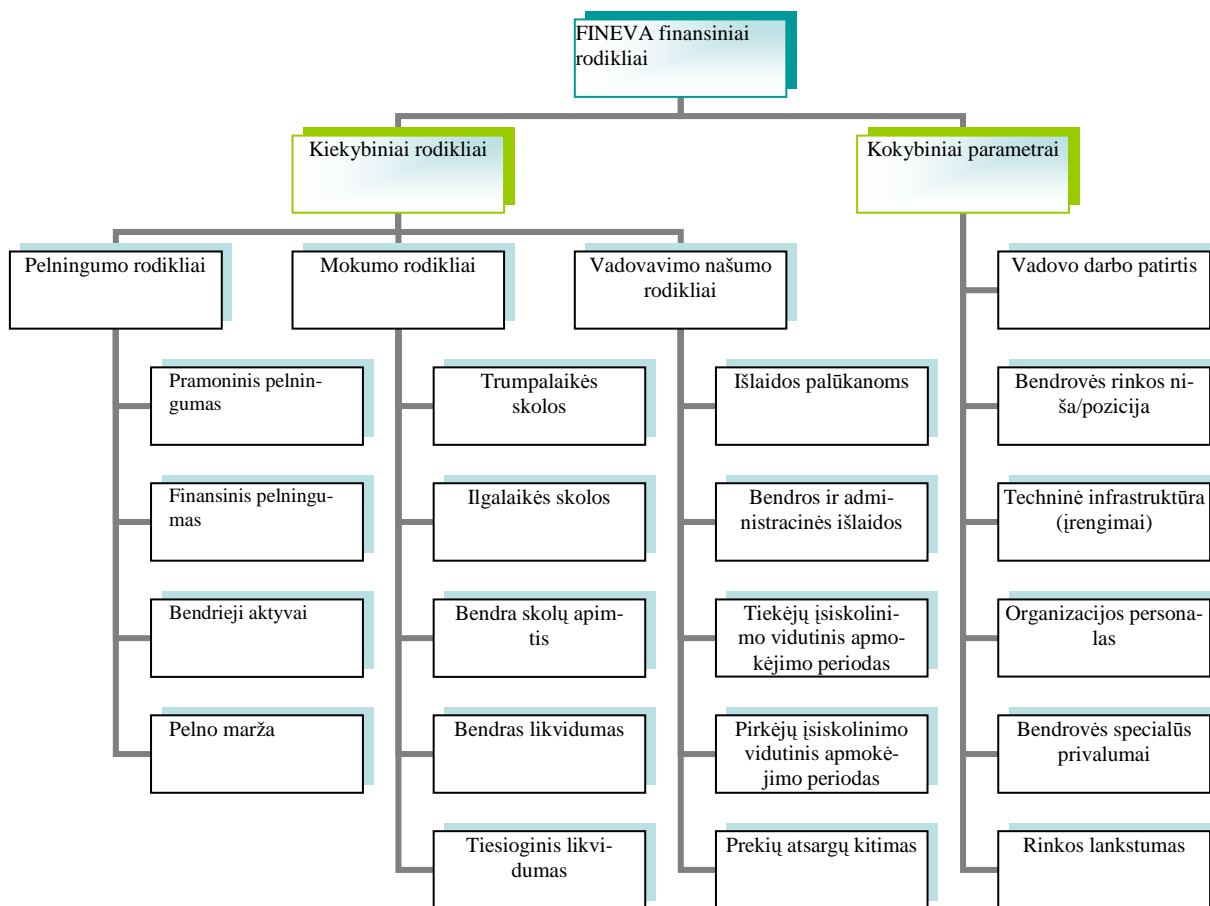
Šaltinis: sudaryta autoriaus.

4. *Laukiamo nuostolio tikimybės apskaičiavimas.* Kadangi fuzzy logikos taisyklės apima visus pagrindinių rizikos indikatorių galimus variantus, galima suskaičiuoti kiekvieną rizikos indikatorius reikšmę atitinkantį kredito rizikos tikimybės dydį.

*Fuzzy* logika suteikia daugiau informacijos ir padeda efektyviau valdyti riziką, nes nustatant rizikos indikatorius galima gauti tikslesnę informaciją. Apibendrintai vertinant įsipareigojimų nevykdymo galimybę yra sunkiau rasti priklausomybę tarp rizikos indikatorių; naudojant *fuzzy* logiką, galima patikrinti visas galimas priklausomybes tarp indikatorių ir pagal jas nustatytiems faktoriams skirti daugiau dėmesio.

### 1.1.5.3. Ekspertinių taisyklių taikymo finansinės rizikos vertinime pavyzdys

Ši žinių išgavimo bei pateikimo metodologija finansinės analizės rityje realizuota FINEVA (FINAncial EVAluation) sistemoje (Matsatsinis, Doumpos & Zopounidis 1997; cituojama pagal Nedović, Devedžić, 2001). Tai daugelio kriterijų žiniomis paremta sprendimų paramos sistema bendrovių našumo, silpnųjų bei stipriųjų pusių ir galimybių įvertinimui, sukurta, naudojant M4 programavimo terpę Kretos universitete.



Šaltinis: sudaryta autoriaus pagal L. Nedović, V. Devedžić. Expert systems in finance – a cross-section of the field.

**8 pav. FINEVA naudojami finansiniai įverčiai ir rodikliai**

Joje yra galimybė naudoti tiek kokybinį vertinimą, tiek ir finansinių rodiklių interpretavimą

bei jų kombinavimą; šiuo atveju ekspertai pasiūlė priskirti lygius svorius. Ją kurdami, ekspertai naudojo žinias iš tarptautinės literatūros bei informaciją, gautą iš daugybės interviu su finansiniais ekspertais. Produkcinės taisyklės leido gautas euristines žinias realizuoti žinių bazėje. Autorių teigimų, žinių bazėje yra 1693 taisyklės, aprašančios daugiau kaip 13.000 galimų kombinacijų.

Finansiniai rodikliai buvo suskirstyti į tris pagrindines grupes – pelningumo rodikliai, mokesčių rodikliai bei vadovavimo našumo rodikliai. 8 pav. pateikiama šių rodiklių klasifikacija, o 5 lentelėje – kai kurios euristinės taisyklės, realizuotos FINEVA. Šiuo atveju finansiniai rodikliai išreikšti procentiniais pokyčiais, lyginant su prieš tai buvusiais metais. 6 lentelėje pateikiami kokybinių rodiklių įverčiai.

5 lentelė. Kiekybinių finansinių rodiklių modeliavimo taisyklės FINEVA

<b>Pramoninis pelningumas A1:</b> A1 < 10% nepatenkinamas 10% < A1 <= 20% vidutinis 20% < A1 <= 30% patenkinamas A1 > 30% labai patenkinamas	<b>Finansinis pelningumas A2:</b> A2 <= 17.5% nepatenkinamas 17.5% < A2 <= 20% vidutinis 20% < A2 <= 23% patenkinamas 23% < A2 labai patenkinamas	<b>Bendris pelnas/Bendrieji aktyvai A3:</b> A3 <= 0% nepatenkinamas 0% < A3 <= 50% vidutinis 50% < A3 <= 75% patenkinamas A3 > 75% labai patenkinamas
<b>Pelno marža A4:</b> A4 <= 0% nepatenkinamas 0% < A4 <= 50% vidutinis 50% < A4 <= 100% patenkinamas A4 > 100% labai patenkinamas	<b>Trumpalaikės skolos B1:</b> B1 < 25% nepatenkinamas 25% < B1 <= 50% vidutinis 50% < B1 <= 75% patenkinamas 75% < B1 <= 100% labai patenkinamas	<b>Bendros skolos B2:</b> B2 > 80 % nepatenkinamas 60% < B2 <= 80% vidutinis 40% < B2 <= 60% patenkinamas B2 <= 40% labai patenkinamas
<b>Ilgalaikės skolos B3:</b> B3 <= 0.5 patenkinamas B3 > 0.5 nepatenkinamas	<b>Bendras likvidumas B4:</b> B4 >= 2 patenkinamas B4 < 2 nepatenkinamas	<b>Tiesioginis likvidumas B5:</b> B5 <= 1 nepatenkinamas 1 < B5 < 1.5 patenkinamas B5 >= 1.5 labai patenkinamas
<b>Finansinės išlaidos C1:</b> C1 > 5% nepatenkinamas 3% < C1 <= 5% vidutinis 2% < C1 <= 3% patenkinamas C1 <= 2% labai patenkinamas	<b>Bendrosios ir administracinės išlaidos C2:</b> C2 > 8% nepatenkinamas 6% < C2 <= 8% vidutinis 4% < C2 <= 6% patenkinamas 2% < C2 <= 4% labai patenkinama C2 <= 2% puikus	<b>Tiekėjų išskolinimo vidutinis apmokėjimo periodas C3, pirkėjų išskolinimo vidutinis apmokėjimo periodas C4:</b> C3 > C4 nepatenkinamas C3 <= C4 patenkinamas
<b>Prekių atsargos C5:</b> C5 didėja nepatenkinama C5 mažėja ar stabilus patenkinamas		

Šaltinis: Ljubica Nedović, Vladan Devedžić. Expert systems in finance – a cross-section of the field.

6 lentelė. Kokybinių finansinių rodiklių modeliavimo taisyklės FINEVA

<b>Vadovo darbo patirtis:</b>	
Neigiama patirtis	Nepatenkinamas
Nėra jokios patirties	Vidutinis
Teigiama patirtis iki 5 metų	Patenkinamas
Teigiama patirtis 5-10 metų	Labai patenkinamas

<b>Teigiama gaugiau nei 10 metų patirtis</b>	Puikus
<b>Bendrovės rinkos niša/pozicija:</b>	
<b>Didelė konkurencija, silpna bendrovės pozicija</b>	Nepatenkinamas
<b>Didelė konkurencija, stabili ir konkurencinga bendrovė</b>	Vidutinis
<b>Vidutinė konkurencija, stipri bendrovės pozicija</b>	Patenkinamas
<b>Silpna konkurencija, bendrovė tarp pirmaujančių</b>	Labai patenkinamas
<b>Bendrovė monopolistė</b>	Puikus
<b>Techninė infrastruktūra (įrenginiai):</b>	
<b>Sena ir netinkama įranga, pasenę gamybos metodai</b>	Nepatenkinamas
<b>Vidutiška techninė infrastruktūra, nekonkurencingi gamybos kaštai</b>	Vidutinis
<b>Pakankamai modernizuota įranga</b>	Patenkinamas
<b>Gera techninė infrastruktūra, planuojama pilna modernizacija</b>	Labai patenkinamas
<b>Puiki infrastruktūra, modernūs gamybos metodai</b>	Puikus
<b>Organizacijos personalas:</b>	
<b>Bendrovės personalo samdymo politikos trūkumas</b>	Nepatenkinamas
<b>Pakankamas personalo politikos samdymo lygis</b>	Vidutinis
<b>Pakankamas personalo politikos samdymo lygis, noras jį tobulinti</b>	Patenkinamas
<b>Geras personalo politikos samdymo lygis</b>	Labai patenkinamas
<b>Puikus personalo politikos samdymo lygis</b>	Puikus
<b>Bendrovės specialūs konkurenciniai privalumai</b>	
<b>Bendrovė neturi kompetencijos gamybos metodams</b>	Nepatenkinamas
<b>Bendrovė turi mažai kompetencijos gamybos metodams</b>	Vidutinis
<b>Bendrovė turi pakankamai kompetencijos gamybos metodams</b>	Patenkinamas
<b>Bendrovė turi išskirtinę kompetenciją gamybos metodams</b>	Labai patenkinamas
<b>Rinkos lankstumas:</b>	
<b>Bendrovė neseka rinkos tendencijų, gamina produkciją, turinčią mažą paklausą</b>	Nepatenkinamas
<b>Bendrovės lankstumas ribotas</b>	Vidutinis
<b>Bendrovės lankstumas pakankamas</b>	Patenkinamas
<b>Bendrovė seka rinkos tendencijas</b>	Labai patenkinamas
<b>Bendrovė yra lyderė savo gaminamos produkcijos šakoje</b>	Puikus

Šaltinis: Ljubica Nedović, Vladan Devedžić. Expert systems in finance – a cross-section of the field.

Taigi naudojantis šiomis ir kitomis taisyklėmis, galima sukurti ir ekspertinį posistemį, kuris galėtų atlikti įvertinimą, naudodamas tiek realius finansinius rodiklius, tiek ir ekspertų žinias, nusakančias riziką iš kokybinės pusės. Šių taisyklių realizacija CLIPS kalba, kurią galima naudoti ir JAVA Jess ekspertinių sistemų kūrimo aplinkoje, kas leistų jų integraciją į bendrą kredito rizikos SPS, pateikiama 5 priede.

#### 1.1.6. Evoliuciniai algoritmai

Genetiniai algoritmai - tai metodas, kuriuo remiantis, problemos sprendžiamos evoliucionuojančio proceso pagrindu, kai gaunamas geriausias sprendimas, kitaip tariant, sprendimas yra išvystomas. Ši metodika buvo pirmąsyk aprašyta 1960 m. I. Rechenberg darbe „Evoliucijos strategijos“ (orig. „*Evolutionsstrategie*“). Toliau jo idėja buvo vystoma kitų mokslininkų, ypač John Hol-



land ir jo kolegų. Būtent šis mokslininkas ir sukūrė pirmuosius genetinius algoritmus, kurie buvo aprašyti jo knygoje "Adaption in Natural and Artificial Systems" 1975 m. 1992 m. genetinis algoritmas pirmą sykį buvo panaudotas praktiškai (John Koza) (Obitko, 1998).

Šios DI krypties pagrindas apima tiek biologinį (genetinį), tiek ir matematinį požiūrius. Pirmasis yra paremtas tuo, kad organizmą sudarančios ląstelės savyje turi chromosomų *rinkinį (genomą)*, savo ruožtu sudarytą iš genų rinkinio (*genotipo*), kurių kiekvienas chromosomoje turi savo vietą (*padėtį*, angl. *locus*). Kaip žinoma iš biologijos mokslo, kiekvienas genas savyje saugo informaciją apie tam tikrą bruožą, pavyzdžiui, žmogaus akių spalvą Dauginimosi metu pirma įvyksta rekombinavimas (arba kryžminimasis), ko pasekoje tėvinių objektų genai suformuoja naują chromosomą. Tada naujai sukurta chromosoma gali „mutuoti“, ką sąlygoja DNR elementų pasikeitimas dėl netikslios tėvų genų kopijos (kitais tariant, vyksta evoliucionavimas arba evoliucija), todėl uždavinio sprendimas naudojant genetinius algoritmus yra evoliucinis procesas.

Sprendžiant tam tikrą problemą, paprastai ieškoma geriausio sprendimo iš galimų. Visa galimų sprendimų aibė vadinama paieškos erdve (*search space*). Kiekvieną galimą sprendimą atspindi vienas paieškos erdvės taškas, kuris gali būti pažymėtas atitinkama verte (arba tinkamumu). Genetinio algoritmo tikslas – surasti paieškos erdvėje optimalų sprendimą, pažymėtą tam tikru tašku. Tokiu atveju sprendimo ieškojimas tampa kurio nors ekstremumo (minimumo ar maksimumo) paieška tam tikroje galimų sprendimų aibėje uždaviniu. Pagrindinė problema, išskylanti šiuo atveju, yra ta, kad ne visada paieškos erdvė yra visiškai apibrėžta, pavyzdžiui, kokia nors funkcija – dažnai žinoma tik keletas jos taškų. Genetinio algoritmo vykdymo metu sprendimo ieškantis procesas pats evoliucionuodamas generuoja kitus galimus sprendimus (paieškos erdvės taškus) (Obitko, 1998).

Pagrindinė algoritmo problema – neaiškumas, kur ieškoti sprendimo, todėl dažniausiai ieškoma ne teisingo sprendimo, bet geriausio sprendimo. Rastas sprendimas laikomas geriausiu, kadangi ne visada galima nuspręsti, kas yra optimalus sprendimas. Jo paieškai yra daug evoliucinių metodų, čpatys svarbiausi bei populiariausi pateikiami 5 priede. Iš jų dažniausiai nagrinėjamas ir taikomas **genetinis algoritmas** bei įvairios jo modifikacijos. Pradiniai algoritmo duomenys – sprendimų aibė, atvaizduojama chromosomomis, vadinama *populiacija*. Vienos populiacijos sprendimai naudojami formuoti naujai populiacijai, tikintis, kad nauja populiacija bus geresnė. Naujos populiacijos sprendimai pasirenkami atsižvelgiant į jų tinkamumą – kuo jie tinkamesni, tuo daugiau galimybių jiems daugintis. Laikoma, kad kuo geresnis tėvinių chromosomų tinkamumas, tuo didesnė tikimybė sugeneruoti „geresnę“ „atžalą“.

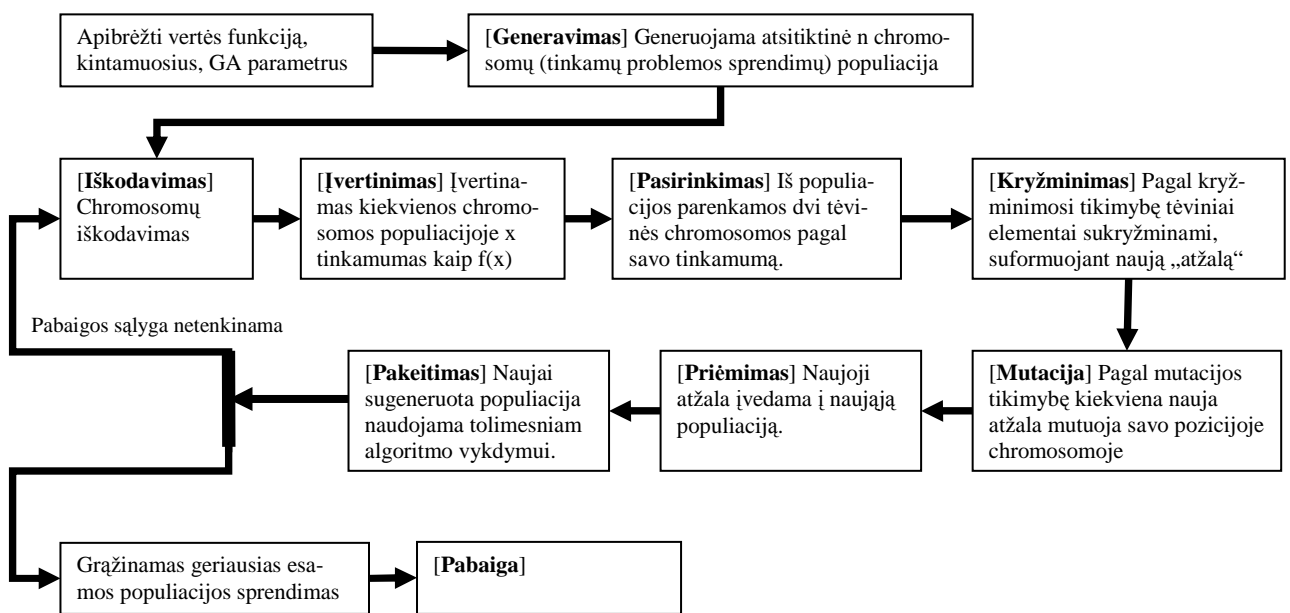
Formaliai šios sąvokos ir pats genetinis algoritmas gali būti aprašomas taip (Dunham, 2003):

- Jei turime alfabetą  $A$ , tai chromosoma arba individas yra eilutė  $I = I_1, I_2, \dots, I_n$ , kur  $I_j \in A$ . Kiekvienas eilutės simbolis  $I_j$  vadinamas genu. Reikšmės, kurias kiekvienas simbolis gali

turėti, vadinamos alelėmis. Populiacija P yra individų aibė.

- Jei duota populiacija P, tai tinkamumo funkcija (*fitness function*) yra susiejimas  $f : P \rightarrow R$ .

- Genetinis algoritmas (GA) – tai skaičiavimo modelis, susidedantis iš 5 esminių dalių:
  1. Individų aibės P inicializavimo;
  2. Kryžminimosi metodo;
  3. Mutacijos algoritmo;
  4. Tinkamumo funkcijos;
  5. Algoritmo, kuris kryžminimosi ir mutacijos metodus aibei P iteratyviai, naudodamas tinkamumo funkciją nustatyti geriausiems individams, kurie turi likti. Algoritmas pakeičia iš anksto apibrėžtą individų skaičių iš populiacijos kiekvienos iteracijos metu ir baigia darbą, kai pasiekama tam tikra riba.



Šaltinis: sudaryta autoriaus.

### 9 pav. Genetinio algoritmo principinė schema

Yra du esminiai genetinio algoritmo parametrai:

1. *Kryžminimosi tikimybė* – nusako, kaip dažnai atliekamas kryžminimasis. Jei šis parametras lygus 100%, tai kryžminimo atveju visos „atžalos“ susidaro kryžminimosi būdu, jei 0% - visa nauja populiacija sudaroma iš tikslų senos populiacijos chromosomų kopijų (tai nereiškia, kad naujoji populiacija tokia pati, kaip senoji);

2. *Mutacijos tikimybė* – nusako, kaip dažnai chromosomos dalys mutuoja. Jei nėra jokios mutacijos (tikimybė lygi 0%), tai chromosomoje niekas nekeičiama, jei mutacijos tikimybė lygi 100%, tai pakeičiama visa chromosoma. Pati mutacija neturėtų vykti dažnai, kadangi ji gali pavirsti į atsitiktinę paiešką.

Dar vienas svarbus parametras yra *populiacijos dydis*, nuo kurio priklauso paieškos erdvės dydis bei galimybės atlikti kryžminimąsi. Reikia atsižvelgti į tai, kad per didelis populiacijos dydis

gali smarkiai sulėtinti genetinio algoritmo veikimą, be to, tyrimai parodė, kad didelės populiacijos naudojimas, palyginti su vidutine, nėra efektyvus sprendimas (Obitko, 1998).

Evoliuciniuose skaičiavimuose naudojama gausybė kryžminimosi ir mutacijų strategijų ir metodų. Paprastai kryžminami du atsitiktinai parinkti individai (chromosomos) arba vienas atsitiktinai parinktas individas kryžminamas su vis kitu atsitiktinai parinktu individu; pastarasis atvejis vadinamas mutacija. Populiariausi metodai pateikiami 15 lentelėje.

Genetiniai algoritmai šiuo metu yra viena iš perspektyviausių ir veiksmingiausių dirbtinio intelekto sričių, plačiai taikoma įvairiems uždaviniams spręsti. Ji gali būti naudojama (ir paprastai yra kartu naudojama) ir su kitomis DI sritimis, tokiomis, kaip neuroniniai tinklai ar fuzzy logika.

Gali būti išskirti šie GA privalumai (Schlotmann, 2001):

- priešingai nuo kitų metodų, sprendimo proceso metu įmanoma saugoti nebaigtus ir ne-optimalius sprendimus, kurie gali būti artimi globaliam optimumui, kol yra daug lokalių optimumų, žymiai nutolusių nuo globalaus optimumo;

- galima spręsti netiesines, nelanksčias problemas;
- konceptualus paprastumas;
- plačios taikymo galimybės;
- geresnis našumas nei klasikinių metodų;
- galimybė naudoti problemos žinias bei hibridizaciją;
- lygiagretumas;
- lankstumas dinaminiam pasikeitimams;
- savybė pačiam optimizuotis;
- galimybė spręsti problemas be žinomų sprendimų.

Ši dirbtinio intelekto sritis gali būti pritaikyta ir kredito rizikos vertinimui. Ją tyrę autoriai pasiūlė kredito portfelių vertinimo hibridinius metodus (Schlottmann, Seese, 2001, 2002), taip pat metodą kreditų rizikos reitingavimo patobulinimui (Apotheker, Barthelemy, 2000). Pirmasis algoritmas apjungia genetinio algoritmo metodologiją su godžiu algoritmu, naudojančiu kiekybinę portfelio kredito rizikos modelio informaciją, panašiu į rizikos optimizavimą, naudojant gradientu paremtus metodus (šis metodas taip pat aprašomas Schlottmann ir Seese darbe). Kitame šių autorių darbe aprašomas hibridinis evoliucinis algoritmo karkasas ribotų efektyvių rizikos atžvilgiu kredito portfelių struktūrų vertinimui. Šio algoritmo paskirtis yra efektyvių Pareto optimalumo atžvilgiu portfelio struktūrų radimas, naudojant netiesinius rizikos vertinimo parametrus bei diskrečius sprendimo kintamuosius (Schlottmann, Seese, 2001, 2002). Tuo tarpu Apotheker ir Barthelemy tyrė globalesnę šiame kontekste problemą – jų tyrimo tikslas buvo rizikos reitingų vertinimo galimybių

patobulinimas, tiriant šalių kreditavimo rizikos pasikeitimus. Šiam tikslui jie pasinaudojo genetiniais algoritmais bei *simulated annealing* metodu, skirtu globalaus optimizavimo problemos gvilde-  
nimui. Genetiniais algoritmais naudojantis, galima prognozuoti ir finansinį stabilumą/nestabilumą  
(Zumbach, Pictet, Masuti, 2001).

## **1.2. Kredito rizikos vertinimo problema bei pagrindinės sąvokos**

### **1.2.1. Rizikos sąvoka, klasifikacija ir galimi vertinimo būdai**

Rizikos vertinimas yra vienas svarbiausių klausimų tiek finansų analitikui, tiek ir investuo-  
tojui, siekiančiam investuoti pinigus į kokią nors įmonę ar jos vertybinius popierius. Rizika turi du  
išmatavimus - kiekį (galimo nuostolio dydį) ir kokybę (potencialaus nuostolio patyrimo tikimybę).  
Rizikos kiekį galima riboti nustatant limitus prisiimamoms rizikoms, o kokybę galima vertinti kre-  
dito reitingais.

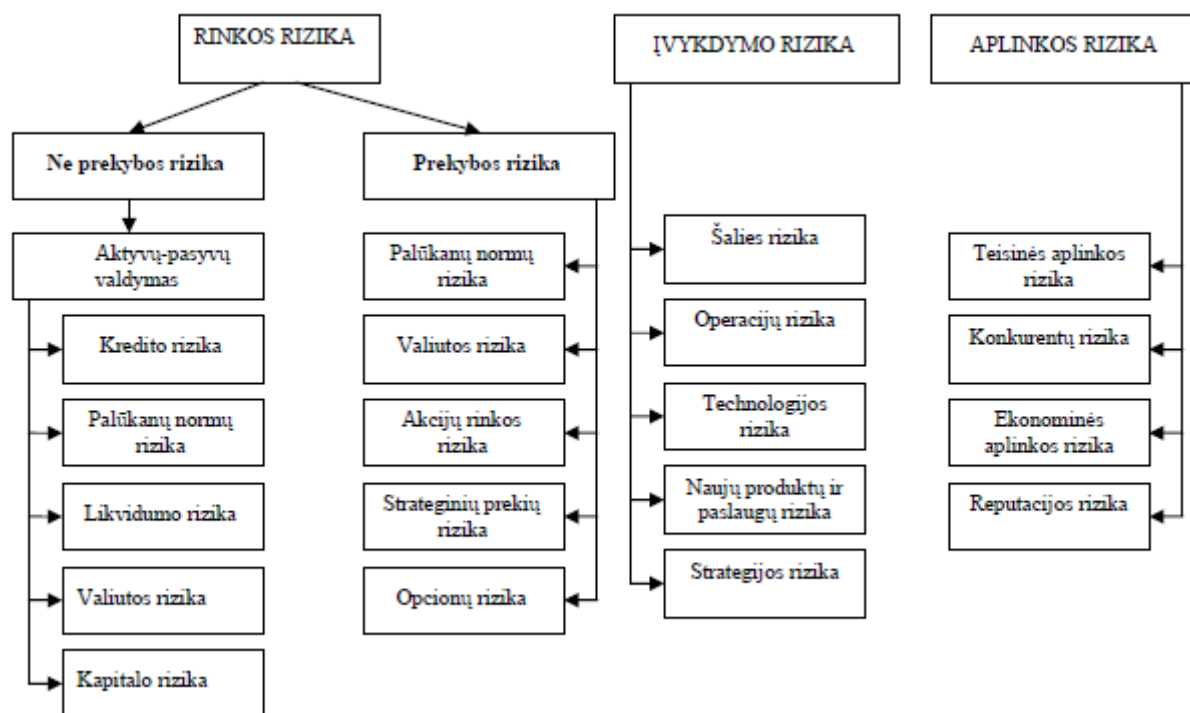
Riziką geriausia nagrinėti, kaip pavyzdį imant bankus ar kitas finansines institucijas. Tam  
yra bent 4 prielaidos:

1. Jos veikiama rinkos rizikų, tiesiogiai kylančių iš finansinių institucijų veiklos;
2. Šių organizacijų stabilumas yra visuomenės svarbos reikalas, dėl ko joms privalu laiky-  
tis oficialių reikalavimų (normatyvų), skirtų riboti rizikai;
3. Jos suteikia kreditus, taip sukurdamos planuojamus ir tuo pačiu rizikingus pinigus srau-  
tus. Tokiu būdu automatiškai atsiranda kredito rizika;
4. Bankas turi mažiau akcinio kapitalo, lyginant su įsipareigojimais, nei nefinansinės įmo-  
nės, kas reiškia, kad bet koks pelnas ar nuostolis turės didelę įtaką bankų akcininkų nuosavybės ver-  
tei. Norint didinti akcininkų nuosavybę (kapitalą), reikia uždirbti pelną, taigi ir prisiimti riziką; kuo  
didesnis pelnas, tuo didesnė ir rizika, kurią reikia prisiimti. Todėl bankai stengiasi kuo tiksliau įver-  
tinti prisiimamą riziką, pirmieji diegia rizikų valdymo modelius. Iš esmės, rizikos valdymas yra  
pelno maksimizavimo ir rizikos minimizavimo uždavinys akcininkų nustatytuose rizikos priimti-  
numo rėmuose.

Norint tinkamai įvertinti riziką, kurią kreditorius gali patirti išduodamas kreditą, reikia api-  
brėžti visas su kreditoriaus bei debitoriaus teisine, makroekonominė ir kita aplinka susijusias bei  
patiriamas rizikas (10 pav.). Svarbu pabrėžti tai, kad vienokia rizika (kredito, likvidumo) išmatuo-  
jama, kitokia (ekonominės aplinkos) įvertinama tik prognozuojant, o dar kita (sisteminė rizika, api-  
manti įmonės valdymą, įvairius jos veikimo aspektus) neišmatuojama; ją tinkamai įvertinti gali tik  
analitikai, įvertinę kitus aplinkinius veiksnius, įtakojančius debitoriaus makroekonominę aplinką.  
Rizikos valdymas dažniausiai apima išmatuojamą riziką; ji ir yra šio darbo tyrimo objektas, todėl  
šiuo darbe čia ir toliau bus nagrinėjama tik kredito rizika.

Reikia pabrėžti, kad tiek diagramoje, tiek ir 7 lentelėje nagrinėjamos rizikos, labiau susijusios su bankų veikla, tačiau jos (ar bent tam tikra jų dalis) lengvai gali būti pritaikytos ir kitoms organizacijoms su panašiu ar skirtingu veiklos profiliu.

Detali rizikos klasifikacija ir analizė pateikiama Vaškelačio knygoje (Vaškelaitis, 2003); čia apsiribojama tik rizikos klasifikacijos pateikimu ir trumpa lyginamąja analize (7 lentelė).



Šaltinis: Vaškelaitis, Vytautas. Pinigai: komerciniai bankai ir jų rizikos valdymas.

### 10 pav. Banko rizikos klasifikacija

Kaip matyti iš diagramos, gali būti išskiriamos trys pagrindinės rizikos klasės: aplinkos rizika, įvykdymo rizika bei rinkos rizika. Aplinkos rizika apibrėžiama kaip galima nuostolių rizika dėl teisinės ir ekonominės aplinkos pakitimų, nuo kurių bankas siekia apsisaugoti, tačiau kuriuos kontroliuoti turi mažiausiai galimybių. Įvykdymo rizika – tai nuostolių rizika, susijusi su banko paslaugų teikimo bei šių paslaugų pelningumo užtikrinimu; rinkos rizika apibūdinama kaip galimų nuostolių rizika dėl rinkos kainų ir normų svyravimo finansų ir kapitalo rinkose, jai priskiriamos tos rizikos kategorijos, kurios susijusios su šiais svyravimais (Vaškelaitis, 2003)

7 lentelė. Rizikų analizė ir galimi vertinimo metodai

RIZIKOS PA-VADINIMAS	ATSIRADIMO PRIEŽASTIS	VALDYMO TIKSLAS	VERTINIMO METODAI
Teisinės aplinkos rizika	Valdžios institucijų teisinių aktų, reglamentuojančių ekonominius santykius, pakitimai	Tinkamai pasiruošti naujų teisės aktų įgyvendinimui, kad būtų išvengta nuostolių, neiškiltų pavojus stabiliai veiklai ar ne-	Ekspertinis, analitinis

<b>RIZIKOS PA- VADINIMAS</b>	<b>ATSIRADIMO PRIEŽASTIS</b>	<b>VALDYMO TIKSLAS</b>	<b>VERTINIMO ME- TODAI</b>
		sumažėtų pasitikėjimas banku	
<b>Ekonominės aplinkos rizika</b>	Vyriausybės vykdomos ekonominės politikos ir centrinio banko vykdomos pinigų politikos pokyčiai	Nuostolių dėl Vyriausybės ir CB politikos pokyčių mažinimas	Ekspertinis, analitinis
<b>Konkurentų rizika</b>	Konkurencinių pranašumų neišnaudojimas, neadekvati reakcija į konkurentų veiksmus rinkoje ir aplinkos pokyčius	Užtikrinti banko konkurencingumą rinkoje visose teikiamų paslaugų ir produktų srityse ir didinti banko užimamos rinkos dalį pagal verslo plane numatytą strategiją	Ekspertinis, analitinis, rinkos tyrimai
<b>Reputacijos rizika</b>	Banko gero vardo praradimas dėl dalyvavimo pinigų plovimo operacijose, nelegalaus verslo finansavimo bei banko klientų komercinių paslapčių atskleidimo	Skaidrumo politika	Ekspertinis, analitinis, remiantis istoriniais faktais
<b>Įvykdymo rizika</b>			
<b>Šalies rizika</b>	Užsienio valstybių nenoras ar negalėjimas vykdyti savo įsipareigojimų kreditoriams. Gali būti tiek ekonominė, tiek politinė	Vertinama pagal elgseną praeitame laikotarpyje	Specialios rodiklių sistemos; šalies ekonominės, socialinės ir politinės padėties bei reitingų sekimas ir analizė
<b>Operacijų rizika</b>	Darbuotojų daromos klaidos, neatidumas arba darbuotojų nesugebėjimas laiku ir tinkamai atlikti banko operacijų	Užtikrinti, kad stabilus ir nepertraukiamo darbo nepaveiktų darbuotojų klaidos ir skaičiavimai bei tinkamą vidinę sistemos kontrolę	Pagal veiklos klaidas ar iškilusias problemas bei jų skaičių, veiklos efektyvumą
<b>Technologijos rizika</b>	Banko technologijų sistemų ir naudojamų technologinės įrangos netinkamas darbas ar neefektyvumas bei negebėjimas aptarnauti išaugusios banko teikiamų paslaugų apimtį.	Užtikrinti normalų darbą technologijų sistemų gedimo, gaisrų ar kitokių stichinių nelaimių atvejais bei siekti, kad bankas naudotųsi naujausia technine įranga ir užtikrintų efektyvų ir kokybišką klientų aptarnavimą. Neįprastų situacijų planuose reglamentuojama, kaip užtikrinti nepertraukiamą darbą krizės situacijose	Pagal veiklos klaidas ar iškilusias problemas bei jų skaičių, veiklos efektyvumą
<b>Naujų produktų ir paslaugų rizika</b>	Atsiranda įdiegiant ar teikiant į rinką naujus produktus ir paslaugas dėl netiksliai įvertintos paklausos rinkoje ir išlaidų lygio, dėl to šie produktai bus nepelningi	Užtikrinti naujų produktų ir paslaugų įdiegimo pelningumą bei taip pasiruošti šiam įdiegimui, kad jis nesukeltų pavojaus banko stabilumui	Atliekant rinkos tyrimus, įvertinant galimą paklausą, būsimas pajamas ir išlaidas bei produktų/ paslaugų pelningumą
<b>Strategijos rizika</b>	Negebėjimas kompleksiskai įvertinti ekonominę aplinką ateityje, pasirenkant veiklos plėtros sritis	Užtikrinti nuolatinį ir pelningą augimą, plėtrą bei vystymąsi	Verslo plano analizė
<b>Rinkos rizika</b>			
<b>Ne prekybos rizika</b>	Turto ir įsipareigojimų struktūros svyravimai, dėl kurių gali sumažėti banko pelnas ir/ar	Nustatyti tinkamą banko turto ir įsipareigojimų bei nebalansinių straipsnių struktūrą, kad	Fundamentalioji ir techninė analizė

<b>RIZIKOS PA-VADINIMAS</b>	<b>ATSIRADIMO PRIEŽASTIS</b>	<b>VALDYMO TIKSLAS</b>	<b>VERTINIMO METODAI</b>
	turto bei įsipareigojimų dabartinė vertė. Šios rizikos klasės veiksniai turi didžiausią įtaką banko veiklai. Su šių rizikos kategorijų valdymu yra susijęs vidinių rizikų ribojančių limitų nustatymas.	svyravimai finansų rinkose nesumažintų pelno ir turto bei dabartinės vertės įsipareigojimų	
<b>Kredito rizika</b>	Banko skolininko dėl tam tikrų priežasčių (ekonominiai, politiniai pakitimai, bankrotas ir kt.) negalėjimas įvykdyti savo įsipareigojimų bankui	Išvengti nuostolių skolinant pinigus skolininkams	Fundamentaliaji ir techninė skolininko finansinių rodiklių analizė, įvairūs matematiniai, ekonometriniai, finansiniai, statistiniai modeliai
<b>Palūkanų normų rizika</b>	Išskyla tuomet, kai, keičiantis palūkanų normoms, palūkanų pajamos gali sumažėti arba išlaidos palūkanoms tampa didesnės negu pajamos	Šis rizikos vertinimas ypač svarbus bankams, kadangi didžioji dalis jų gaunamų pajamų bei padaromų išlaidų yra būtent palūkanų pajamos ir išlaidos.	Techninė analizė, prognostiniai metodai
<b>Likvidumo rizika</b>	Tai rizika, kad bankas tam tikru momentu neturės pakankamai likvidžių lėšų, kad įvykdytų visus savo įsipareigojimus mažiausiomis išlaidomis.	Pakankamo likvidžių lėšų kiekio užtikrinimas bei gali-mybių finansuoti lėšų trūkumą skolinantis iš rinkos arba banko gebėjimo vykdyti mokėjimus garantijos užtikrinimas	Techninė rodiklių analizė
<b>Valiutos rizika</b>	Tai rizika, kad bankas patirs nuostolių dėl nepalankiai pakitusio užsienio valiutos kurso. Valiutos kurso rizika labai didelė tuose bankuose, kurie siekia gauti spekuliacinių pajamų iš kurso nesutapimo skirtingose rinkose ir skirtingais laiko momentais, t.y. bankuose, vykdančiuose arbitražines operacijas.	Valiutos kurso rizika valdoma pasirenkant savo nacionalinę valiutą kaip mokėjimo valiutą; įtraukiant į paskolos sutartį sąlygas, keičiančias piniginių įsipareigojimų sumas priklausomai nuo valiutos kurso pokyčio ar apsidraudžiant būsimais valiutos sandoriais	Techninė analizė, prognostiniai metodai
<b>Kapitalo rizika</b>	Nepakankamo kapitalo dydis, kai bankas gali prarasti pasitikėjimą ar veiklos stabilumą. Kadangi prisiimamos rizikos dydis lemia patiriamų nuostolių dydį, kuris atsispindi ir kurį padengia kapitalas, rizika susijusi su kapitalu tiesiogiai.	Tinkamo kapitalo kiekio užtikrinimas bei prisiėmimas tik tų įsipareigojimų, kuriuos bankas pajėgus įvykdyti	Kapitalo rodiklių (sudėties, kiekio) analizė
<b>Prekybos rizika</b>	Dėl rinkos kainų ir kursų pakitimo sumažėjusi prekybinio portfelio vertė	Tinkamas prekybos užtikrinimas	Rizikuojamos vertės statistiniai metodai (rizikos vertės nustatymas (VaR), Markowitz ir kt.), modernūs ekonometriniai modeliai, testavimas nepalankiausiomis sąlygomis, Monte Karlo analizė

Šaltinis: sudaryta autoriaus, remiantis Vaškelaitis, Vytautas. Pinigai: komerciniai bankai ir jų rizikos valdymas.

Svarbu pabrėžti, jog nors formaliai visos rizikos rūšys gali būti atskirtos viena nuo kitos, bet realiai to padaryti beveik neįmanoma, kadangi kiekvienoje banko operacijoje dažniausiai pasireškia kelios skirtingos rizikos kategorijos ir veikia viena kitą. Todėl, vertinant tiek banko, tiek ir bet kurios kitos finansinės, prekybinės ar kito tipo institucijos finansinį pajėgumą, svarbu atsižvelgti ne tik į finansinius rodiklius, bet ir kitus aspektus, susijusius su kredituojama organizacija, t.y., vertinant kredito riziką, svarbu įvertinti ir kitas rizikas.

### 1.2.2. Kredito rizikos sąvoka ir pagrindiniai jos valdymo aspektai

Kaip jau minėta, kredito rizika yra viena iš svarbiausių banko veikloje, nes paskolų portfelis paprastai sudaro pačią didžiausią banko aktyvų dalį ir dėl jos galimi dideli potencialūs nuostoliai. Ji apibrėžiama kaip nuostolis, kurį patiria bankas, kai sandorio šalis negali įvykdyti savo sutartinių įsipareigojimų bankui (Kancerevyčius, 2004).

Kredito rizika būna tokiose srityse:

- paskolų (skolų) išdavimas (grąžinimas);
- skolos instrumentų (pvz., obligacijų) nominalios vertės atgavimas;
- palūkanų už paskolą arba obligacijos kupono (atkarpos) mokėjimų gavimas;
- depozitų (indelių) priėmimo/davimo;
- išvestinių finansinių instrumentų (pvz., išankstiniai, pasirinkimo sandoriai ir kt.).

Bankas, išdavęs paskolą, nuolat stebi ir vertina kliento kredito mokumą ir tikrąją užstato vertę, tam naudodamas vidinių reitingų sistemą, padedančią įvertinti skolininko kredito kokybę, nustatyti atitinkamus kredito limitus ir stebėti rizikos dinamiką.

Vertinant bet kokią riziką, svarbu atsiminti, kad kiekvienu atveju ji apima du pagrindinius aspektus: nepatikimumą ir pavojų prarasti investuotas lėšas bei neapibrėžtumą. Kredito rizikos atveju, ją vertindama finansinė institucija turi apsvarstyti tris klausimus:

- įsipareigojimų nevykdymo tikimybė (angl. *probability default*) – kokia galimybė, kad kita šalis nevykdys savo įsipareigojimų jų gyvavimo metu arba tam tikru jų laikotarpiu, pavyzdžiui, metais? Vienu metų laikotarpio šis rodiklis gali būti apibrėžiamas kaip *tikėtinas įsipareigojimų nevykdymo dažnis*;

- kredito praradimas (angl. *credit exposure*) – kokio dydžio bus neapmokėta skola, kai atsisiras įsipareigojimų nevykdymas iš skolininko pusės;

- atgavimo rodiklis (angl. *recovery rate*) – kokia prarasto kredito dalis įsipareigojimų nevykdymo atveju gali būti susigrąžinta per bankroto procedūrą ar kitą atsiskaitymo būdą?

Kredito rizikoje taip pat naudojama ir sąvoka „kredito kokybė“, apibūdinanti skolininko galimybę atsiskaityti už įsipareigojimą. Ji apima tiek skolininko įsipareigojimų nevykdymo tikimybę,



ties ir atgavimo rodiklį.

Algoritminių metodų naudojimą kredito rizikos vertinimui galima apibrėžti sąvoka „kredito rizikos modeliavimas“, kuri apima tokias sritis, kaip tradicinė kreditų analizė, metodai ir modeliai, naudojami įvertinti derivatyvus, bei portfelių kredito rizikos matavimus, naudojamus viso obligacijų portfelio analizei[26]. Svarbu pabrėžti ir tai, kad kai kuriais atvejais kredito rizikos vertinimas yra komplikuoatas, unikalus, taip pat gali būti ir tokių atvejų, kai riziką geriau vertinti mažiau formaliu metodu. Pats kreditų kokybės vertinimo procesas apibrėžiamas kaip *kredito analizė* ir apima kredito rizikos vertinimą tiek algoritmiškai, tiek ir mažesnio formalumo metodais, o asmenys, atliekantys šį procesą, įvardinami kaip *kredito analitikai*. Šie asmenys pagal gautą informaciją apie būsimą skolininką įvertina kredito išdavimo galimybę, remdamiesi metiniais ir ketvirčių balansais, pajamų ataskaitomis, verslo šakos, kurioje veikia tiriamas subjektas, galimybėmis, ekonomine šalies situacija, įvairiais rodikliais ir kt. kriterijais, ir tuo remdamiesi sudaro *kredito reitingą*, kuriuo skolininką priskiria vienai ar kitai grupei pagal jo galimybes gražinti kreditą. Šiuo reitingu remiantis, kreditorius gali atlikti kredito išdavimo sprendimus. Tokius reitingus kitoms įmonėms kuria ir pasaulyje žinomos bendrovės *Fitch ratings, Moody's, Dunn & Bradstreet, Standard & Poor's* ir kt., kurių reitingai pasaulyje yra vertinami ir pripažįstami įvairių finansinių institucijų, teikiančių kreditavimo paslaugas.

### 1.2.3. Kredito rizikos vertinimui naudojami rodikliai ir kintamieji

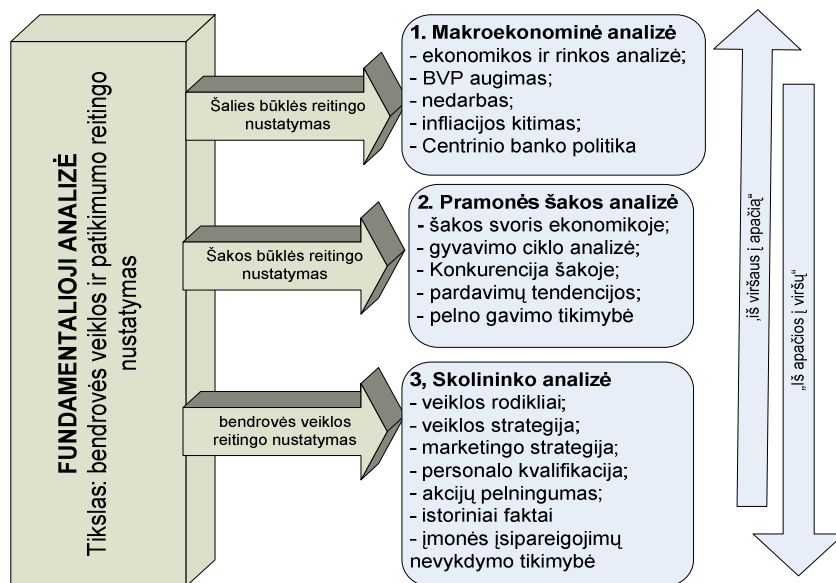
Kredito rizikos vertinimui gali būti naudojama įvairus kiekis kintamųjų, kurie yra tam tikro laikotarpio skolininko finansiniai rodikliai<sup>2</sup>. Kaip vėliau bus pastebėta, šis kiekis nėra konkrečiai apibrėžtas – klasikiniuose vertinimo balais ir reitingavimo modeliuose naudojama paprastai iki 10 kintamųjų, tačiau moderniuose metoduose šis skaičius gali siekti 30 ir daugiau. Kaip pavyzdį galima pateikti faktą, kai kurie autoriai, dirbę su neuroniniais tinklais, paprastai naudojo nuo 15 iki 25 kintamųjų; galima rasti ir atvejų, kai buvo naudotas 41 kintamasis – tiek kintamųjų savo neuroninių tinklų tyrime naudojo Lee ir kt. 1996 m. (Vellido, Lisboa, Vaughan). Didesnis kintamųjų skaičius leidžia įvertinti skolininko kredito riziką pagal didesnę kriterijų skaičių. Tai gali lemti ir tikslesnį kredito rizikos įvertinimą; tačiau reikia atsižvelgti ir į faktą, kad tada sumažėja duomenų prieinamumo ir gaunamumo galimybė, o tuo pačiu ir pilno duomenų rinkinio gavimo tikimybė, nes tokiu atveju gali smarkiai padidėti duomenų kiekis ir apimtis.

Vienareikšmiškai įvertinti atskirų kriterijų įtakos kreditingumui neįmanoma, kadangi vienuose modeliuose labiau atsižvelgiama į vienus kriterijus, kituose – į kitus. Šie parametrai priklauso ir nuo tokių aspektų, kaip:

---

<sup>2</sup> Išsamus rodiklių ir kriterijų, galimų naudoti kredito rizikos vertinimui, sąrašas pateikiamas 1 priede.

- kas vertinama (pramonės šakos būklė, bendrovės veikla, skolininko patikimumas),
- kokie yra jo šalies įstatymai, reglamentuojantys tiek jo bendrovės veiklą,
- mokesčiai tam tikrose šalyse,
- šalies finansinė būklė,
- pramonės ar verslo šakos, kurioje veiklą vysto kredituojamas asmuo ar bendrovė, perspektyvos, prioritetas šalyje, kurioje įsikūrusi kredituojama bendrovė, organizacija.



Šaltinis: sudaryta autoriaus.

### 11 pav. Skolininko analizės procesas

Atsižvelgiant į šiuos ir kitus parametrus, kreditorius (dažniausiai tai būna bankas ar kita bendrovė, teikianti tokio pobūdžio paslaugas) pasirenka tinkamiausią modelį skolos negrąžinimo galimybės įvertinimui arba pagal savo prioritetus bei vertinimo kriterijus susikuria ir pritaiko savo paties metodiką.

Svarbu paminėti ir tai, kad šie parametrai bus skirtingi, vertinant tiek atskirą šaką, tiek ir kredituojamojo kaip fizinio asmens ar bendrovės veiklą. Be to, dažniausiai siekiant įvertinti kredito riziką konkretaus subjekto atžvilgiu, svarbu įvertinti ne tik paties subjekto veiklos galimybes, bet ir jo veiklos makroekonominę aplinką – tai jau minėti šalies finansinė būklė, jos įstatymai, konkurentų veikla, verslo šakos ateities perspektyvos ir kt. Todėl kredito rizikos vertinimui šiuo atveju gali būti naudojamas fundamentinės analizės principas – ekonomikos ir rinkos analizė, šakos analizė, bendrovės veiklos analizė ir reitingo nustatymas (11 pav.). Ši analizė gali būti atlikta dviem būdais;

- Iš pradžių vertinama bendra šalies, kurioje yra skolininkas, bendra ekonomika ir makroekonominė aplinka, įvertinant tokius rodiklius, kaip infliacija; jei skolininkas yra įmonė, užsiamanti prekių/paslaugų gamyba, analizuojama ir prekybinė jo gaminamų prekių/teikiamų paslaugų

rinka. Jei šis subjektas yra ir emitentas (t.y., išleidęs akcijas, kuriomis prekiaujama rinkoje, gali būti analizuojama ir vertybinių popierių rinka. Po to analizuojama pramonės šaka, kuriai atstovauja skolininkas, tiriant konkurenciją, prekių realizavimo galimybes ir kitus panašius kriterijus. Vėliausiai analizuojama pati įmonė, įvertinant jos atskirus veiklos finansinius rodiklius bei veiklos, marketingo strategijas ir kitus jos veiklos aspektus, galiausiai apskaičiuojant įmonės išsipareigojimų tikimybę. Toks požiūris vadinamas „iš viršaus į apačią“ (angl. top-down).

- Atitinkamai iš pradžių įvertinama skolininko būklė, po to analizuojama šakų ir bendrą ekonominę situaciją. Toks požiūris vadinamas „iš apačios į viršų“ (angl. down-top).

Rekomenduojama kreditavimo įstaigoje atskirai vertinti bendras ekonomines tendencijas atspindinčių (t.y. pasižyminčių didele koreliacija su bendrais visų įmonių rodikliais) rodiklių kitimą ir konkrečios šakos specifinius rodiklius, pastariesiems suteikiant didesnę svorį (Dzidzevičiūtė, 2005).

Įvertinus šakos būklės vertinimo kriterijus, būtų gaunamas skolininko verslo šakos būklę atspindintis balas, leidžiantis įvertinti kredituojamo subjekto veiklos krypties perspektyvas ir galimybę ateityje padengti išsipareigojimus pagal veiklos srities vystymosi galimybes. Vėliau lygiai taip pat svarbu yra įvertinti ir bendrovės veiklą. Šiuo atveju būtų nagrinėjami tokie rodikliai, kaip pajamų ir pelno kitimas, išlaidų dydis ir struktūra, finansinių rodiklių dinamika, pinigų srantai, apyvarta, pelnas ir kt. Galima pastebėti, kad šių kriterijų nėra nei viename iš klasikinių kredito rizikos vertinimo modelių, tačiau į juos taip pat turi būti atsižvelgta, todėl jų analizė gali būti atlikta atskirai. Lygiai taip pat turi būti įvertintos paskolos grąžinimo galimybės, atliktas kredituojamo subjekto finansinių ataskaitų prognozavimas bei įvertinta skolininko išsipareigojimų nevykdymo tikimybės kitimo tendencija pagal prognozuojamų finansinių ataskaitų duomenis (Dzidzevičiūtė, 2005).

Įmonės (kredituojamojo subjekto) finansinė krizė gali būti įvertinta pagal tai, kaip kinta pagrindiniai finansiniai rodikliai. Absoliutų rizikos lygį lemia ekonomikos ciklas ir kiti išoriniai veiksniai. Pavyzdžiui, krizės metu net aukščiausio reitingo įmonių skolos išsipareigojimų nevykdymo tikimybė padidėja, lyginant su kitais laikotarpiais. Tačiau agentūros tikisi, kad aukštesnio reitingo emisijų išsipareigojimų nevykdymas įvyks rečiau negu žemesnių reitingų. Be to, agentūros stengiasi įvertinti ekonominius, politinius, kredito ir žaliavinių prekių kainų ciklus, kad vien tik dėl jų netektų keisti reitingo.

Panašius kriterijus naudoja ir tokios reitingavimo agentūros, kaip „Standard & Poor’s“ ir kt. Pastaroji vertina tiek šalis, tiek ir atskirus juridinius subjektus. Be to, anksčiau minėtoji „Standard & Poor’s“ agentūra atlieka ir šalių reitingavimą. Šalies reitingai (angl. *sovereign ratings*) gali būti suteikiami vyriausybės skolos emisijoms vietine ir užsienio valiuta. Šie reitingai gali turėti didelę įtaką kitiems šalies ūkio subjektų reitingams, nes nustato reitingams „šalies lubas“ (angl. *sovereign ceiling*). Paprastai šalyje veikiančios įmonės negauna geresnių reitingų negu šalies reitingai, nors

tokių atvejų paskutiniu metu pasitaiko vis daugiau. Šalies valiuta denominuotai skolai ūkio subjektas gali turėti daugiau likvidumo ir turto, lyginant su įsipareigojimais negu vyriausybė, ir tada jos reitingas gali būti aukštesnis. Esant įsipareigojimams užsienio valiuta, situacija yra keblesnė. Ūkio subjektas gali turėti geresnę padėtį negu vyriausybė, tačiau pastaroji gali nustatyti kapitalo judėjimo kontrolę, kuri apribotų valiutos konvertavimą. Tokiu atveju privačios įmonės kredito reitingą gali pakelti trečiosios aukštą reitingą turinčios šalies garantijos arba užstatas (garantas) užsienyje.

Apžvelgus kredito rizikos vertinimui naudojamus rodiklius, lengvai galima pastebėti, kad dauguma iš jų yra **kiekybiniai**, t.y., kintamieji, kuriuos galima išmatuoti iš anksto fiksuotoje matavimų skalėje pagal reikšmės dydį (V. Boguslauskas, 2007). Bet yra ir kita kintamųjų grupė, kurios kiekybinė išraiška gali būti įvertinta tik apytikriai ir subjektyviai vertinančio asmens atžvilgiu; šiuo atveju tai gali būti apibūdinama kaip ekspertinių žinių pritaikymas arba **kokybiniai** parametrai - kintamieji, galintys įgyti tik tam tikras fiksuotas, iš anksto apibrėžtas reikšmes (Boguslauskas, 2007). Tokie parametrai gali apimti priklausomybę nuo asmens amžiaus (tarkime, rizika nevykdyti įsipareigojimų bus didelė, jei asmens amžius daugiau nei 70 metų, arba žymiai mažesnė, jei asmuo 30-40 metų), darbo kvalifikacijos ir kitų subjektyvių kriterijų, ne visada išreiškiamų kiekybiškai. Be to, juos ne visada galima išreikšti algebriskai modelio matematinėje išraiškoje atskiru kintamuoju ar funkcija, kas sukelia keblumą juos įtraukiant į taikomą modelį bei pritaikant realiai.

8 lentelėje pateikiami tiek kiekybinių, tiek ekspertinių (nekiekybinių) parametrų pavyzdžiai.

8 lentelė. Kiekybiniai ir nekiekybiniai parametrai

<b>Kiekybiniai parametrai</b>	<b>Nekiekybiniai parametrai</b>
<b>Įmonės užimama rinkos dalis (procentais)</b>	Veiklos strategija
<b>Naudojamų technologijų galimybės</b>	Marketingo strategija
<b>Investicijų atsipirkimo periodas</b>	Planavimo lygis
<b>Grynoji dabartinė vertė</b>	Personalo kvalifikacija
<b>Investicijų pelningumas</b>	Prekių (paslaugų) realizavimo galimybės
<b>Alternatyviųjų investicijų grąža</b>	Prekių (paslaugų) gamybos ir gyvavimo ciklas
<b>Pajamų augimo tempas, pajamų augimo standartinis nuokrypis</b>	Turimos informacijos patikimumas
<b>Paskolos vieta kliento balanso struktūroje, jos gražinimo galimybės</b>	Apskaitos politikos vertinimas
<b>Finansinių ataskaitų prognozės</b>	Finansavimo šaltinių pritraukimo galimybės
<b>Pelno augimo tempas, pelno augimo tempo svyravimai</b>	Įvairių rizikų (vidinių konfliktų, gamybinės, technologinės veiklos) vertinimas
<b>Išlaidų struktūra</b>	Stabilumas
<b>Bendrasis, veiklos pelningumas</b>	Šaunaudų pastovumas
<b>Finansų struktūros rodikliai</b>	Žaliavų išlaidų įtaka šakos pelnui
<b>Prekybos skolų bei atsargų apyvartumas</b>	Įėjimo į sektorių barjerai, teisinis reglamentavimas
<b>Likvidumo, finansų struktūros, atsargų, veiklos rodikliai</b>	Tarptautinės ir vidaus konkurencijos vertinimas
<b>Turto grąža</b>	Šakos įmonių pardavimų kitimo tendencijos, perspektyvos

<b>Investicijų graža</b>	Gaminamų produktų ar teikiamų paslaugų specifika
<b>Bendrojo ir grynojo pelningumo rodikliai</b>	Sektoriaus koncentracija
<b>Nepaskirstyto pelno ir turto santykis</b>	BVP, užsienio valiutos ir palūkanų normų įtaka sektoriui
<b>Gamybos išlaidų ir pardavimų santykis</b>	Politinio reguliavimo rizikos vertinimas
<b>Gautinų sumų ir grynojo apyvartinio kapitalo apyvartumo rodikliai</b>	
<b>Grynojo apyvartinio kapitalo apyvartumas</b>	
<b>Darbo jėgos kaštai</b>	
<b>Skolinto kapitalo lygis</b>	
<b>Turto vertės kitimas</b>	

Šaltinis: sudaryta autoriaus pagal įvairius šaltinius 1 priede

#### 1.2.4. Kredito rizikos vertinimas, naudojant reitingavimą

Rizikos, atsirandančios išduodant paskolas individualiems asmenims ar mažoms organizacijoms, vertinimo atveju kredito kokybė paprastai įvertinama kredito vertinimu balais (angl. *credit scoring*): kreditorius iš turimos informacijos apie subjektą, kuriam išduodama paskola, apskaičiuoja kredito įvertinimą balais, tam naudodamas tam tikrą standartizuotą formulę. Pagal šį rodiklį kreditorius gali spręsti, ar jam verta suteikti kreditą konkrečiam subjektui, ar to daryti neverta. Kaip jau minėta, reitingavimą naudoja ir kredito reitingavimo agentūros, tam naudodamos skirtingus rodiklius. Pavyzdžiui, „Standard & Poor’s“ reitinguodama skaičiuoja bendrą išsipareigotojo pajėgumą įvykdyti finansinį išsipareigojimą, ir tai atlieka apskaičiuodama išsipareigojimų neįvykdymo rodiklį (*probability default*). Tuo tarpu „Moody’s“ į vertinimą įtraukia ir sprendimą apie skolos atgavimą esant nuostoliams, todėl skaičiavimai artimi laukiamų praradimų (angl. *expected loss*) rodiklio skaičiavimams. Šios bendrovės, reitinguodamos bendroves, vadovaujasi tokiais kriterijais, kaip verslo šakos ar pramonės šakos rizika, subjekto vieta šioje šakoje, jo veiklos ir efektyvumo rodikliai, finansinis pajėgumas bei lankstumas ir kt.

**Kredito reitingų sistema** suskirsto skolininkus/paskolas į homogenines grupes (reitingus) pagal jų kredito riziką. Gali būti reitinguojamas tiek skolininkas, kaip asmuo, įvertinant visas paskolas, kurias jis turi; o taip pat gali būti reitinguojama konkreti paskola ir jos rizika. Pastaruoju atveju paprastai įvertinama ir paskolos užtikrinimo priemonė, tam naudojant 2 lygių reitingavimo sistemą – paskola ir paskolos užtikrinimo priemonė). Reitingavimas paprastai naudojamas juridiniams asmenims. Fiziniais asmenims reitingavimo atitikmuo būtų vertinimas balais (angl. „scoring“), kuris pagrįstas iš esmės tik kiekybine analize.

Beveik visos įmonių ir municipalinės obligacijos turi reitingą. Tik kai kurios specifinės, pvz. bankų obligacijos neturi reitingų. Tai neturinčios reitingo obligacijos (angl. „nonrated bonds“). Obligacija yra sutartis, pagal kurią obligacijos investuotojas skiriasi su savo pinigais esamu momentu, o skolininkas pasižada, kad už tai, jog gauna finansavimą iš investuotojo dabar, jis apmokės skolą

paskirstytai mokėjimais ateityje. Reitingavimo paslaugas teikiančios agentūros teigia, jog jos gali suteikti investuotojui informaciją apie skolininko galimybes ir norą įgyvendinti sutarties sąlygas. Gali būti įtraukiama ir tikimybė, kad finansavimas bus gražintas su prieaugiu pagal nustatytą tvarkaraštį, o jeigu skola nebegalės būti apmokėta, bus numatyta, kokią dalį paskolintos sumos galima susigražinti ir kada. Šios agentūros traktuoja reitingus kaip nuorodą į ateitį apie santykinę riziką, kad skolos emitentas pajėgs ir norės atlikti pilnus ir savalaikius palūkanų bei nominalo mokėjimus; jos nesiekia, kad reitingai suteiktų informaciją apie nemokumo momentą arba apie absoliutų rizikos lygį, susijusį su atskiru instrumentu.

### 1.2.5. Kredito rizikos vertinimas balais

Šiuo metu žinoma daug įvairių modelių kredito rizikos vertinimui balais; kaip galima pastebėti iš 4 lentelės, beveik visi populiariausi ir plačiausiai naudojami (arba dauguma iš jų) paremti tiesine diskriminantine analize arba netiesine logit priklausomybės forma (pvz. Shumway, Grigaravičiaus, Ohlson modeliai). Pastarosios priklausomybės išraiška: 
$$p(x) = \frac{1}{1 + e^{-(a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n)}}$$

Skirtingose šalyse naudojami skirtingi modeliai, atitinkantys konkrečios šalies finansų sistemos specifiką. Šių modelių palyginimai bei jų naudojimo specifika plačiai aprašomi įvairiuose užsienio autorių darbuose. Šiuo metu plačiai žinomi ir taikomi Altmano Z, Altmano modelis gamybinėms ir negamybinėms įmonėms, Springate, Fulmer, Shumway, Zmijewski, Ohlson, taip pat Lietuvoje sukurtas Grigaravičiaus modeliai<sup>3</sup> (L.Dzidzevičiūtė, 2005).

Kai kurių iš paminėtų modelių taikymas Lietuvos sąlygomis plačiau nagrinėjamas L. Dzidzevičiūtės darbe, kuriame autorė pateikia ir jų privalumus ir trūkumus. Šie kriterijai, atmetus duomenų prieinamumo ir trūkumo aspektus, pateikiami 9 lentelėje.

9 lentelė. Vertinimo balais modelių esminiai privalumai ir trūkumai

Modelis	Privalumai	Trūkumai
<b>Altmano modelis (nuosavybės rinkos vertės vertinimas)</b>	1. Įvertinama nuosavybės rinkos vertė. 2. Didelis tikslumas, turint audituotas finansines ataskaitas	1. Nėra galimybės įvertinti įmonių, kurių akcijomis neprekiuojama biržoje 2. Tiesinė priklausomybė
<b>Altmano modelis (negamybinėms įmonėms)</b>	1. Nevertinamas pardavimų santykis - galimybė palyginti skirtingų šakų kreditingumo rodiklius. 2. Didelis tikslumas, turint audituotas finansines ataskaitas.	1. Pardavimų dydžio neįvertinimas, nagrinėjant įmonės kreditingumo balą skirtingais laikotarpiais. 2. Tiesinė priklausomybė
<b>Shumway modelis</b>	1. Logistinės regresijos forma 2. Koreliuoja su paskolų kokybės rodiklių dydžiais 3. Įsipareigojimų nevykdymo tikimybės	1. Galimas mažesnis tikslumas, neturint išsamesnių duomenų.

<sup>3</sup> 2 priede pateikiami šių modelių kintamųjų išraiškos ir matematiniai aprašai.

Modelis	Privalumai	Trūkumai
	radimas be papildomų transformacijų	
<b>Zmijewski modelis</b>	1. Įsipareigojimų nevykdymo tikimybės radimas be papildomų transformacijų 2. Koreliuoja su paskolų kokybės rodiklių dydžiais	1. Galimas mažesnis tikslumas, neturint išsamesnių duomenų
<b>Springate modelis</b>	1. Galimybė adaptuoti modelį bankams, pakoregavus vertinimo skalių reikšmes.	1. Per griežtas skolininkų diferencijavimas 2. Tiesinė priklausomybė

Šaltinis: Laima Dzidzevičiūtė. Kredito rizikos vertinimo bankuose ypatumai: vertinimo balais modelių taikymas Lietuvos bankuose.

Galima pastebėti, kad iš aukščiau aprašytų modelių daugiausia rodiklių apima Grigaravičiaus ir Fulmerio modeliai, kas leidžia daryti prielaidą, kad šie modeliai geriausiai įvertina kredito riziką, kai žinoma daugelis skolininko finansinių rodiklių (tai nereiškia, kad šie modeliai yra patys tiksliausi ir patikimiausi). Be šių modelių pasaulyje naudojama ir gausybė kitų mažiau žinomų modelių vertinimui balais, kai kada apimančių ir dar daugiau kriterijų, leidžiančių tuo pačiu apibrėžti skirtingų finansinių rodiklių spektrą bei svarbą kredito rizikos vertinimo procese šiose šalyse. Šių modelių sudarytą klasifikaciją galima rasti Dzidzevičiūtės darbe. Akivaizdu, kad tarp šių modelių autorių vyrauja Altman pavardė, kas leidžia daryti prielaidą, kad dalis iš šioje lentelėje išvardintų šalių naudoja modifikuotą ir jų sąlygoms adaptuotą klasikinį Altman modelį. Šių modelių aprašymus ir panaudojimo galimybes galima rasti įvairiuose užsienio autorių darbuose.

#### 1.2.6. Modernieji kredito rizikos vertinimo modeliai

Besivystant tiek kredito rizikos vertinimui kaip mokslo sričiai, tiek ir atsirandant naujoms techninėms galimybėms, tokioms, kaip modernios informacinės technologijos, leidžiančios daugelį skaičiavimų atlikti žymiai greičiau, o tuo pačiu ir naudoti žymiai sudėtingesnius metodus bei algoritmus, buvo sukurti ir praktikoje pritaikyti nauji ir tobulesni modeliai, skirti kredito rizikos vertinimo modeliavimui ir prognozavimui.

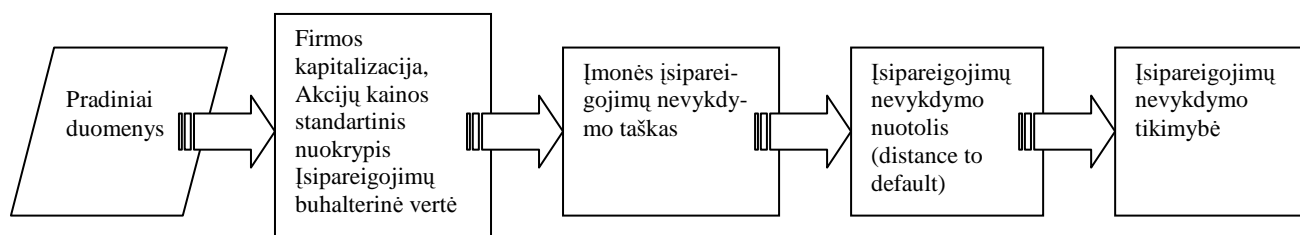
Šie modeliai, remiantis N.Arora, J.Elizande (2003), L. Allen (2002) ir kitais autoriais, gali dar būti suskirstyti į struktūrinius ir supaprastintos formos modelius. Struktūriniai modeliai naudoja firmos vertę praradimų laikotarpiui vertinti, kai tuo tarpu supaprastintos formos modeliai skirti įsipareigojimų neįvykdymo kaip šuolinio proceso, kuriame šuolio laikas yra neįvykdymo laikas, intensyvumui nustatyti (J.Elizande, 2003). Supaprastintos formos modeliuose naudojami intensyvumu grįsti modeliai, siekiant apskaičiuoti stochastinius rizikos lygius. Struktūrinio požiūrio modeliai (pvz. KMV Credit Manager, Moody's RiskCalc), kurių pradininku laikomas Mertonas, modeliuoja ekonominių įsipareigojimų nevykdymo procesą. (Allen L., 2002). Taip pat reikia pabrėžti ir tai, kad šie požiūriai skiriasi ir pagal įsipareigojimų nevykdymo tikimybės skaičiavimą.

Pačiu pirmuoju moderniu modeliu ir apylamai įsipareigojimų neįvykdymo vertinimo modeliu

laikomas Mertono modelis (1974 m.). Jame skolininkas turi neapmokėtą nulinę skolą, o firmos aktyvų vertės vystymasis auga pagal geometrinę Brauno judėjimo funkciją. Šiuo atveju firma turi įsipareigojimų, jei aktyvų vertė yra žemiau jos neapmokėtos skolos t.y., priimama, kad įsipareigojimų nevykdymo riba yra nominalioji skolos vertė. Taigi svarbiausia Mertono modelio charakteristika yra įsipareigojimų neįvykdymo laiko apribojimas iki skolos termino, pašalinant galimybę ankstyvam neįvykdymui, nepaisant to, kas atsitinka su firmos verte iki skolos termino. Jei firmos vertė iki skolos termino nukrenta iki minimalaus lygio, bet gali “atsigauti” ir atitikti skolos mokėjimą termino suėjimo metu, pagal šį požiūrį įsipareigojimų nevykdymo turėtų būti vengiama. Kitas Mertono modelių aspektas yra kredito rizikos prognozavimas. Šiuo atveju firmos aktyvų vertė modeliuojama kaip Brauno judėjimas ir įsipareigojimai nevykdomi tik tada, kai yra skolos terminas; jie gali prognozuojami su didėjančiu tikslumu, artėjant skolos grąžinimo laikotarpiui. To rezultatas – generuojami labai maži trumpalaikiai kreditų pasiskirstymai, kadangi kredito rizikos tikimybė labai trumpame laiko intervale yra arti nulio (J.Elizande, 2003).

Dar vieno iš šių požiūrių įsipareigojimų nevykdymo atvejų atveju įsipareigojimų nevykdymas atsitinka tada, kai firmos aktyvų vertė nukrenta žemiau tam tikro slenksčio; priešingai nuo Mertono požiūrio, tai gali atsitikti bet kada (Black & Cox modelis, sukurtas 1976 m.; J.Elizande, 2003). Tokie modeliai dar vadinami pirmojo perėjimo modeliais (angl. *First Passage Models*) (J.Elizande, 2003).

Iš šių metodų išsivystė daugelis plačiai taikomų modelių, naudojančio vieno ar daugiau žingsnių schemas. Pavyzdžiui, Moody’s KMW modeliui būdinga trijų žingsnių schema, skirta (angl. *Expected Default Frequency*, sutr. EDF) parametro apskaičiavimui (3 pav.)



**12 pav. Laukiamo įsipareigojimų nevykdymo dažnio apskaičiavimas pagal Moody’s KMW**

Šaltinis: sudaryta autoriaus.

Šiuo metu plačiai nagrinėjami ir taikomi tokie modernūs modeliai, kaip:

- CreditMetrics;
- Credit Risk Plus;
- KMW/Moody’s (Merton OPM);
- Sumažintos formos KPMG/Kamakura.



Šių modelių pagrindinės charakteristikos ir palyginimas pateikiami 10 lentelėje. Galima pasitebėti, kad daugumoje atvejų rizikos klasifikavimui naudojamas grupavimas arba reitingai; tai sudaro prielaidas šiuose modeliuose panaudoti ir klasifikavimo/klasterizavimo metodiką bei jos elementus.

10 lentelė. Moderniųjų kredito rizikos vertinimo modelių palyginimas

<b>Charakteristikos</b>	<b>CreditMetrics</b>	<b>Credit Risk Plus</b>	<b>Merton OPM KMW/Moody's</b>	<b>Sumažintos formos KPMG/Kamakura</b>
<b>Rizikos apibrėžimas</b>	Rinkos vertės modelis (mark-to-market)	Įsipareigojimų nevykdymo (default mode)	Rinkos vertės arba įsipareigojimų nevykdymo modelis	Rinkos vertės modelis (mark-to-market)
<b>Rizikos faktoriai</b>	Turto vertės	Laukiami įsipareigojimų nevykdymo lygiai	Turto vertės	Skolos ir nuosavybės kainos
<b>Duomenų reikalavimai</b>	Pasikeitimo tikimybių matricos, kredito maržų ir pelningumų kreivės, LGD, koreliacijos, kredito pozicijos dydis	Įsipareigojimų nevykdymo lygiai ir kintamumas, makroekonominiai faktoriai, LGD, kredito pozicijos dydis	Nuosavybės kainos, kredito maržos, koreliacijos, kredito pozicijos dydis	Skolos ir nuosavybės kainos, pasikeitimo tikimybių matricos, koreliacijos, kredito pozicijos dydis
<b>Kredito įvykių charakteristika</b>	Kreditingumo pasikeitimas	Aktuarinis atsitiktinių įsipareigojimų nevykdymų lygis	Atstumas iki įsipareigojimų nevykdymo: struktūrinis ir empirinis	Įsipareigojimų nevykdymo intensyvumas
<b>Kredito įvykių kintamumas</b>	Pastovus arba kintamas	Kintamas	Kintamas	Kintamas
<b>Kredito įvykių koreliacija</b>	Turto grąžos	Nepriklausomumo prielaida arba koreliacija su laukiamu įsipareigojimų nevykdymo lygiu	Turto grąžos	Poisono intensyvumo procesai su jungtiniais sistemiais faktoriais
<b>Paskolos grąžinimo lygis</b>	Atsitiktinis (beta pasiskirstymas)	Pastovus grupėje	Pastovus arba atsitiktinis	Pastovus arba atsitiktinis
<b>Numerologinis požiūris</b>	Simuliacija arba analitinis požiūris	Analitinis požiūris	Analitinis arba ekonometrinis požiūris	Ekonometrinis požiūris
<b>Palūkanų normos</b>	Pastovios	Pastovios	Pastovios	Stochastinės
<b>Rizikos klasifikacija</b>	Reitingai	Grupės pagal kredito pozicijas	Empirinis įsipareigojimų nevykdymo dažnis (EDF)	Reitingai arba kredito maržos
<b>Privalumai</b>	Rinkos informacijos naudojimas padidintų rezultatų tikslumą	Labiausiai tinkamas metodas paskolų registro mastu	Rinkos informacijos naudojimas padidintų rezultatų tikslumą	Rinkos informacijos naudojimas padidintų rezultatų tikslumą
<b>Trūkumai</b>	Skolininkų kredito reitingų, reitingų pasikeitimo tikimybių, LGD	Duomenų įsipareigojimų nevykdymo santykiniams dažniui ir grąžinimo lygio	Rinkos duomenų stygius; sunku taikyti paskolų registro mastu	Rinkos duomenų stygius; sunku taikyti paskolų registro mastu

Charakteristika	CreditMetrics	Credit Risk Plus	Merton OPM KMW/Moody's	Sumažintos formos KPMG/Kamakura
	ir koreliacijos statistikos nebuvimas. Diskonto palūkanų normų nebuvimas (neišvystytos finansų rinkos problemos).	skaičiavimui stygius		

Šaltinis: sudaryta autoriaus pagal (Allen, 2002; Dzidzevičiūtė, 2005).

Kiekvienas iš nagrinėjamų modelių turi savo taikymo ypatumus, o taip pat ir privalumus bei trūkumus. Šie aspektai nagrinėjami įvairių autorių darbuose. Siekiant išnagrinėti modernių modelių privalumus ir trūkumus, reikėtų apžvelgti pagrindinių modelių ypatumus ir kritiką, kadangi, kaip jau prieš tai buvo minėta, būtent šių modelių pagrindu ir yra sukurti šiuolaikiniai kredito rizikos vertinimo metodai. J. Elizande taip pat aprašo Mertono modelio privalumus ir pateikia jo kritiką; taip pat pateikiama ir pirmojo perėjimo modelių kritika. Kaip teigiama jo darbe, pagrindinis Mertono modelio privalumas yra tai, kad jis leidžia tiesiogiai taikyti 1973 m. Black ir Scholes sukurta Europos opcionų įvertinimo teoriją, tačiau tam reikia, kad būtų įvertinti svarbūs kriterijai, siekiant pritaikyti įmonės aktyvų vertės procesų dinamiką, palūkanų normas ir kapitalo struktūrą pagal Black-Scholes modelio reikalavimus.

Praktinis pritaikymas buvo nagrinėtas daugelyje šaltinių. L. Dzidzevičiūtė, atlikusi analizę, kuria buvo siekiama iširti šių modelių galimybes pritaikyti Lietuvos bankuose, pateikė išvadą, kad Lietuvos sąlygomis tinkamiausias yra Credit Risk+ vertinimo modelis, kuris galėtų būti pritaikytas atidėjimų sudarymui, limitų nustatymui, kreditų portfelio valdymui (Dzidzevičiūtė, 2005). Šiuo atveju rizika klasifikuojama pagal kredito pozicijų grupes; tai sudaro prielaidas sukurti hibridinį kredito rizikos vertinimo modelį, geriausiai pritaikytą Lietuvos bankų sistemai, naudojant vieną iš dirbtinio intelekto metodikų, skirtų klasifikavimui, bei CreditRisk+ modelį.

### 1.3. Ekspertinės sistemos ir sprendimų priėmimas

Kredito rizikos vertinimas tapo efektyvesnis, pradėjus naudoti informacines technologijas, kas leido automatizuoti daugumą skaičiavimų ir atlikti tokio tipo analizę ir prognozavimą efektyviau ir greičiau. Paskutiniame XX a. dešimtmetyje padėti naudoti ir tuo metu plačiai vystytos ir šiuo metu tebevystomos kompiuterių mokslo šakos – dirbtinio intelekto – metodai, leidę šiai sričiai pritaikyti visai naujo pobūdžio metodiką, kuri kai kuriais atžvilgiais pasirodė efektyvesnė nei iki tol naudoti klasikiniai modeliai. Norint apibūdinti šiuos metodus, tikslinga visų pirma apibrėžti pačią dirbtinio intelekto (angl. *Artificial Intelligence*, sutr. AI) sąvoką bei jos naudojimo kredito rizikos

vertinime istorines prielaidas.

Pati dirbtinio intelekto sąvoka yra plati ir galbūt todėl sunkiai apibrėžiama; dėl šios priežasties nėra vienareikšmio jos apibrėžimo, nes beveik kiekvienas šaltinis pateikia savo apibrėžimą ar net kelis apibrėžimus. Čia pateikiami tik keletas jų (visi, išskyrus paskutinįjį, pateikiami Russell ir Norwig knygoje "Artificial Intelligence: A Modern Approach"):

- „Nuostabus naujas bandymas priversti kompiuterius mąstyti... mašinos su protais, pilna ja ir literatūrine prasme“ (*Haugeland, 1985*);
- „[Automatizavimas] veiklos, kurią mes siejame su žmogaus mąstymu, tokios kaip sprendimų priėmimas, problemų sprendimas, mokymasis...“ (*Bellman, 1978*);
- „Mašinų, kurios atlieka funkcijas, reikalaujančias intelekto tada, kai jos atliekamos žmonių, kūrimas“ (*Kurzweil, 1990*);
- „Mokslas, kaip priversti kompiuterius atlikti darbus, kuriuose šiuo metu žmonės yra geresni“ (*Rich and Knight, 1991*);
- „Žmogui būdingų pažintinių ar suvokimo pojūčių studija, naudojant skaičiavimo modelius“ (*Charniak and McDermott, 1985*);
- „Skaičiavimų, įgalinančių suprasti, samprotauti ir veikti, studija“ (*Winston, 1992*);
- „Studijų kryptis, kuri siekia paaiškinti ir kopijuoti intelektinį elgesį skaičiavimo procesu terminais“ (*Schalkoff, 1990*);
- „Kompiuterių mokslo šaka, susijusi su protingo elgesio automatizavimu“ (*Luger and Stubblefield, 1993*);

„Dirbtinis intelektas yra dalis kompiuterių mokslo, susijusio su kompiuterinių sistemų, kurios demonstruoja charakteristikas, kurias mes siejame su protu žmogus elgesyje – supratimu, kalba, mokymusi, samprotavimu, problemų sprendimu ir kt., kūrimu...“ (*Barr and Feigenbaum, Handbook of AI*).

### **1.3.1. Ekspertinių sistemų evoliucija ir jų taikymo kredito rizikos vertinime prielaidos**

• Dirbtinio intelekto sąvoka atsirado prieš pusę amžiaus, 1956 m., nors ši sritis kaip intelektika buvo nagrinėjama dar maždaug prieš 2000 m. 6-ajame šio amžiaus dešimtmetyje pradėti naudoti kompiuteriai šią sąvoką pavertė rimta teorine ir praktine disciplina, kuri pasirodė gerokai sudėtingesnė nei manyta iki tol. Tuo metu buvo sukurtos pirmosios dirbtiniu intelektu pagrįstos programos, leidusios panaudoti šios srities žinias praktiniam problemų sprendimui.

Visi dirbtinio apibrėžimai gali būti suvesti į keturias grupes (12 lentelė).

<b>Sistemos, mąstančios kaip žmonės</b>	<b>Sistemos, mąstančios racionaliai</b>
<b>Sistemos, besielgiančios kaip žmonės</b>	<b>Sistemos, besielgiančios racionaliai</b>

Šaltinis: Graham Kendall. History of AI .

Kaip galima pastebėti nagrinėjant DI istoriją, vienu ar kitu metu buvo laikomasi visų keturių sampratų. Sistemų, kurios elgiasi kaip žmonės, sąvoką atspindi *Tiuringo testas*<sup>4</sup>, kurio pagrindinis subjektas yra pripažinimas, kad mašina gali mąstyti, tuo tarpu požiūriui, esančio lentelėje viršuje kairėje, svarbiausia yra pats žmogaus mąstymo metodas, tačiau tai gali būti nusakyta tik patekus į patį žmogaus protą, kas šiuo metu dar neįmanoma, arba vykdant psichologinius eksperimentus, kas susiję su psichologija ir ne visada tinka kuriant atitinkamas sistemas. Sistemos, galinčios mąstyti racionaliai, yra paremtos logikos taisyklėmis, pavyzdžiui, iš teiginių „Universitetas yra Kaune“ ir „Kaunas yra Lietuvoje“ gauname teiginį „Universitetas yra Lietuvoje“. Tas turėtų leisti nesunkiai sukurti tokias sistemas, tačiau iškyla dvi problemos: gana sudėtinga atitinkamą problemą transformuoti į formalias taisykles, ypač jei ji mažiau nei 100% teisinga, bei gana daug skaičiavimų reikalaujantis procesas, kas neretai problemą sprendžia ilgiau nei ji to verta. Todėl dažniausiai laikomasi paskutiniojo požiūrio, teigiančio, kad sistemos turi elgtis racionaliai.

1976 m. Newell ir Simon pabandė apibrėžti intelektą taip, kaip jis turėtų būti atvaizduotas kompiuteryje. Jie teigė: „Fizinė sistema susideda iš aibės esybių, vadinamų simboliais, kurie yra fiziniai šablonai, galintys būti kaip komponentai kito esybės tipo, vadinamo išraiška (arba simbolinė struktūra). Tada simbolinė struktūra yra sudaryta iš tam tikro skaičiaus simbolių, susijusių tam tikru fiziniu būdu (kaip, pavyzdžiui, vieno simbolio buvimas šalia kito. Bet kuriuo metu sistema turi tokių simbolių struktūrų rinkinį. Be šių struktūrų, sistema taip pat turi procesų, valdomų tam tikromis išraiškomis, leidžiančiomis sukurti kitas išraiškas, rinkinį; tai sukūrimo, modifikavimo, pakartotino sukūrimo bei sunaikinimo procesai. Fizinė simbolinė sistema yra mašina, kuri per tam tikrą laiką sukuria besivystantį simbolių struktūrų rinkinį; tokia sistema egzistuoja objektų pasaulyje, platesniame, nei šios simbolinės išraiškos“ (Kendall, 2005). Remdamiesi šiuo teiginiu, jie iškėlė fizinių simbolių sistemos hipotezę (*The Physical Symbol System Hypothesis*, sutr. *PSSS*), kuri teigia, kad fizinė sistema turi būtinas ir pakankamas priemones bendrajam intelektiniam veiksmui. Reikia akcentuoti, kad ši hipotezė taip ir liko tik hipoteze, ji nėra įrodyta, remiantis logika ar pan., tačiau,

<sup>4</sup> Testas, pasiūlytas 1950 m. Alan Turing. Jo esmė yra tokia: yra du stebėtojai A ir B bei kompiuteris; vienas iš šių stebėtojų sėdi gretimame kambaryje, o antrasis tuo metu uždavinėja klausimus kompiuteriui, į kuriuos kompiuteris pateikia atsakymus. Stebėtojas A, nežinodamas, kuris iš pokalbio dalyvių yra kompiuteris, iš gaunamų atsakymų bando tai atspėti. Jei kompiuteriui pasiseka apgauti stebėtoją A, laikoma, kad jis praėjo Tiuringo testą, ir galima daryti išvadą, kad kompiuteris gali mąstyti.

remiantis šiuo teiginiu, buvo atlikta dauguma darbo dirbtinio intelekto srityje. Vėliau buvo sukurta nemažai intelektualių sistemų, kurios galėjo atlikti tam tikrus veiksmus, būdingus žmogui, bei galėjo vienu ar kitu aspektu imituoti jo veiklą (12 lentelė).

12 lentelė. Pirmosios dirbtiniu intelektu paremtos programos

Sistema	Autorius, metai	Paskirtis
<i>Logic Theorist</i>	Allen Newell, Herbert Simon, and J.C. Shaw, 1955 – 1956	Pirmoji DI programa, skirta teoremų įrodymui.
<i>Geometry Theorem Prover</i>	Herbert Gelernter, 1957	DI programa, kurios pagrindinė paskirtis – geometrinių teoremų įrodymas. Ši programa milijardą alternatyvų trijų žingsnių geometrijos teoremos įrodymui galėjo susiaurinti iki 25 alternatyvų
<i>General Problem Solver (GPS)</i>	Newell and Simon, 1963	Pirmoji dirbtinio intelekto programa, realizuojanti Means-End analizės idėją. Šios idėjos esmė yra ta, kad problema išskaidoma į mažesnes problemas, o šių problemų sprendimai apjungiami į vieną. Tai yra, randami skirtumai pradinės ir galutinės problemos padėčių ir tada bandoma rasti operatorių, kuris judės nuo pradinės padėties į galutinę; jei tokio operatoriaus rasti nepavyksta, tai problema skaidoma į mažesnes problemas, kurios sprendžiamos remiantis tuo pačiu principu.
<i>ELIZA</i>	Weizenbaum, 1966	Viena pirmųjų DI programų. Jai buvo galima užduoti klausimus, ir ji į juos atsakydavo.
<i>MYCIN</i>	Shortliffe, 1976	Programa, galinti diagnozuoti infekcinius susirgimus. Ji laikoma viena pirmųjų ekspertinių sistemų, kadangi gydytojų žinios buvo atvaizduojamos kompiuteryje.
<i>DENDRAL</i>	Edward Feigenbaum, Robert K. Lindsay, 1965-1975	Ekspertinė sistema organinių junginių struktūros nustatymui ir analizei.
<i>XCON/RI</i>	DEC (Digital Equipment Corporation), 8 deš.	Pirmoji komercinė DI sistema, užtikrinusi, kad vartotojas būtų aprūpintas visais komponentais ir programine įranga, reikalinga jo užsakytai kompiuterinei sistemai. Tai sudėtingas uždavinys, kadangi skirtingai sistemai reikalingi skirtingi komponentai, pavyzdžiui, užsakius kietąjį diską, turi būti pristatyti atitinkamas kontroleris ar laidai, o pati sistema gali susidėti iš tūkstančių įvairių komponentų.

Šaltinis: sudaryta autoriaus pagal [Kendall,2005; Henke, 2005].

Svarbus lūžis įvyko, DEC kompanijai pradėjus naudoti XCON sistemą prekyboje. Tai ne tik kompanijai sutaupė milijonus JAV dolerių ir kilstelėjo tokio tipo sistemų kartelę į daug aukštesnį lygį, bet ir pradėjo ekspertinių sistemų naudojimo komercijoje erą. Atsiradus pirmosioms dirbtiniu intelektu paremtoms komercinėms sistemoms, į šią kompiuterių mokslo šaką buvo atkreiptas didesnis dėmesys ir ji buvo imta rimčiau taikyti sistemų, naudojančių bazę tam tikrų sprendimų priėmimui, arba *ekspertinių sistemų* kūrimui. Jau 9-ojo dešimtmečio pradžioje buvo sukurtos ir kompanijose įdiegtos komercinės ekspertinės sistemos, tokios, kaip mineralų stebėjimo sistema PROSPECTOR, radusi molibdeno nuosėdas, Bell sistema telefonų tinklų analizei, audrų stebėjimo sistema

WILLARD meteorologams ir kt. Tarp jų buvo ir pirmoji finansinė ekspertinė sistema FOLIO, skirta vertybinių popierių portfelio analizei.

Maždaug tuo pačiu metu atsirado ir naujos dirbtinio intelekto sritys – neraiškioji logika (*fuzzy logic*), neuroniniai tinklai (*neural networks*), kurie 1982 m. John Hopfield buvo pritaikyti skaičiavimams. Būtent ši technologija 9-ojo dešimtmečio viduryje ir tapo viena labiausiai komercinės kompanijas sudominusių sričių, kas pradėjo tikrą ekspertinių sistemų kūrimo bumą. Reikia pabrėžti, kad svarbus kriterijus šios srities vystymuisi buvo ir statistikos mokslo vystymasis, leidęs įvairiose ekspertinėse sistemose naudoti tiek statistinius, tiek ir hibridinius DI modelius, kuriuose galėjo būti taikomi šie statistiniai metodai. Visa apibendrinta dirbtinio intelekto, duomenų gavybos mokslo ir statistikos metodų atsiradimo seka pateikiama 13 lentelėje.

13 lentelė. DI, duomenų gavybos ir statistikos mokslų vystymosi evoliucija

Metai	Sritis	Metodas
XVIII a.	Statistika	Bajeso tikimybės teorema
XX a. pradžia	Statistika	Regresinė analizė
XX a. 3 deš.	Statistika	Maksimalios galimybės įvertis ( <i>Maximum likelihood estimate</i> )
XX a. 5 deš.	DI	Neuroniniai tinklai
XX a. 6 deš.		„Artimiausio kaimyno“ metodas
XX a. 6 deš.		Vienas ryšys ( <i>Single link</i> )
XX a. 6 deš.	DI	Perceptronas
XX a. 6 deš.	Statistika	Pavyzdžių perrinkimas ( <i>resampling</i> ), paklaidos mažinimas ( <i>bias reduction</i> ), <i>jackknife</i> įvertis
XX a. 7 deš.	DI	Mašininis mokymas
XX a. 7 deš. vidurys	DB	„Paketinės“ ataskaitos ( <i>batch reports</i> )
	DM	Sprendimų medžiai
	Statistika	Tiesiniai klasifikavimo modeliai
	IR	Panašumo įverčiai
	IR	Klasterizavimas
	Statistika	Tiriamoji duomenų analizė ( <i>Exploratory Data Analysis</i> )
XX a. 7 deš. pabaiga	DB	Reliacinis duomenų modelis
XX a. 8 deš. pradžia	IR	„Protingos“ informacijos išgavimo sistemos ( <i>SMART IR systems</i> )
XX a. 8 deš. vidurys	DI	Genetinis algoritmas
XX a. 8 deš. pabaiga	Statistika	Vertinimas pagal nepilnus duomenis (EM algoritmas)
XX a. 8 deš. pabaiga	Statistika	<i>K-means</i> klasterizavimas
XX a. 9 deš. pradžia	DI	Kohoneno saviorganizuojantis žemėlapis
XX a. 9 deš. vidurys	DI	Sprendimų medžio algoritmai
XX a. 10 deš. pr.	DI	Asociatyvinių taisyklių algoritmai
		Interneto ir paieškos sistemos
XX a. 10 deš.	DB	Duomenų sandėliavimas
XX a. 10 deš.	DB	Online analitinis duomenų apdorojimas (OLAP)

Šaltinis: Margaret H. Dundam. Data Mining: Introductory and Advanced Topics

Eksploatuojant taisyklėmis pagrįstas ekspertines sistemas, buvo pastebėti ir pagrindiniai jų trūkumai – nelankstumas, leidęs priimti tik riboto lankstumo sprendimus, negalėjimas atsižvelgti į panašius praeities įvykius, kas paprastai būdinga žmogiškajam mąstymui, neteisingų sprendimų

galimybė, galimybės tokioms sistemoms bendrauti tarpusavyje nebuvimas, sudėtingas tokių sistemų palaikymas ir kt. XCON sistemai sukaupus apie 10 tūkst. taisyklių, jos išlaikymas ir tobulinimas tapo labai sudėtingas ir brangus. Nepaisant to, tokių sistemų poreikis didėjo, ir jos buvo naudojamos vis plačiau. Jau 10-ajame dešimtmetyje buvo naudojamos tokios sistemos, kaip automatinė tvarkaraščių sudarymo programinė įranga projektų valdymui, mokymo įranga, dirbanti kaip korepetitorius su studentu, kalbos ir veido atpažinimo programos, automatinės dokumentų paieškos ir automatinių jų apibendrinimo programinė įranga. Atsirado nemažai ir finansinių bei vadybinių sprendimų, tokių, kaip automatinės nekilnojamojo turto įkeitimo sistemos, automatiniai investavimo sprendimų sudarytojai, programinė įranga kasdieninių bendrovės pajamų ir personalo poreikių planavimui, kreditinių sukčiavimų nustatymo sistemos (*credit fraud detection systems*), pagalbos vartotojui sistemos, leidžiančios pateikti teisingą atsakymą į bet kurį vartotojo klausimą greičiau, Interneto prekybos agentai (*shopping bots*), duomenų gavybos (*data mining*) įrankiai ir kt. Akivaizdu, kad tokio tipo sistemos gali būti labai naudingos žmogui, siekiančiam priimti svarbų sprendimą. Čia galima paminėti ir tai, kad Amerikoje kai kurios finansinės institucijos jau dabar linkusios pasikliauti ne tik žmogiškąja išmintimi, bet ir programine įranga, paremta dirbtinio intelekto metodais (neuroniniais tinklais, neraiškąja logika, saviorganizuojančiais tinklais ir kt.). Šiuo atveju gali būti priimami tokie sprendimai, kaip investavimas realiu laiku, akcijų kainų bei indeksų prognozavimas, vertybinių popierių pasirinkimas ir pan. Į tokių sprendimų grupę galima įtraukti ir nagrinėjamą sritį, kadangi tiek bankroto prognozavimą, tiek kredituojamų subjektų pasirinkimą galima priskirti prie tokio tipo uždavinių.

Kaip svarbų faktą reikia pabrėžti tai, kad ši sritis vis dar tirama ir plėtojama, todėl ekspertinės sistemos gali būti naudojamos tik kaip patarėnės vartotojui, jų sprendimu negalima visiškai pasitikėti. Tolimesnė šios srities plėtra ir plačiai atliekami jos metodų taikymo galimybių tyrimai leidžia tikėtis, kad ateityje bus sukurta sistemos, kurios galės pateikti optimalų sprendimą.

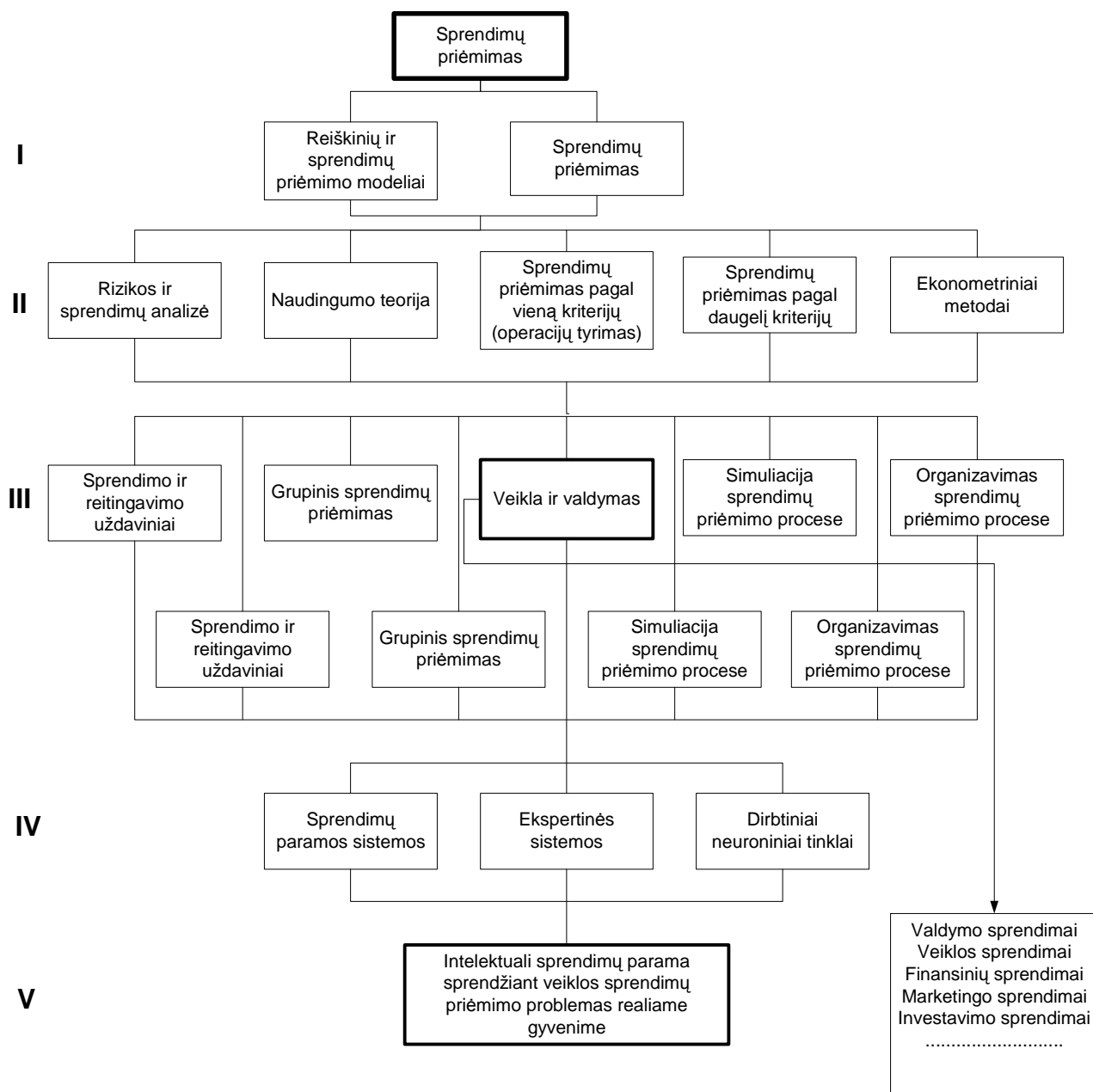
### **1.3.2. Ekspertinės sistemos ir jų vaidmuo sprendimų priėmimo procese**

Šio tipo sistemos gali būti apibrėžtos įvairiai, bet jų esmė yra ta pati – sprendimų priėmimas ir jų pasiūlymas vartotojui, tam naudojant tam tikrą žinių bazę. Čia pateikiami du jų apibrėžimai:

- *Ekspertinės sistemos* – tai tokios informacijos sistemos, kurios atsižvelgia į tam tikrus kriterijus ir visas galimas išvadas, renka informaciją apie kriterijus ir pasiūlo tinkamiausią tolimesnių veiksmų seką. Skirtingai nei žmogus, ekspertinės sistemos niekada nepamiršta svarbiausių detalių ir pateikia tikslesnius rezultatus (Abner);

- Tai programos, naudojančios žmogiškojo patyrimo žinių bazę, skirtos padėti problemų sprendimui. Problemų sprendimo laipsnis pagrįstas duomenų kokybiškumu bei žmogaus eksperto

suformuluotomis taisyklėmis. Ekspertinės taisyklės sukurtos veikti žmogaus eksperto lygmenyje  
(Free On-Line Dictionary of Computing)



**13 pav. Sprendimų priėmimas ir jį apimantys procesai**

Ekspertinių sistemų vaidmenį sprendimų priėmimo iliustruoja 13 pav., atspindintis tiek įvairių sričių vietą veiklos procesų modeliavime, tiek ir jų įtaką sprendimų priėmimo. Ši diagrama aprašo ir tai, kokių mokslo šakų, tiek matematikos, tiek ir kompiuterių mokslo, žinios reikalingos tokių sistemų modeliavimui bei kūrimui, bei nurodo sritis, kuriose tokio tipo sistemos gali būti taikomos ir naudojamos sprendimų priėmimui.

Čia aprašomos sistemos paprastai naudojamos kaip pagalbininkai sprendimų priėmimui. Tokio tipo sistemos paprastai vadinamos sprendimų paramos sistemomis (angl. *Decision Support Systems*, sutr. DSS) arba sutrumpintai SPS.



Peter Keen išskiria tokius SPS tikslus (Hamilton, 2004):

1. Padėti vadovaujamojo lygmens darbuotojui sprendžiant nestruktūrizuotas arba pusiau struktūrizuotas problemas (tačiau ne pakeisti patį darbuotoją). Tokio tipo sistemos leidžia išbandyti įvairius sprendžiamos problemos sprendimo variantus ir strategijas, išbandyti alternatyvas be jokių pasekmių modeliavimo lygmenyje, kuriant įvairias sprendimų pasekmes prognozuojančias aplinkynes (jautrumo analizė „kas būtų, jei“, tikslų paieškos metodas)

2. Prisidėti prie sprendimo efektyvumo, tačiau ne produktyvumo, taip padidinant teigiamo galimo rezultato galimybę. Šiuo atveju išskirta koncentruota informacija tokį sprendimą leidžia priimti greičiau, objektyviau ir tiksliau.

Svarbu pabrėžti tai, kad intelektualių metodų taikymas ir praėjusių sprendimų bei jų pasekmių sekimas apmoko sprendimų priėmimo sistemą tolesnių sprendimų priėmimui, kas leidžia panašių klaidų išvengti ateityje.

Gali būti išskirti tokie SPS tipai pagal jų paskirtį (Powers, 1997; cituojama pagal Merkevičius, Garšva, Cepkovataja, 2005):

- **Komunikacijos valdymo (grupės valdymo) SPS** (angl. *communication-driven, group-driven*) skirtos daugiau nei vieno žmogaus tikslams, užduotims suderinti, komunikuoti tarpusavyje, koordinuoti atskirų grupių veiklą.

- **Duomenų SPS** (angl. *data-driven*) realizuoja priėjimą ir panaudojimą prie istorinių arba laiko eilučių duomenų organizacijos vidiniuose arba išoriniuose resursuose.

- **Žinių valdymo SPS** (angl. *knowledge-driven*) - jos pateikia žinias apie tam tikrą sritį, padeda suvokti tos srities problemas. Čia integruojamos ekspertinės sistemos.

- **Dokumentų SPS** (angl. *document-driven*) orientuotos į dokumentų panaudojimą veiklos sprendimų priėmime.

- **Modelių valdymo SPS** (angl. *model-driven*) akcentuoja priėjimą prie modelio ir jo valdymą. Paprastos statistinės ir analitinės priemonės pateikia paprasčiausio lygio funkcionalumą. Ap-skritai modelio valdymo DSS sprendimų priėmimui naudoja kompleksinius - finansinius, modelia-vimo ir/arba optimizavimo - modelius. Modelio valdymo DSS naudoja sprendimų priėmėjų pateik-tus duomenis ir parametrus, pagal kuriuos analizuoja situaciją ir padeda priimti sprendimus.

### 1.3.3. Dirbtinio intelekto modelių finansinių sprendimų modeliavimui kūrimas

Dirbtiniu intelektu paremtų modelių kūrimui gali būti naudojama septynių žingsnių Kastrą ir Boyd metodologija (I. Kastrą, M.Boyd, 1995). Ši metodologija buvo sukurta ir aprašyta neuroni-niais tinklais pagrįstų modelių kūrimui, tačiau ją lengvai galima modifikuoti ir pritaikyti ir kitų pa-našaus tipo bei hibridinių mašininio mokymu paremtų modelių kūrimui.

**1. Kintamųjų parinkimas.** Pirmasis žingsnis yra labai svarbus, kadangi svarbu žinoti, kurie kintamieji yra svarbūs rinkoje. Finansų tyrinėtojai, suinteresuoti rinkos kainų prognozavimu, turi nuspręsti ar naudoti formalius ar fundamentalius įeigos duomenis iš vienos ar kelių rinkų. Formalūs įeigos duomenys apibrėžti kaip priklausomų kintamųjų ar indikatorių senesnės reikšmės, kurios buvo apskaičiuotos naudojantis šiek tiek senesniais duomenimis, o fundamentalūs įeigos duomenys - tai ekonominiai kintamieji, darantys įtaką priklausomiems kintamiesiems.

Kainų prognozavime dažniausia naudojami istoriniai, praeities duomenys apie rinkos kainas. Geresniam rezultatui gauti, naudojami taip vadinami tarpiniai rodikliai tarp rinkų, kurie sudaro ryšį tarp skirtingų rinkų: šalies ir tarptautinės. Pavyzdžiui, tarpinių rinkos rodiklių: Latvijos lato/jenos ir Lietuvos lito valiutos kurso ir palūkanų normos diferencialas gali būti naudojamas kaip neuroninių tinklų įvedimo duomenys prognozuojant lito kursą. Fundamentali informacija, tokia kaip einamasis sąskaitos balansas, pinigų pasiūla, ar didmeninės kainos indeksas, taip pat gali būti naudingi prognozavime.

Imamų duomenų dažnis priklauso nuo to, ką reikia prognozuoti. Jei paprastas vertybinių popierių prekybininkas modeliuotų neuroninį tinklą, jis tikriausiai naudotų kasdien gaunamus duomenis; investuotojas, kuris nori investuoti ilgesniam laikui, naudotų savaitinius arba mėnesio duomenis, taip tikėdamasis gauti geriausią vertybinių popierių rinkinį ir bandydamas išvengti pasyvios „Pirk ir valdyk“ strategijos. Ekonomistas, prognozuojantis BVP, nedarbingumo lygį ar kitus ekonominius indikatorius, naudosis mėnesio ar metų ketvirčio duomenimis. Kredito rizikos vertinime dažniausiai naudojami įvairūs skolininko finansiniai rodikliai, apimantys pusmečio ar metų rodiklius ar finansinių ataskaitų duomenis.

**2. Duomenų rinkimas.** Renkant duomenis kintamiesiems, aprašytiems praetame žingsnyje, būtina įvertinti duomenų prieinamumą ir gavimo kainą. Formalius duomenis yra lengviau gauti iš prekybininkų ar tarptautinių duomenų bazių už prieinamą kainą. Fundamentalius duomenis gauti yra šiek tiek sunkiau. Abiem atvejais turi būti pateikta kokybiška informacija, tačiau duomenys vis tiek turi būti dar kartą patikrinti.

Dažnai pasitaiko, kad tikrinant duomenis pastebimas tarpas tarp duomenų nuoseklumo, t.y. tam tikrų reikšmių nėra. Jei pastebima tokia klaida, trūkstamiems duomenims suteikiamos gretimų duomenų reikšmės. Yra ir kitas būdas išspręsti problemą – trūkstamiems duomenims priskirti vieną neuroną, kuriam suteikiama 1 reikšmė.

**3. Duomenų apdorojimas.** Tai įvedamų ir išvedamų kintamųjų analizė ir transformacija, siekiant sumažinti „triukšmą“, nustatyti svarbiausius ryšius, aptikti kryptingumus (tendencijas). Norint panaikinti tiesinę duomenų kryptį, naudojami diferenciacija ir kintamojo natūrinis logaritmavimas. Techninei analizei gali būti naudojami slenkamieji vidurkiai, generatoriai, nepastovumo filtrai.

Gerai rezultatai gaunami naudojant kelis metodus iš karto. Taip sumažinamas kintamųjų dubliavimasis, o neuroniniai tinklai geriau prisitaiko prie kintamų sąlygų, aišku, nuolatos mokydami. Dažnai yra taikomas įvedamų ir išvedamų duomenų glotninimas, naudojant paprastąjį arba eksponentinį slenkamųjų vidurkių metodą.

Duomenų atrinkimas ir filtravimas labai priklauso nuo to, koks yra prognozavimo arba, kredito rizikos vertinimo atveju, klasifikavimo tikslas. Tai galima iliustruoti prekių kainų prognozavimo pavyzdžiu. Šiuo atveju histograma su prekių kainų pasikeitimais parodys labai daug nežymių kainų pasikeitimų, iš kurių eilinis prekybininkas nelabai galės pasipelnyti, iš gautų pajamų atėmus prekių savikainą, tačiau tai stipriai įtakos kuriamų modelių apmokymą, kadangi menki kainų pasikeitimai vis tiek įtakoja duomenis, naudojamus apmokymams. Pašalinus sąlyginai nedidelius kainų skirtumus, būtų gaunami geresni rezultatai, kadangi modelio apmokymas būtų nukreiptas vien tik į didesnius, potencialiai pelningesnius kainų pasikeitimus. Aišku, įmanoma, kad sistema nepasiteisins net tada, kai prognozavimo tikslumas yra 85%, nes prognozės rezultatus gali sudaryti maži, nereikšmingi kainų pasikeitimai. Kitu atveju, smulkus prekybininkas gali būti suinteresuotas mažų kainų pasikeitimais, tačiau jis turi tvirtai žinoti ką modelis turi išmokti.

Reikia atkreipti dėmesį, kad duomenų apdorojimas reikalauja atlikti daug testavimų.

**4. Mokymasis, testavimas ir patvirtinimas.** Dirbtinio intelekto modelių apmokymus galima apibūdinti kaip jų sugebėjimą atpažinti duomenų struktūras. Testavimas yra naudojamas įvertinti tariamai apmokyto modelio apibendrinimo galimybes. Patvirtinimas – tai paskutinis sistemos įvertinimas, pripažinimas, kad sistema įvykdė uždavinius ir yra gauti norimi rezultatai.

Pats geriausias ir tiksliausias testavimo būdas yra „Ėjimas pirmyn“ (*walk-forward*), taip pat žinomas kaip slenkamųjų ar judančių langelių testavimas, atkartojantis realaus pasaulio prekybą ir testuojantis modelį, nuolat mokydamas modelį bei tam įvesdamas didelius kiekius duomenų (Kaastara, Boyd, 1995). Dažnas apmokymas leidžia greičiau adaptuotis prie nuolat besikeičiančių rinkos sąlygų.

**5. Įvertinimo (pripažinimo, angl. *evaluation*) kriterijai.** Dažniausiai naudojama paklaidų minimizavimo funkcija yra kvadratinių paklaidų suma. Taip pat gali būti naudojamos absoliučios deviacijos (nukrypimo), procentų skirtumų funkcijos. Tačiau šios funkcijos nėra paskutinis žingsnis įvertinant sistemos tinkamumą, kadangi kiti prognozavimo įvertinimo metodai, tokie kaip MAPE, paprastai nėra minimizuojami neuroniniuose tinkluose.

Kaip pavyzdį galima pateikti tą pačią prekių pardavimo sistemą. Joje sistemos prognozės yra paverčiamos į pirkti/parduoti signalus. Pavyzdžiui, visos prognozės didesnės nei 0.8 ar 0.9 gali būti traktuojamos kaip pirkimo signalai, o visos prognozės mažesnės nei 0.2 ar 0.1 – pardavimo signalai. Pirkimo ir pardavimo signalai įvedami į programą, kad ši suskaičiuotų pelno riziką. Tada

mažiausiai rizikingą pelną išrenka modeliuojančios sistemos.

**6. Sstemos apmokymas.** Sistemos apmokymas spręsti problemas, pvz., prognozuoti akcijų kainas, vyksta nuolatos apmokant tinklą. Sistemos mokymas yra panašus į vaiko mokymą atpažinti raides, pavyzdžiui, parodžius piešinėlį su raide „A“, vaikas klausiamas, kokia tai raidė. Jei atsakymas teisingas, pereinama prie kitos raidės, priešingu atveju vaikui pasakomas teisingas atsakymas. Toliau parodoma kita raidė ir taip procesas kartojamas. Šiuo atveju sistemai yra parodomi duomenys, tada ji spėja, koks turėtų būti rezultatas. Iš pradžių spėjimai ne visiškai atitinka tiesą, tačiau, kai modelis suklysta, jis pataisomas. Kai kitą kartą sistema pamato tuos pačius duomenis, jos spėjimų rezultatai bus kur kas tikslesni. Kaip ir žmogus, tokio tipo modelis apibendrina duomenis, taip sumažindama klydimų skaičių sekantį kartą parodžius kitokius duomenis. Tiesiog reikia nustatyti kokią informaciją suteikti sistemai, o ši atpažįsta struktūras, tendencijas ir paslėptus ryšius tarp elementų (Kaastra, Boyd, 1995).

#### **1.3.4. Intelektuali SPS kredito rizikos vertinimo procese**

Apžvelgus ekspertinių sistemų struktūrą, pritaikymo galimybes sprendimų priėmimo procese bei jose naudojamų modelių kūrimo pagrindinius principus, teliko nusakyti, kokį vaidmenį šios sistemos vaidina kredito rizikos procese.

Naudojant ekspertines sistemas šiame procese, gali būti išskiriami 6 etapai (Merkevičius, Garšva, Cepkovataja, 2005):

1) inicijavimo etape (klientui kreipiantis arba planinio kliento peržiūros įvykiui atsitikus) vadybininkas gauna užduotį atlikti kliento kredito rizikos įvertinimą;

2) informacijos paieškos etape vykdomas informacijos rinkimo procesas, kurį atlieka vadybininkas ir analitikas. Vadybininkas surenka pirminę informaciją iš kliento pateiktų duomenų, vidinių (jau esančių DB) resursų, gauna teisinį, rinkos, operacijų įvertinimą ir susistemines patalpina ją į MDDB bei pateikia analitikui. Analitikas atlieka platesnę paiešką išoriniuose informacijos šaltiniuose ir taip pat patalpina susistemintą informaciją MDDB.

Naudojantis ISPS atliekama kliento rizikos analizė. Kredito rizikos vertinimo procese analizuojami šie klientą liečiantys aspektai:

- *Įmonės esamos būklės įvertinimas* (darbo su klientu istorija, kliento pristatymas, pelno skirstymo politika, susijusios įmonės, susidaranti įmonių grupė, prekių/paslaugų įvertinimas, padėtis rinkoje, tiekėjai, savikainos aspektai, vadovavimo kokybė, trečiųjų šalių informacija);

- *Finansinė informacija* (apskaitos kokybė, sezoniškumo įtaka, pajamų/išlaidų struktūra, turto įvertinimas, išipareigojimų įvertinimas, mokumo, likvidumo rodikliai, kapitalo struktūra, finansiniai rodikliai, pinigų srautų analizė, buvusių prognozių atitikimas rinkos įvertinimas);

- *Prognozuojamos įmonės būklės įvertinimas* (biudžeto prognozė, investicijų poveikis, pasiskolos aptarnavimo galimybės, jautrumo analizė);

- *Rizikos mažinimo priemonės* (užstato (nekilnojamo turto) kokybės įvertinimas, užstato (įkeičiamų reikalavimo sutarčių) sąlygos, laidavimo kokybė).

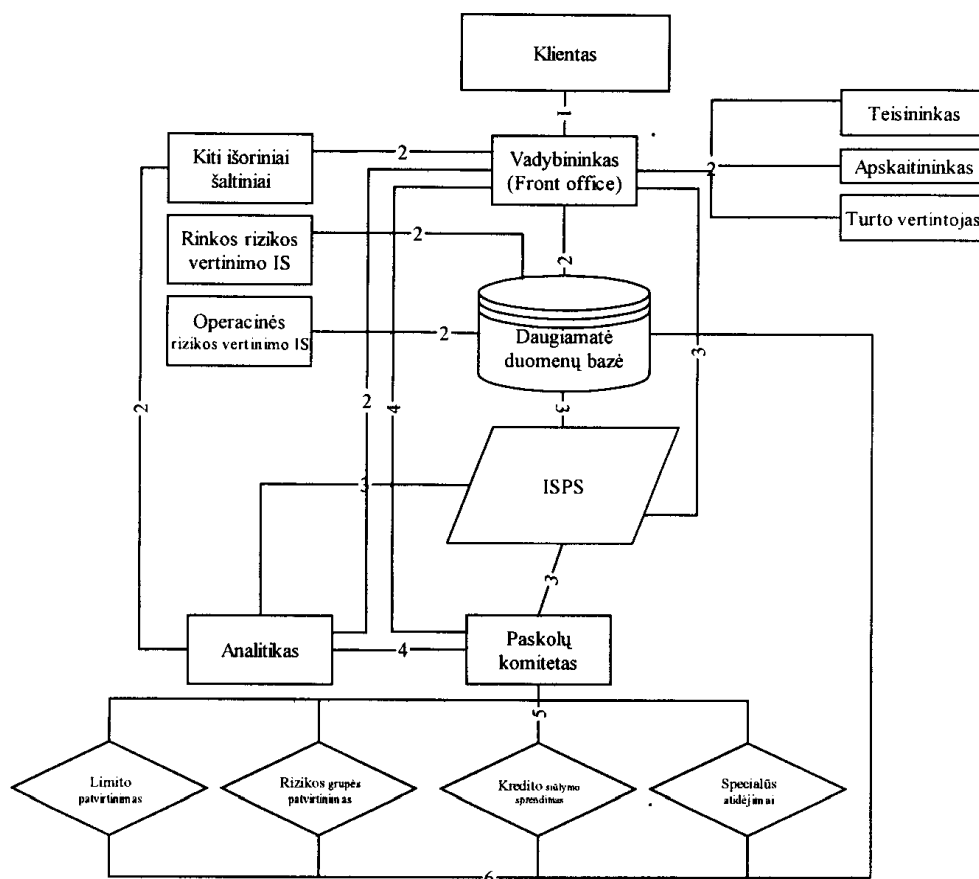
Analizės etape naudojamosi statistiniais, ekspertiniais ir dirbtinio intelekto metodais.

4) šiame etape vyksta kliento svarstymas ir ekspertinis vertinimas pagal vadybininko, analitiko ir ISPS pateiktas išvadas;

5) priimami sprendimai, kurių esminiai aspektai pateikti 14 pav.

6) grįžtamasis ryšys - dėl priimtų sprendimų atnaujinami duomenys MDDB.

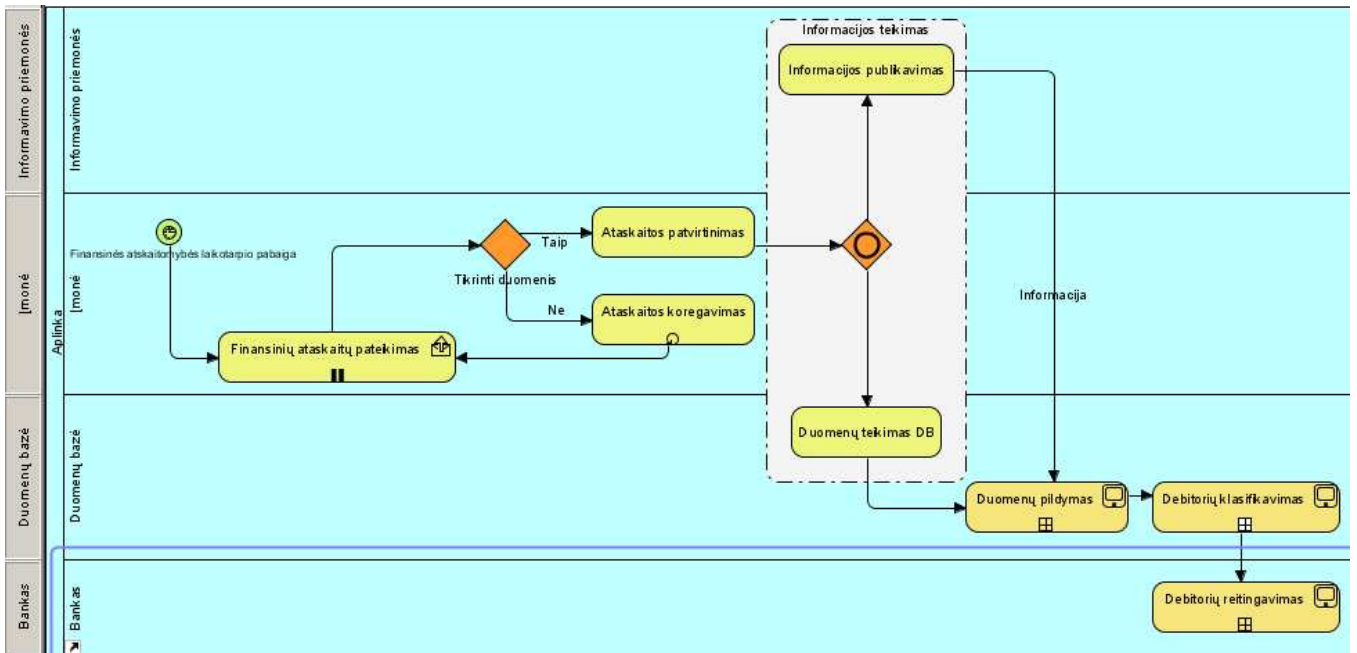
Visas kredito rizikos vertinimo procesas bei jo etapai pateikiami 14 pav.



Šaltinis: Merkevičius, E., Garšva, G., Cepkovataja, O. Intelektualios sprendimų paramos sistemos kredito rizikos vertinimui struktūra.

#### 14 pav. Intelektualios SPS vieta kreditų rizikos vertinimo procese

Pats informacijos pateikimo DSS procesas grafiškai pateikiamas 15 pav. Svarbu pabrėžti tai, kad ryšys tarp procesų “Duomenų pildymas” ir “Duomenų klasifikavimas” yra labiau sąlyginis, kadangi jie vyksta nebūtinai iškart vienas po kito.



Šaltinis: sudaryta autoriaus.

### 15 pav. Informacijos pateikimas DSS ir jos panaudojimas reitingavimo procese

#### 1.4. Atskirų DI metodų taikymo kredito rizikos vertinimui privalumai ir trūkumai

Išanalizavus pagrindinių dirbtinio intelekto sričių – neuroninių tinklų, genetinių algoritmų, *fuzzy* logikos, paramos vektorių mašinų – pritaikymo kredito rizikos analizei atliktus tyrimus, buvo nustatyti tokie faktoriai:

- Daugiausia tyrimų atlikta, taikant neuroninius tinklus, todėl galima daryti išvadą, kad ši DI sritis kredito rizikos tyrimui pritaikoma lengviausiai. Tą sąlygoja ir šio metodo lankstumas (pavyzdžiui, gana nesudėtinga klasifikuoti duomenis į keletą ar daugiau klasių) bei nesudėtingas matematinis pagrindas, leidžiantis jį paprastai modifikuoti, siekiant gauti tikslesnius rezultatus (ištyrimų akivaizdu, jog buvo taikoma daugybė neuroninių rinklų variacijų).

- Tiek evoliuciniai skaičiavimai, tiek *fuzzy* logika daugumoje tyrimų naudojama su kitais metodais, todėl dažniausiai šios sritys taikomos hibridiniuose modeliuose. Galima išskirti evoliucinių algoritmų našumą, sprendžiant įvairaus tipo optimizacinius skaičiavimus. Fuzzy logika gali būti panaudota ekspertinių žinių pritaikymui bei ekspertinio posistemio kūrimui.

- SVM metodas pradėtas taikyti neseniai, tačiau, remiantis atliktais tyrimais, galima teigti, jog jis yra vienas iš perspektyviausių ir teikiančių tiksliausius rezultatus. Atlikus jau atliktų šios srities tyrimų analizę, buvo nuspręsta orientuotis į atraminių vektorių mašinų metodo pritaikymą nagrinėjamos problemos sprendimui.

Pagrindiniai šių metodų panaudojimo privalumai ir trūkumai pateikiami 14 lentelėje.

14 lentelė. Dirbtinio intelekto metodų taikymo kredito rizikos vertinimui privalumai ir trūkumai

Metodas	Privalumai	Trūkumai
<b>Neuroniniai tinklai</b>	<ul style="list-style-type: none"> <li>tinkami darbui su nepilnais, trūkstamais ar neaptriukšmintais duomenimis;</li> <li>nereikalingi pirminiai skirstinio ar duomenų susiejimo įvertinimai;</li> <li>lankstumas - galimybė spręsti netiesines problemas, susieti bet kokią sudėtingą netiesiškumą bei aproksimuoti bet kokio sudėtingumo funkciją;</li> <li>atlikta daugybė tyrimų, kas reiškia, kad ši sritis pritaikoma lengvai;</li> <li>geresni klasifikavimo rezultatai ir našumas, lyginant su klasikineis metodais;</li> <li>plačios taikymo galimybės;</li> <li>galimybė spręsti problemas be žinomų sprendimų (tinklai, naudojantys mokymo be mokytojo principą, pvz.: Kohoneno tinklai);</li> <li>galimybė pritaikyti ekspertines žinias, naudojant fuzzy logiką</li> </ul>	<ul style="list-style-type: none"> <li>sudėtingumas bei neaiškumas (“juoda dėžė”) – nėra galimybės įvertinti tarpusavyje nepriklausomų kintamųjų svarbą bei sugeneruoti taisyklių rinkinį modelio veikimo aprašymui;</li> <li>“taisyklių ištraukimo” problema;</li> <li>sudėtingas modelio architektūros parinkimas;</li> <li>per didelis prisitaikymas ir tinklo persimokymo problema.</li> </ul>
<b>Evoliuciniai skaičiavimai</b>	<ul style="list-style-type: none"> <li>Lankstumas - galimybė spręsti netiesines problemas;</li> <li>Konceptualus paprastumas;</li> <li>Plačios taikymo galimybės;</li> <li>Geras našumas, lyginant su klasikineis metodais;</li> <li>Galimybė naudoti problemos žinias;</li> <li>Galimybė juos naudoti su kitais metodais;</li> <li>Lygiagretumas;</li> <li>Lankstumas dinaminiam pasikeitimams;</li> <li>Savybė pačiam optimizuotis;</li> <li>Galimybė spręsti problemas be žinomų sprendimų.</li> </ul>	<ul style="list-style-type: none"> <li>Retai taikomi atskirai; dažniausiai taikomi kartu su kitais dirbtinio intelekto metodais, kas reiškia papildomas laiko sąnaudas kuriant sistemą;</li> </ul>
<b>Fuzzy logika</b>	<ul style="list-style-type: none"> <li>Galimybė integruoti ekspertines žinias</li> </ul>	<ul style="list-style-type: none"> <li>Reikalingas taisyklių kūrimo, saugojimo ir vykdymo posistemis</li> </ul>
<b>Paramos vektorių mašinos (SVM)</b>	<ul style="list-style-type: none"> <li>Nėra lokalaus minimumo, kas yra daug efektyviau ir našiau bei nesudaro tam tikrų problemų, kaip tai atsitinka kuriant neuroninį tinklą;</li> <li>Optimalus ir platus sprendinių paskirstymas;</li> <li>Erdvės kontrolė, gaunama optimizuojant ribos parametą;</li> <li>Gerai našumo rezultatai;</li> <li>Galimybė pritaikyti ekspertines žinias, naudojant fuzzy logiką</li> </ul>	<ul style="list-style-type: none"> <li>Šiuo metu atlikta palyginti mažai tyrimų konkrečiai kredito rizikos ir bankroto galimybės vertinimo srityse;</li> <li>Sudėtinga taikyti klasifikuojant į daugiau nei dvi klases</li> </ul>

Šaltinis: sudaryta autoriaus.

Galima taip pat išskirti ir klasikinių modelių bei dirbtinio intelekto metodų privalumus ir trūkumus, juos lyginant tarpusavyje. Čia jie gali būti lyginami pagal tokius kriterijus, kaip galimas priklausomybės ir funkcijos tipas, lankstumas (galimybė modifikuoti, išplėsti ar naudoti su kitais modeliais), sudėtingumas, pritaikymo galimybės, atliktų tyrimų kiekis, vertinamų faktorių skaičius,

eksperto žinių panaudojimas ir konkrečių rodiklių naudojimas (15 lentelė). Ši analizė leidžia įvertinti dirbtinio intelekto metodų privalumus prieš klasikinius reitingavimo ir modernius kredito rizikos vertinimo modelius.

15 lentelė. Klasikinių ir DI metodų taikymo kredito rizikos vertinimui lyginamoji analizė

Metodas Kriterijus	Vertinimas balais	Modernūs modeliai	Neuroniniai tink- lai	Genetiniai algoritmai	SVM
Galimas pri- klausomybės (funkcijos) ti- pas	Tiesinė, netiesinė (logit)	Įvairi	Bet kokio sudėtingumo	Bet kokio sudėtingumo	Bet kokio sudėtingumo
Lankstumas	Yra galimybė modifikuoti, pritaikant konkrečiai sri- čiai ar šaliai	Galima išplėsti ar atlikti mo- difikacijas	Galima naudoti su kitais dirbtinio inte- lekto metodais (ge- netiniais algorit- mais, fuzzy logika ir kt.). Žinoma daug tipų	Galima naudoti su kitais dirbti- nio intelekto metodais (neu- roniniais tink- lais, fuzzy lo- gika ir kt.)	Galima naudoti su kitais dirbti- nio intelekto metodais. Ži- noma daug modifikacijų
Atliktų tyrimų kiekis	Didelis	Didelis	Didelis	Nedidelis <sup>5</sup>	Nedidelis
Vertinamų kri- terijų skaičius	Priklauso nuo konkreto mo- delio	Priklauso nuo kon- kreto mo- delio	Nuo 6 iki 40, gali būti ir daugiau	Neapribotas	Neapribotas
Eksperto žinių naudojimo ga- limybė	Galimas, išplė- tus modelį ir įvedus papil- domą kriterijų	Nėra	Yra (naudojant kartu su fuzzy logi- ka)	Yra (naudojant kartu su fuzzy logika)	Yra (naudojant kartu su fuzzy logika)
Konkrečių ro- diklių naudoji- mas	Taip	Taip	Ne	Ne	Ne

Šaltinis: sudaryta autoriaus.

Atsižvelgiant į tyrimo objektą, metodo efektyvumą bei ankstesnių bandymų pritaikyti skaičių, nuspręsta orientuotis į atraminių vektorių mašinų metodą ir jo taikymo galimybes.

### 1.5. Analitinės dalies išvados

1. Išskirti pagrindiniai ir plačiausiai naudojami dirbtinio intelekto metodai (neuroniniai tinklai, evoliuciniai skaičiavimai, fuzzy logika bei atraminių vektorių mašinos), apžvelgti jų ankstesni taikymai nagrinėjamai problemai bei jų privalumai ir trūkumai.
2. Aptartos pagrindinės kredito rizikos sąvokos bei vertinimo būdai. Svarbu atsižvelgti į tai, kad kredito rizikos vertinimas gali apimti ir kitų rizikų vertinimą. Šiame darbe tiriamas tik pačios kredito rizikos vertinimas.
3. Apžvelgti pagrindiniai klasikiniai kredito rizikos vertinimo metodai, tiek diskriminantiniai

<sup>5</sup> Daugiau tyrimų atlikta, kai šis metodas buvo naudojamas su kitais metodais



(Altman, Springate ir kt.), tiek ir modernūs, aptarti jų naudojimo privalumai ir trūkumai, įvertintos dirbtinio intelekto metodų (skirtų klasifikavimui bei klasterizavimui) panaudojimo su jais galimybės. Vėlesniuose tyrimuose nuspręsta apsiriboti diskriminantinių modelių naudojimu.

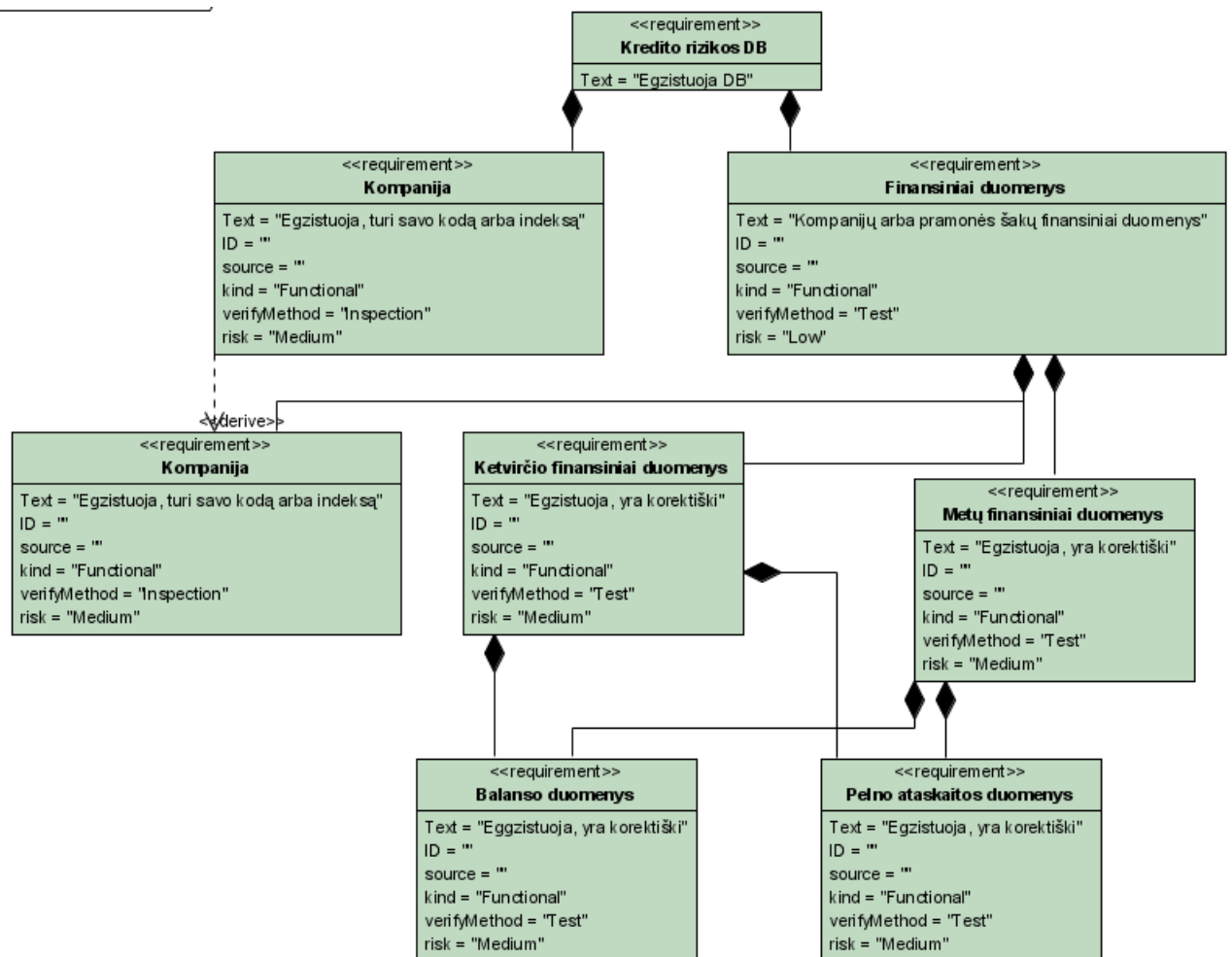
4. Aptartos ekspertinės sprendimų priėmimo sistemos, jų taikymo kredito rizikos vertinime bei sprendimo priėmimo procese prielaidos. Taip pat pateikta metodologija, skirta dirbtinio intelekto modelių finansinių sprendimų modeliavimui kūrimui, kuria bus remiamasi tolimesniame tyrimo, bei galima intelektualios sprendimų paramos sistemos kredito rizikos vertinimui struktūra ir informacijos šio tipo sistemai pateikimo ir veiksmų sekos diagrama.
5. Atlikus klasikinių ir DI metodų taikymo kredito rizikos vertinimui lyginamoji analizę, išryškinti dirbtinio intelekto metodų privalumai.
6. Vėlesniems tyrimams pasirinktas atraminių vektorių mašinų (SVM) metodas, motyvuojant pasirinkimą jo efektyvumu, pritaikymo galimybėmis bei palyginti nedideliu ankstesnių tyrimų skaičiumi.

## 2. SVM METODU PAREMTOS EKSPERIMENTINĖS SISTEMOS TYRIMAS

Šiame skyriuje aprašoma siūlomos sistemos koncepcinė schema, prototipas ir jo įvairūs veiklos aspektai, aptariami klausimai, susiję su jo realizacija bei jame naudojamomis technologijomis. Aptariama, kaip šią sistemą vėliau būtų galima išplėsti iki stambesnės, daugiau matematinių, statistinių, ekonometrinių ir dirbtinio intelekto metodų apimančios sistemos.

### 2.1.1. Kuriamos sistemos prototipo aprašymas

Kuriamos sistemos realizacijai pasirinkta JAVA programavimo kalba, leidžianti realizuoti jos aplinkoje įvairius specifinius modulius (tokius, kaip duomenų importavimo galimybė, duomenų įvedimas ir pan.) bei sąveiką su duomenų baze. Kaip dar vieną iš jos pasirinkimo kriterijų galima išskirti ir faktą, kad šia kalba realizuota daug įvairių dirbtinio intelekto ir mašininio mokymo bibliotekų bei algoritmų, kas leistų ateityje šią sistemą dar labiau plėsti bei tobulinti, jai pritaikant naujus algoritmus.



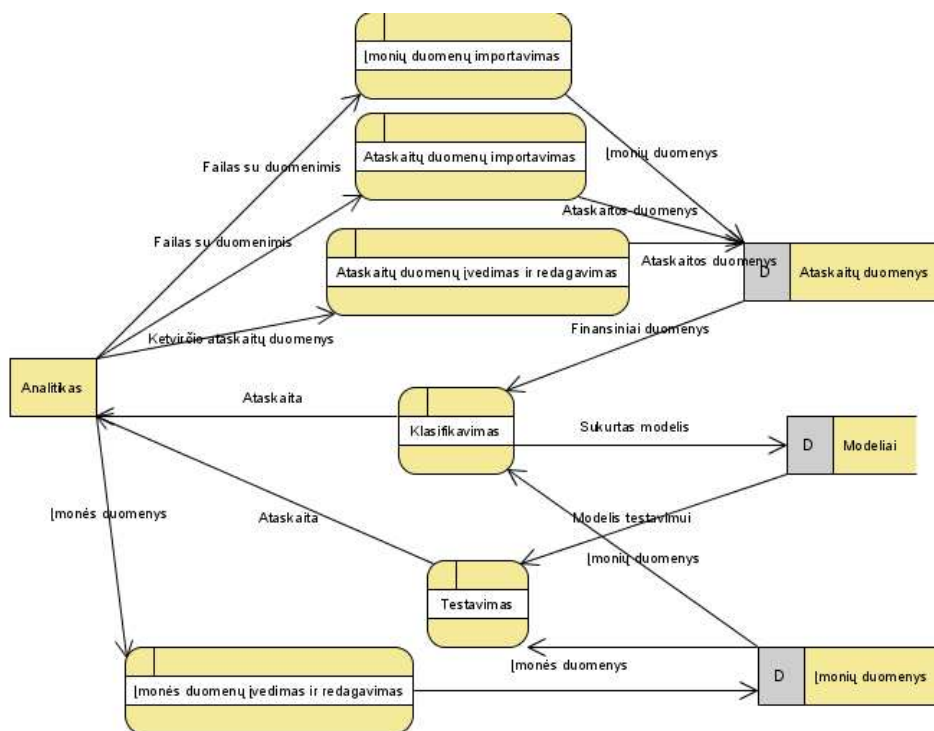
Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

16 pav. Sistemos reikalavimų diagrama

Duomenų saugojimui pasirinkta atvirojo kodo MySQL duomenų bazė<sup>6</sup>, motyvuojant tuo, kad ji leidžia greitai apdoroti didelius duomenų kiekius bei pasižymi dideliu patikimumu, o taip pat yra nevienplatformė (t.y., ją galima naudoti tiek Windows, tiek Linux, tiek ir kitose operacinėse sistemose).

Kuriama sistema naudoja EDGAR tarptautinės duomenų bazės 1999 – 2003 m. duomenis, kurie apima beveik 10000 tarptautinių įmonių įvairius finansinius duomenis; pagrindą sudaro metinių ir ketvirtinių balansų bei pelno ataskaitų duomenys. 6 priede pateikiamas ir šių duomenų sąrašas bei jų apibūdinimas.

Sistemos naudojamus duomenis ir jų sąryšį puikiai iliustruoja SysML<sup>7</sup> sistemos reikalavimų diagrama (16 pav.). Šioje diagramoje taip pat atspindimi ir reikalavimai šioms duomenims (korektiškumas, tikslumas, pilnumas, integralumas ir kt.), o taip pat ir galimas rizikos lygis. Informacijos srautai iliustruojami duomenų srautų diagrama (17 pav.). Šiuo atveju duomenų srautų nėra daug, tačiau kai kurie jų apima didelį parametrų arba informacijos kieki.



Šaltinis: sudaryta autoriaus.

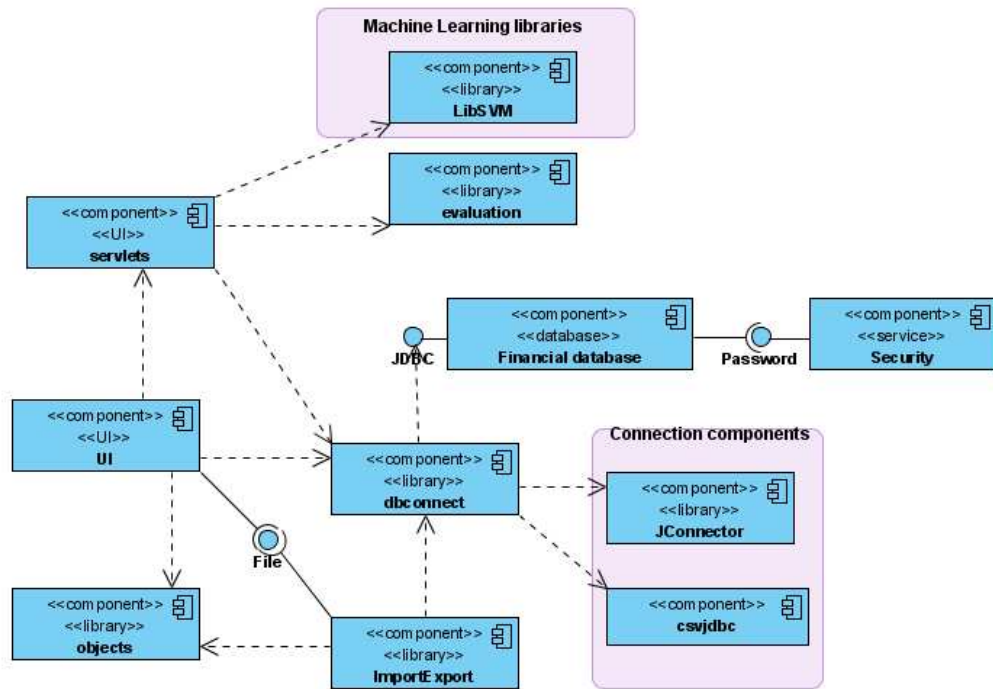
### 17 pav. Sistemos duomenų srautų diagrama

Architektūriniame lygmenyje sistemą galima aprašyti komponentų diagrama. Tai diagrama, skirta ne tik objektiniam, bet ir komponentiniam modeliavimui ir projektavimui. Ši diagrama leidžia

<sup>6</sup> Populiariausių DBVS privalumai ir trūkumai pagal pagrindinius jų vertinimo kriterijus pateikiami 7 priede

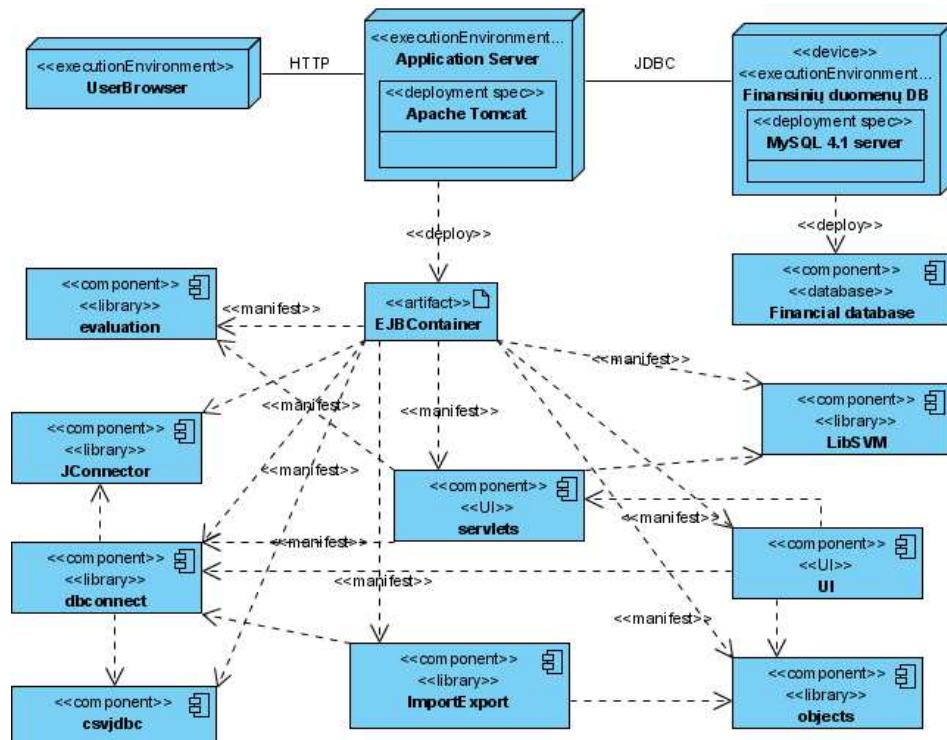
<sup>7</sup> OMG SysML v. 1.0 (Available Specification). Adresas Internete: <http://www.sysml.org/docs/specs/OMGSysML-v1.0-07-09-01.pdf>

apibrėžti tokius elementus, kaip komponentas, sąsaja (*interface*), jungtis (*port*) ir kt.



Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulio, skirto intelektualiu modeliu kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

18 pav. Sistemos komponentų diagrama

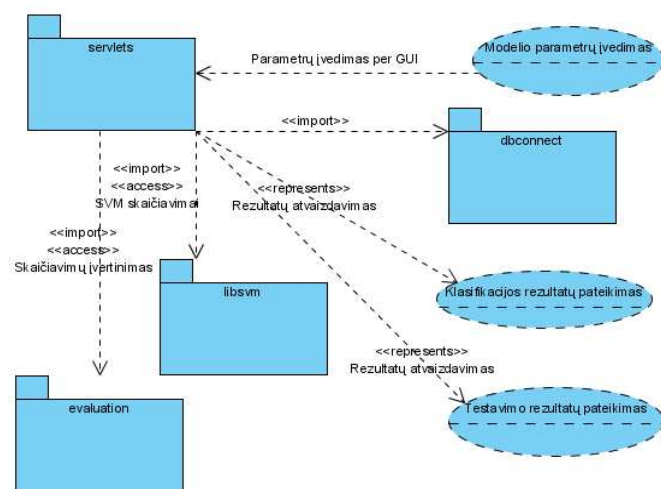


Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulio, skirto intelektualiu modeliu kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

19 pav. Sistemos diegimo diagrama

Sistemos paskirstymą tarp įvairių jos aplinkų gerai iliustruoja diegimo diagrama 19 pav. Galima išskirti dvi jos vykdymo aplinkas: aplikacijų serverį, kuriame įdiegtas visas kuriamos sistemos kodas, bei DBVS, kurioje saugomi visi naudojami duomenys ir modeliai. Šie du komponentai gali būti nebūtinai viename kompiuteryje, kas leidžia dar didesnę sistemos naudojamų resursų paskirstymą. Šiuo atveju pateikiama jau sukurtos sistemos diagrama; jos veikimo aplinkos ateityje gali keistis (pavyzdžiui, sistema gali būti pritaikyta kitoms DB ar kitiems Java aplikacijų serveriams, pavyzdžiui, Sun Java Application Server).

20 pav. pateikiama sistemos UML kompozicijos diagrama, aprašanti sistemos komponentų (modulių) bei uždavinių tarpusavio ryšius.



Šaltinis: sudaryta autoriaus

**20 pav. Modelio kūrimo kompozicijos diagrama**

16 lentelė. Sistemą sudarančių komponentų aprašymas

Komponento pavadinimas	Aprašymas
Financial database	Duomenų bazė, kurioje yra visi naudojami finansiniai duomenys
Security	Komponentas, kuris realizuoja sistemos saugumą
Dbconnect	Klasės, kuriose realizuotas prisijungimas prie MySQL duomenų bazės, užklausų apdorojimas ir vykdymas DB lygmenyje, o taip pat ir prisijungimas prie CSV failo kaip prie duomenų šaltinio
ImportExport	Paketas su klasėmis duomenų importavimui ir eksportavimui
Objects	Java Beans klasės įrašų duomenų redagavimui
UI	Abstraktus komponentas; tai labiau Interneto puslapių rinkinys, realizuojantis vartotojo sąsają
Servlets	Java klasės (servletai) kurios atlieka visas esmines sistemos operacijas – duomenų importavimą, eksportavimą, redagavimą, modelių kūrimą ir testavimą, jų efektyvumo vertinimą
Evaluation	Klasės klasifikavimo efektyvumo vertinimui pagal klaidų matricą (confusion matrix) bei jos parametrus
LibSVM*	Kiek pakeistas originalus LibSVM paketas, atliekantis visus SVM skaičiavimus
JConnector	Originalios bibliotekos prisijungimui prie MySQL duomenų bazės

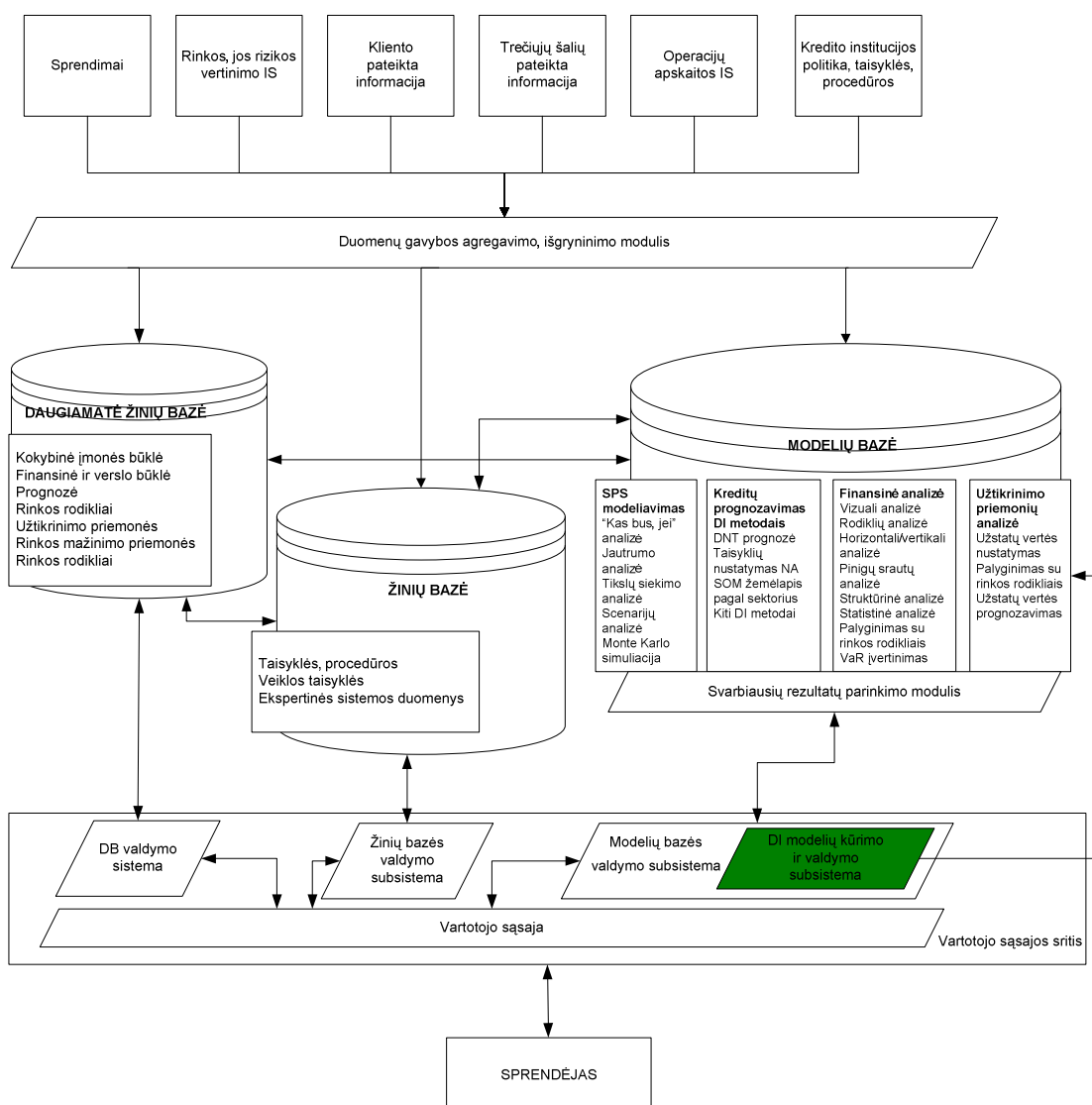
Komponento pavadinimas	Aprašymas
Csvjdbc	Biblioteka prisijungimui prie CSV failo kaip duomenų šaltinio (reikalinga duomenų importavimui)

Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

16 lentelėje pateikiamas sistemą sudarančių komponentų funkcijų aprašymas.

### 2.1.2. Kuriamos sistemos paskirtis ir jos panaudojimo galimybės

Kaip jau minėta skyriaus pradžioje, šios sistemos pagrindinis tikslas yra metodo, pagrįsto atramos vektorių mašinomis, realizavimas ir pritaikymas skolininkų (organizacijų ir pramonės šakų) vertinimui pagal jų pastarojo laikmečio finansinę informaciją, juo suskirstant į mokius ir nemokius.

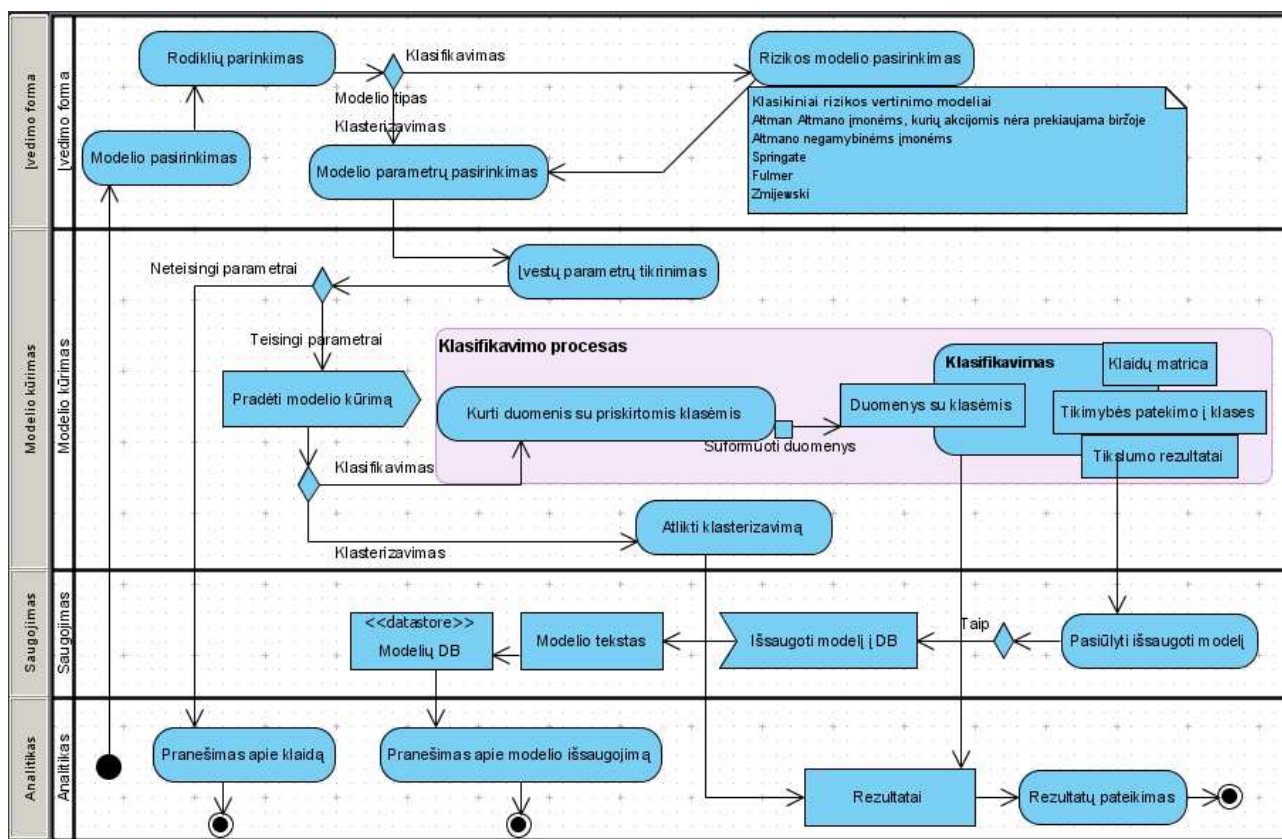


Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

21 pav. Galimos ISPS kredito rizikos vertinimui architektūra

Šią sistemą dar galima naudoti ir tam naudojamos informacijos kaupimui duomenų bazėje su tikslu ją panaudoti čia minimiems ar panašioms skaičiavimams. Ji gali būti taikoma ir kaip stambesnės sprendimų paramos sistemos dalis, kuri atlieka tam tikrą funkciją. Kaip tokios sistemos pavyzdį galima pateikti E.Merkevičiaus ir kt. siūlomą intelektualios sprendimų sistemos kredito rizikos vertinimui pavyzdį (Merkevičius, Garšva, Cepkovataja, 2005) (21 pav.). Šiuo atveju siūlomas sprendimas būtų integruojamas kaip modelių bazės subsistemos dalis (paveikslėlyje pažymėta žaliai).

Sukurtos sistemos veikimą galima iliustruoti sekos diagrama (22 pav.). Ši diagrama apima visą veiksmų seką nuo modelio pasirinkimo iki jo išsaugojimo. Galima pastebėti, kad į šią diagramą įtrauktas klasterizavimas; tai irgi viena iš galimybių plėsti sistemą, įdiegiant ir modernius klasterizavimo metodus bei juos pritaikant kredito rizikos vertinimo sričiai. Svarbu atsižvelgti į tai, kad šiuo atveju nebus sukuriamas ir išsaugojamas joks modelis; klasterizavimo proceso metu duomenys tik suskirstomi į klasterius pagal ryšius tarp jų.

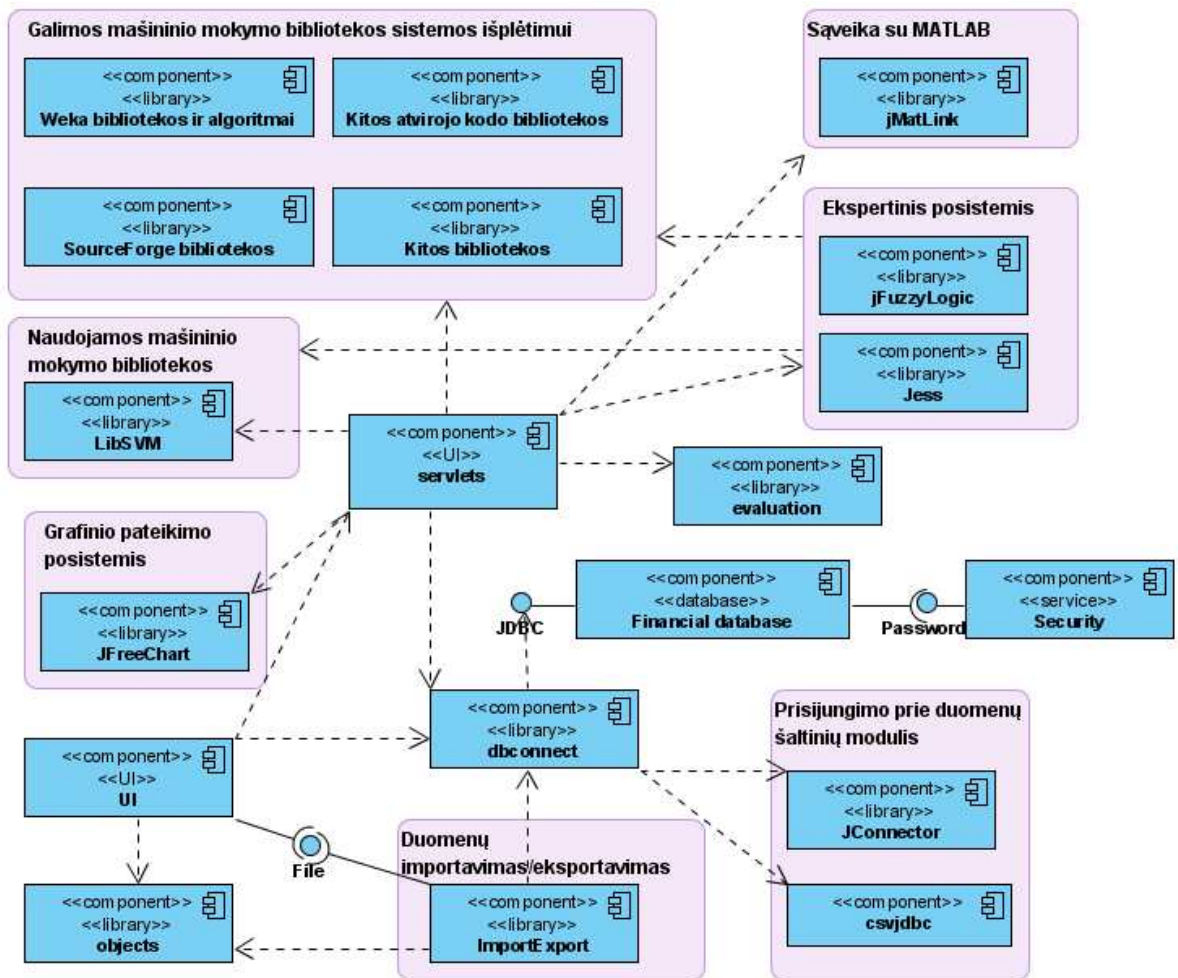


Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

22 pav. Galima modelio kūrimo veiksmų sekos diagrama

Sistema ateityje gali būti ir plečiama, integruojant naujus modelius bei metodus (22 pav.). Tam gali būti panaudotos įvairios bibliotekos bei programiniai komponentai, realizuojantys vieną ar

daugiau mašininio mokymo algoritmų. Kaip to pavyzdį galima pateikti Naujosios Zelandijos universitete sukurtą atvirojo kodo Weka mašininio mokymo sistemą, kurioje realizuoti beveik visi populiariausi modernūs mašininio mokymo algoritmai, tokie, kaip Bajeso tinklai, neuroniniai tinklai, sprendimų medžio ir kt. metodai. Integravus šiuos algoritmus, sistema gali būti išplėsta iki sistemos, kuri galėtų netgi palyginti dviejų ar daugiau skirtingų algoritmų rezultatus (Danėnas, Merkevičius, Garšva, 2008).



Šaltinis: Danėnas P., Merkevičius E., Garšva G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra.

### 23 pav. Išplėstinės sistemos komponentų diagrama

Lygiai taip pat galimas ir kitų atvirojo kodo bibliotekų integravimas (Danėnas, Merkevičius, Garšva, 2008):

- daug mašininio kodo algoritmų realizacijų Java kalba galima rasti SourceForge portale, plačiau atvirojo kodo programas bei bibliotekas, be to, daug bibliotekų ir realizacijų Java ir kitomis programavimo kalbomis galima rasti ir kituose Interneto tinklalapiuose;
- yra ir galimybė naudoti MATLAB skaičiavimų variklį, tam naudojant naudojant JMatLink biblioteką – galimas sistemos komunikavimas su MATLAB sistema bei tikėtina (bet dar netirta)



galimybė modelių rezultatus išsaugoti sistemos DB;

- galimybė integruoti kitos atvirojo kodo sistemų algoritmus, juos adaptuojant sprendžiamai problemai. Iš jų galima išskirti kompanijos Rapid-I YALE<sup>8</sup> duomenų gavybos sistemą bei Java duomenų gavybos standartą ir jo realizaciją JDM<sup>9</sup>;
- integravus ekspertinį posistemį (naudojant jFuzzyLogic<sup>10</sup>, Jess<sup>11</sup> ar kitas JAVA ekspertinių sistemų kūrimo priemones), galima sukurti ir galimybę klasifikavimo ir klasterizavimo procesuose naudoti ekspertines žinias bei kurti ir neraiškių aibių modelius.

### 2.1.3. Sistemoje naudojami metodai

Sistemoje naudojamas viena iš populiariausių ir plačiausiai taikomų duomenų gavybos ir statistikos metodų – klasifikavimas, kurio pagrindinė užduotis yra suskirstyti atributų vertes pagal galimas klases. Šis metodas svarbus įvertinant tiek dabartines, tiek ir prognozuojamas vertes; prognozavimas gali būti traktuojamas kaip atributo vertės klasifikavimas į vieną iš galimų klasių (Dunham, 2003).

Pati klasifikavimo problema formaliai apibrėžiama taip (Dunham, 2003): turint duomenų bazę su  $D = \{t_1, t_2, \dots, t_n\}$  įrašų (vienetų) ir klasių aibę  $C = \{C_1, C_2, \dots, C_n\}$ , klasifikavimo problema yra susiejimo  $f : D \rightarrow C$ , kurioje kiekvienas  $t_i$  yra priskirtas kuriai nors klasei. Klasėje  $C_j$  yra tiksliai tik tie įrašai, kurie su ja susieti, t.y.,  $C_j = \{t_i \mid f(t_i) = C_j, 1 \leq i \leq n, t_i \in D\}$ .

Šios problemos sprendimui gali būti naudojami trys pagrindiniai metodai (Dunham, 2003):

1. **Ribų apibrėžimas** – klasifikacija atliekama padalijant galimų įeigos įrašų erdvę į regionus, kur kiekvienas regionas susietas su viena klase;
2. **Tikimybių pasiskirstymo naudojimas** – kiekvienai duotai klasei  $C_j$ ,  $P(t_i / C_j)$  yra tikimybių pasiskirstymo funkcija (*probability distribution function*, sutr. PDF) klasei, vertinamai taške  $t_i$ ;
3. **Vėlesnių (*posterior*) tikimybių panaudojimas** – turint duomenų reikšmę  $t_i$ , reikia apibrėžti tikimybę, kad  $t_i$  yra klasėje  $C_j$ . Tai žymima  $P(C_j/t_i)$  ir vadinama vėlesne tikimybe. Vienas iš tokios klasifikacijos požiūrių būtų apibrėžti vėlesnę tikimybę kiekvienai klasei ir tada priskirti  $t_i$  klasei su didžiausia tikimybe.

Tiek neuroniniai tinklai ar SVM, tiek ir kiti mašininį mokymąsi naudojantys metodai priskiriami trečiajam požiūriui, todėl juo remiamasi ir šiame darbe.

<sup>8</sup> Rapid-I [interaktyvus]. Adresas Internete: <http://rapid-i.com/content/blogcategory/10/69/lang/en/>.

<sup>9</sup> JSR-000073 Data Mining API (Final Release). Adresas Internete:

<http://jcp.org/aboutJava/communityprocess/final/jsr073/index.html>

<sup>10</sup> jFuzzyLogic: Open Source Fuzzy Logic (Java). Adresas Internete: <http://jfuzzylogic.sourceforge.net/html/index.html>

<sup>11</sup> Jess, the Rule Engine for the Java Platform[interaktyvus]. Adresas Internete: <http://herzberg.ca.sandia.gov/>

### 3. EKSPERIMENTO IR SISTEMOS REALIZACIJOS APRAŠYMAS

Šioje dalyje aprašomas metodas, sukurtas naudojant dirbtinio intelekto sritis bei metodikas (atramos vektorių mašinos), jo veikimas, empirinis taikymas. Taip pat pateikiami rezultatai, gauti pagal šį metodą, bei galimybės pasiekti dar geresnius rezultatus.

#### 3.1. Tyrimo metodikos aprašymas

Siūlomas sprendimas apimtų klasikinių diskriminantinių modelių ir vieno iš mašininio mokymo metodų (šiuo atveju atraminių vektorių mašinų) integravimą. Šiuo atveju diskriminantinis modelis būtų naudojamas suskaičiuoti išėigos reikšmėms, t.y., įvertinant įmonės rizikingumą, priskirti ją prie vienos iš dviejų klasių (rizikinga ar nerizikinga). SVM metodas naudojamas modelio gavimui, t.y., apskaičiuojami jo koeficientai, kuriuos naudojant, būtų gaunama duomenų be išėigos reikšmių tiksliausia klasifikacija. Šios problemos sprendimas apima tik binarinį klasifikavimą, t.y., bendrovės pagal savo finansinius duomenis suskirstomos tik į dvi klases – „rizikinga“ ir „nerizikinga“ (SVM leidžia suskirstyti ir į daugiau klasių, tačiau toks klasifikavimas yra daug sudėtingesnis procesas, todėl šiame darbe jis nenagrinėjamas). Kiti tyrimo žingsniai apimtų šiam tinkamiausių diskriminantinių modelių nustatymą bei geriausių SVM parametrų parinkimą.

Skaičiuojant išėigos reikšmes, diskriminantinis modelis pritaikomas tik „rizikinga“ ir „nerizikinga“ klasėms, t.y., jo rezultatas gali būti tik viena iš šių reikšmių. Tai ypač svarbu Altmano modelių atveju; jo atveju rezultatas, rodantis neapibrėžtumo zoną, laikomas „nerizikinga“.

Nustatant tinkamiausius diskriminantinius modelius, atliekamas apmokymas, naudojant visų naudojamų rodiklių (iš viso 56) standartizuotus duomenis su sumažintu jų masteliu (t.y., jų reikšmės turi būti intervale [-1;1]) bei standartinėmis SVM parametrų reikšmėmis. Šiuo atveju kuriami modeliai, naudojant kiekvieną iš taikomų diskriminantinių modelių (Altman, Springate, Zmijewski) bei vieną iš SVM branduolio funkcijų (tiesinė, polinominė, RBF, sigmoidinė); vėliau su geriausiaisiais variantais bandoma parinkti geriausius parametrus. Šiuo atveju naudojamas indukcinis principas, t.y., laikoma, kad modelis, su standartinėmis SVM reikšmėmis parodęs geriausius rezultatus, yra geriausias ir ieškoma būdų, kaip su geriausia kombinacija pasiekti dar geresnių rezultatų.

Skyriaus pabaigoje aprašoma sistema, kurią naudojant, atliktas šis eksperimentas, jos realizacijos aspektai, galimybės bei įdiegimas.

#### 3.2. Eksperimentas ir jo rezultatas

Siekiant įsitikinti šiame darbe atliktų analizės teisingumu, buvo atliktas tyrimas, kurio tikslas – išsiaiškinti SVM metodo tinkamumą kredito rizikos vertinimui. Buvo pasirinkta v-SVC realizacija (LibSVM programinis paketas) ir, integravus su duomenų baze bei kai kuriais diskriminanti-

niais modeliais, sukurtas prototipas pagal pirmuose šio skyriaus poskyriuose aprašomą architektūrinį karkasą. Atliekamam eksperimentui pasirinktas SVM modelio hibridinimas su trimis pagrindiniais ir populiariausiais šiuo metu naudojamais diskriminantiniais modeliais: Altman, Zmijewski bei Springate. Pats eksperimentas susideda iš dviejų žingsnių – pirmuoju siekiama nustatyti, kuris iš šių modelių ir su kuria branduolio funkcija duoda geriausius rezultatus klasifikuojant pateikiamus atvejus, o antruoju siekiama nustatyti geriausius (bent jau apytikriai) v-SVC parametrus  $\text{coef0}$ ,  $\nu$ , branduolio funkcijos laipsnį (jei funkcija netiesinė) ir kt. Eksperimentui naudojami visi galimi rodikliai (iš viso 56; jų sąrašas pateikiamas 7 priede).

Pradiniam testavimui naudojami standartiniai v-SVC parametrai, kuriuos rekomenduoja LibSVM kūrėjai (17 lentelė).

17 lentelė. v-SVM metodo parametrai ir aprašymai

Parametras	Aprašymas	Pagal nutylėjimą
<b>v-SVC parametrai</b>		
<b>Branduolio funkcijos tipas</b>	Branduolio funkcijos tipas (tiesinis, polinominis, RBF, sigmoidinis ar vartotojo aprašytas (precomputed type)). Testavimas atliekamas su pirmaisiais keturiais	Tiesinis
<b>Kernelio funkcijos laipsnis</b>	Jeigu branduolio funkcija polinominė, šis parametras apibrėžia jos laipsnį	3
<b>coef0</b>	coef0 parametras branduolio funkcijoje	0
<b><math>\gamma</math> parametras</b>	$\gamma$ parametras branduolio funkcijoje	1
<b><math>\nu</math> (nu) parametras</b>	$\nu$ parametras	0,05
<b>Toleravimo parametras</b>	Užbaigimo kriterijaus tolerancijos parametras	0,001
<b>Mažėjimo euristicos naudojimas</b>	Nusako, ar naudoti mažėjimo euristicą	Nustatyta
<b>SVC ar SVR apmokymas tikimybių skaičiavimui</b>	Nusako, ar kuriamą modelį apmokyti tikimybių įvertinimui. Jį nustačius, testavimo metu bus apskaičiuotos priskyrimo tam tikrai klasei tikimybės	Nustatyta
<b>Kryžminis validavimas</b>	Jį nustačius, sukuriama modelį, jis bus ištestuotas naudojant kryžminį validavimą	Nenustatyta
<b>Kryžminio validavimo žingsnių skaičius</b>	Nustatoma kryžminio validavimo žingsnių skaičius, t.y., į kiekvieni bus suskirstytas apmokymo duomenų rinkinys tam, kad su kiekviena iš jų būtų atliekamas apmokymas, o su likusiais duomenimis – testavimas	5
<b>Kiti parametrai</b>		
<b>Standartizuoti duomenys</b>	Nustatoma, ar atlikti duomenų standartizavimą	Nustatyta
<b>Atlikti duomenų mastelio sumažinimą</b>	Duomenys „sumažinami“, t.y., padalinami iš tam tikros reikšmės taip, visi duomenys patektų į intervalą [-1;1]	Nustatyta
<b>Duomenų mastelio sumažinimo metodas</b>	Galimi du mastelio sumažinimo metodai, t.y., parametras, iš kurio dalijami visi duomenys, radimo būdai: <ul style="list-style-type: none"> <li>• Maksimali absoliuti reikšmė;</li> <li>• Skirtumas tarp minimalios ir maksimalios reikšmės</li> </ul>	Maksimali absoliuti reikšmė
<b>Vertinimo metodas klasių kūrimui</b>	Vienas iš diskriminantinių modelių, kurio pagrindu kuriamos išėjimo reikšmės: <ul style="list-style-type: none"> <li>• Altmano;</li> </ul>	Altmano

Parametras	Aprašymas	Pagal nutylėjimą
	<ul style="list-style-type: none"> <li>• Altmano įmonėms, neprekiaujančiomis akcijomis biržoje</li> <li>• Altmano negamybinėms įmonėms</li> <li>• Springate</li> <li>• Fulman</li> <li>• Zmijewski</li> </ul>	

Šaltinis: sudaryta autoriaus.

Antroje eksperimento dalyje bus bandoma keisti standartinius parametrus ir rasti tinkamiausius. Duomenys standartizuojami, naudojant klasikinę standartizavimo procedūrą, bei sumažinamas jų mastelis pagal absoliučią reikšmę.

### 3.2.1.1. Įvertinimo metodikos aprašymas

Įvertinimui naudojama standartinė įvertinimo metodika, dar žinoma kaip maišos matrica (angl. confusion matrix), bei pagrindiniai rodikliai, apskaičiuojami iš jos verčių.

Visų pirma, maišos matrica apibrėžiama<sup>12</sup> kaip matrica, turinti tokias reikšmes:

- a yra **teisingų** spėjimų skaičius, kad atvejis yra **neigiamas**;
- b yra **neteisingų** spėjimų skaičius, kad atvejis yra **teigiamas**;
- c yra **neteisingų** spėjimų skaičius kad atvejis yra **neigiamas**;
- d yra **teisingų** spėjimų skaičius, kad atvejis yra **teigiamas**.

		Progozė	
		Neigiamas	Teigiamas
Sistema	Neigiamas	a	b
	Teigiamas	c	d

Iš šios jos gaunami parametrai, kurie gali būti naudojami tikslumo ir efektyvumo vertinimui:

- Teisingumas (accuracy) AC yra teisingų spėjimų skaičiaus ir bendro spėjimų skaičiaus proporcija. Ji apskaičiuojama taip:  $AC = \frac{a+d}{a+b+c+d}$

- *Recall* arba teisingų teigiamų rodiklis (true positive rate) arba TP yra teigiamų atvejų, kurie buvo teisingai identifikuoti, proporcija, apskaičiuojama taip:  $TP = \frac{d}{c+d}$

- Neteisingų teigiamų rodiklis (the false positive rate) arba FP yra neigiamų atvejų, kurie buvo neteisingai suklasifikuoti kaip teigiami, proporcija, apskaičiuojama taip:  $FP = \frac{b}{a+b}$

- Teisingų neigiamų rodiklis (the true negative rate) arba TN apibrėžiamas kaip neigiamų atvejų, kurie buvo teisingai suklasifikuoti, proporcija, apskaičiuojama taip:  $TN = \frac{a}{a+b}$

- Neteisingų neigiamų rodiklis (the false negative rate) arba FN apibrėžiamas kaip teigiamų atvejų, kurie buvo neteisingai suklasifikuoti kaip neigiami, proporcija:  $FN = \frac{c}{c+d}$

<sup>12</sup> Maišos matrica gali būti apibrėžta bet kuriam klasių skaičiui; čia ji apibrėžiama tik sprendžiamos problemos atveju (binariniam kalsifikavimui)

- Tikslumas (precision) P yra prognozuojamų teigiamų teisingų atvejų propocija, apskaičiuojama taip:  $P = \frac{d}{b+d}$
- Teisingumo įvertis gali būti neadekvatus našumo matavimas, kai neigiamų atvejų skaičius yra daug didesnis nei teigiamų atvejų skaičius. Šiuo atveju apbrėžiamas geometrinis vidurkis (g-mean) ir F-Measure rodiklis:

$$g - mean_1 = \sqrt{TP * P} ; g - mean_2 = \sqrt{TP * TN}$$

- $F_1$  rodiklis (F-Measure) yra harmoninis tikslumo ir recall įverčių vidurkis:

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

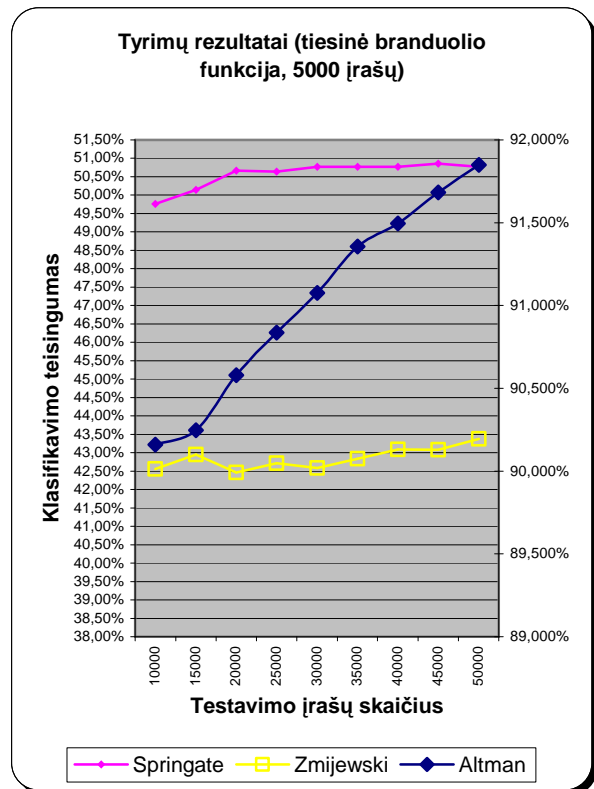
Šiame darbe naudojami tik tesingumo, tikslumo ir *recall* parametrai, tačiau realizacijoje taip pat skaičiuojami ir visa maišos matrica bei kiti rodikliai.

### 3.2.1.2. Tinkamiausio diskriminantinio modelio bei branduolio funkcijos nustatymas

Modelio kūrimui pasirenkame pirmus 5000 įrašų, testavimas atliekamas palaipsniui imant duomenis nuo 10000 iki 50000 atvejų, palaipsniui didinant skaičių 5000 vienetu.

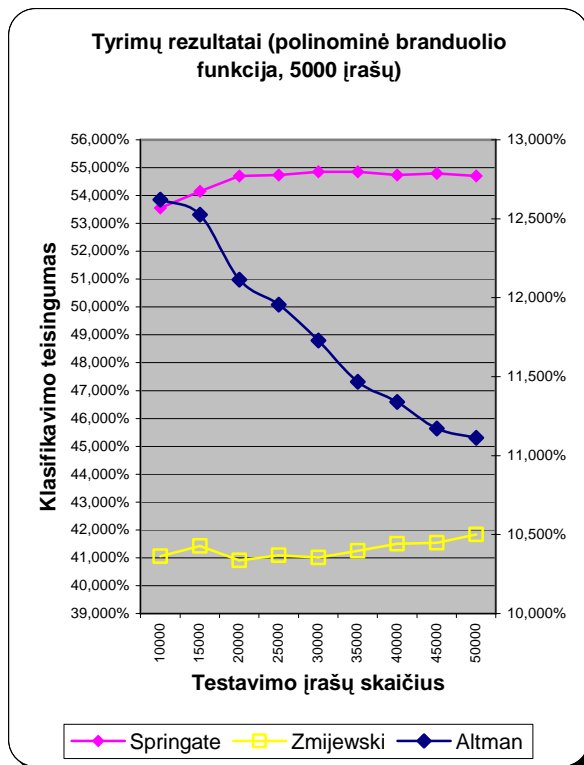
#### 1. Rezultatai, gauti, kai branduolio funkcija yra tiesinė:

Diskriminantinio modelio pavadinimas	Įrašų kiekis testavimui	Bendras teisingumas
Altman	10000	90,160%
Altman	15000	90,2467%
Altman	20000	90,58%
Altman	25000	90,836%
Altman	30000	91,0767%
Altman	35000	91,357%
Altman	40000	91,495%
Altman	45000	91,682%
Altman	50000	91,850%
Springate	10000	49,76%
Springate	15000	50,15%
Springate	20000	50,665%
Springate	25000	50,64%
Springate	30000	50,767%
Springate	35000	50,771%
Springate	40000	50,765%
Springate	45000	50,86%
Springate	50000	50,766%
Zmijewski	10000	42,56%
Zmijewski	15000	42,9533%
Zmijewski	20000	42,465%
Zmijewski	25000	42,712%
Zmijewski	30000	42,583%
Zmijewski	35000	42,834%
Zmijewski	40000	43,09%
Zmijewski	45000	43,08%
Zmijewski	50000	43,376%



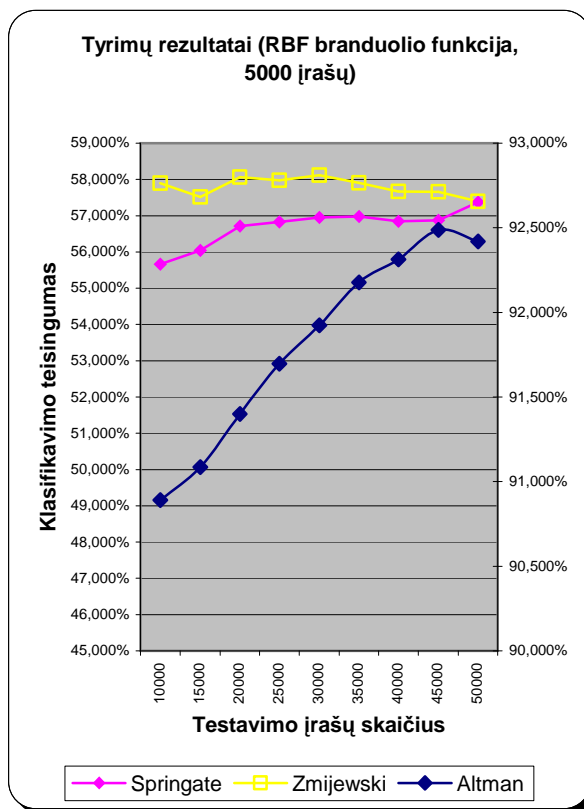
## 2. Rezultatai, gauti, kai branduolio funkcija yra polinominė su laipsniu, lygiu 3:

Diskriminantinio modelio pavadinimas	Įrašų kiekis testavimui	Bendras teisingumas
Altman	10000	12,620%
Altman	15000	12,527%
Altman	20000	12,115%
Altman	25000	11,956%
Altman	30000	11,730%
Altman	35000	11,469%
Altman	40000	11,340%
Altman	45000	11,171%
Altman	50000	11,112%
Springate	10000	53,550%
Springate	15000	54,147%
Springate	20000	54,700%
Springate	25000	54,736%
Springate	30000	54,847%
Springate	35000	54,849%
Springate	40000	54,735%
Springate	45000	54,796%
Springate	50000	54,706%
Zmijewski	10000	41,060%
Zmijewski	15000	41,420%
Zmijewski	20000	40,905%
Zmijewski	25000	41,088%
Zmijewski	30000	41,007%
Zmijewski	35000	41,254%
Zmijewski	40000	41,503%
Zmijewski	45000	41,536%
Zmijewski	50000	41,840%



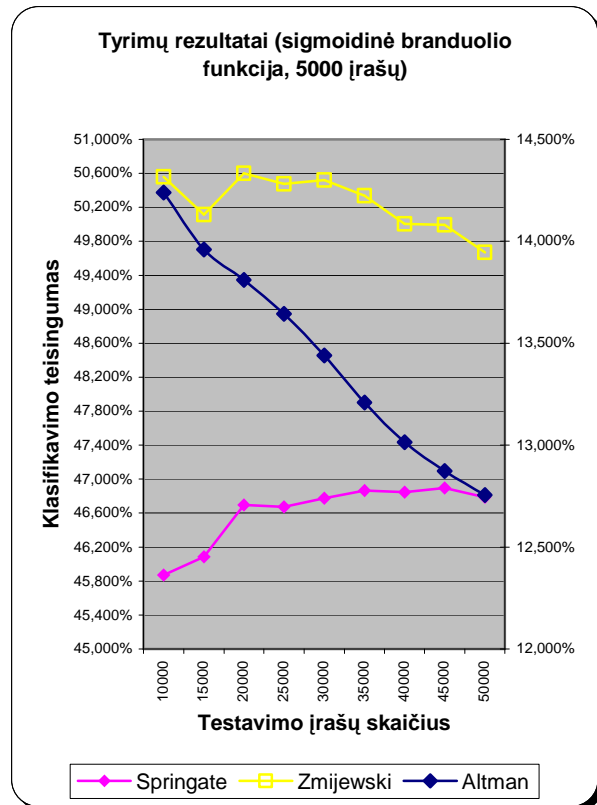
## 3. Rezultatai, gauti, kai branduolio funkcija yra radialinės bazės funkcija (RBF):

Diskriminantinio modelio pavadinimas	Įrašų kiekis testavimui	Bendras teisingumas
Altman	10000	90,890%
Altman	15000	91,087%
Altman	20000	91,400%
Altman	25000	91,696%
Altman	30000	91,923%
Altman	35000	92,177%
Altman	40000	92,313%
Altman	45000	92,487%
Altman	50000	92,420%
Springate	10000	55,660%
Springate	15000	56,040%
Springate	20000	56,710%
Springate	25000	56,832%
Springate	30000	56,950%
Springate	35000	56,977%
Springate	40000	56,842%
Springate	45000	56,873%
Springate	50000	57,380%
Zmijewski	10000	57,900%
Zmijewski	15000	57,513%
Zmijewski	20000	58,065%
Zmijewski	25000	57,976%
Zmijewski	30000	58,113%
Zmijewski	35000	57,903%
Zmijewski	40000	57,668%
Zmijewski	45000	57,656%
Zmijewski	50000	57,392%



#### 4. Rezultatai, gauti, kai branduolio funkcija yra sigmoidinė funkcija:

Diskriminantinio modelio pavadinimas	Įrašų kiekis testavimui	Bendras teisingumas
Altman	10000	14,240%
Altman	15000	13,960%
Altman	20000	13,810%
Altman	25000	13,644%
Altman	30000	13,440%
Altman	35000	13,209%
Altman	40000	13,015%
Altman	45000	12,873%
Altman	50000	12,756%
Springate	10000	45,870%
Springate	15000	46,087%
Springate	20000	46,695%
Springate	25000	46,672%
Springate	30000	46,777%
Springate	35000	46,866%
Springate	40000	46,848%
Springate	45000	46,896%
Springate	50000	46,786%
Zmijewski	10000	50,560%
Zmijewski	15000	50,113%
Zmijewski	20000	50,600%
Zmijewski	25000	50,476%
Zmijewski	30000	50,520%
Zmijewski	35000	50,337%
Zmijewski	40000	50,005%
Zmijewski	45000	49,993%
Zmijewski	50000	49,668%



Visuose grafikuose rezultatų grafikas pagal Altman modelį pateikiamas pagal dešiniąją skalę, kadangi jo rodikliai visais atvejais žymiai skyrėsi nuo kitų. Galima pastebėti, kad Altman modelio rezultatai smarkiai svyruoja – vienais atvejais teisingumas siekė 90% ir daugiau, kitais atvejais jis buvo mažesnis nei 15%. Taigi galima teigti, kad polinominės ir sigmoidinės branduolio funkcijos atveju Altman modelio naudojimas nėra tinkamas, tačiau svarbu paminėti, jog rezultatų skirtumas nebuvo toks žymus, per pastaruosius atvejus nustatant tikimybių neskaičiavimą (šie rezultatai čia nepateikiami). Remiantis tuo, galima teigti, kad Altmano modelio naudojimas šiais atvejais apmokymo reikšmių skaičiavimui ypatingai priklauso nuo naudojamų parametrų.

Iš gautų rezultatų akivaizdu, kad geriausią teisingumą parodė v-SVC modelio hibridinimas su Altman modeliu tada, kai buvo naudojamas tiesinė ir RBF branduolio funkcijos. Būtent šie parametrai (Altman diskriminantinis modelis ir RBF bei tiesinė branduolio funkcijos) ir bus naudojami antrajame tyrimo etape.

#### 3.2.1.3. Kitų parametrų įtakos tyrimas

Modelio kūrimui pasirenkame pirmus 5000 įrašų, testavimas atliekamas su sekančiais 50000

įrašų. Šiame etape skaičiuojami ir pateikiami ir kiti įvertinimo rodikliai (recall, TP, FN, TN)<sup>13</sup>. Tyrimas atliekamas indukcinio būdu, t.y., radus parametą, su kuriuo modelis pateikia geriausius rezultatus, kiti parametrai vertinami, atsižvelgiant būtent į nustatytą šio parametro reikšmę. Svarbu pabrėžti, kad tai nebūtinai pats geriausias sprendimas, kadangi tikslesni rezultatai gali būti gauti ir su kitomis parametų kombinacijomis.

1. **Iš pradžių bandoma įvertinti v parametro reikšmę.** Šis parametras yra vienas svarbiausių, kadangi jis aprašo atraminių vektorių skaičių, taigi ir modelio sudėtingumą. Kuo šis parametras mažesnis, tuo atraminių vektorių skaičius mažesnis, t.y., mažesnis skaičiavimo sudėtingumas ir šiems skaičiavimams reikalingi mažesni išteklių.

18 lentelė. v įtakos tyrimas Altman KRV modelio ir RBF bei tiesinės branduolio funkcijos atveju

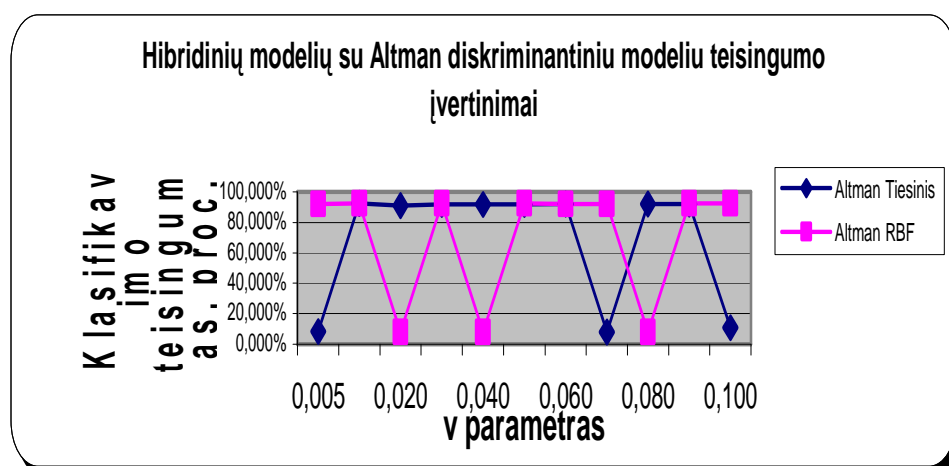
Diskr. modelis	Branduolio funkcija	v	SV skaičius	Accuracy	Recall	FP	TN	FN
Altman	Tiesinis	0,005	26	8,418%	0,072697	0,283632	0,716368	0,994475
Altman	Tiesinis	0,01	50	92,322%	0,264706	0,076332	0,923668	0,006512
Altman	Tiesinis	0,02	100	91,210%	0,15122	0,075214	0,924786	0,158362
Altman	Tiesinis	0,03	150	91,776%	0,203285	0,075213	0,924787	0,094358
Altman	Tiesinis	0,04	200	91,864%	0,220957	0,07518	0,92482	0,084071
Altman	Tiesinis	0,05	250	91,848%	0,283761	0,074006	0,925994	0,102797
Altman	Tiesinis	0,06	300	92,148%	0,230366	0,07587	0,92413	0,037443
Altman	Tiesinis	0,07	352	7,824%	0,074551	0,364742	0,635258	0,997396
Altman	Tiesinis	0,08	400	92,194%	0,24026	0,075954	0,924046	0,029977
Altman	Tiesinis	0,09	450	92,216%	0,183486	0,076226	0,923774	0,022867
Altman	Tiesinis	0,1	500	10,900%	0,081006	0,466019	0,533981	0,967677
Altman	RBF	0,005	26	92,144%	0,433962	0,070684	0,929316	0,114562
Altman	RBF	0,01	50	<b>92,420%</b>	0,51856	0,068457	0,931543	0,112929
Altman	RBF	0,02	100	7,722%	0,066719	0,483733	0,516267	0,987754
Altman	RBF	0,03	150	92,354%	-	0,07646	0,92354	0
Altman	RBF	0,04	200	7,788%	0,070006	0,457383	0,542617	0,991736
Altman	RBF	0,05	250	92,354%	-	0,07646	0,92354	0
Altman	RBF	0,06	300	92,078%	0,246324	0,075531	0,924469	0,051755
Altman	RBF	0,07	352	92,088%	0,23506	0,07566	0,92434	0,048534
Altman	RBF	0,08	400	7,792%	0,070655	0,452037	0,547963	0,992539
Altman	RBF	0,09	450	92,354%	-	0,07646	0,92354	0
Altman	RBF	0,1	500	92,390%	0,666667	0,075822	0,924178	0,004731

Šaltinis: sudaryta autoriaus.

Iš gautų rezultatų galima daryti išvadą, kad didelis atraminių vektorių skaičius nebūtinai reiškia modelio tikslumą, kadangi tiksliausi rezultatai abiejų kerneliu atveju buvo gauti tada, kai  $v=0.01$ , t.y., kai atraminių vektorių skaičius nėra didelis. Šiuo atveju didžiausias gautas tikslumas (teisingumas) lygus 92,420%.

<sup>13</sup> Šie rodikliai pateikiami tik pirmajai klasei (įmonės, turinčios didelę bankroto riziką), kadangi tyrimas orientuotas labiau į rizikingų įmonių identifikavimą.





Šaltinis: sudaryta autoriaus.

**24 pav. v parametro įtakos Altman modelio ir RBF bei tiesine branduolio funkcijomis atveju rezultatai**

2. Bandoma įvertinti  $\gamma$  parametro įtaką. Šis parametras įtaką turės tik RBF branduolio funkcijos atveju.

19 lentelė.  $\gamma$  parametro įtakos tyrimas Altman KRV modelio ir RBF branduolio funkcijos atveju

Diskr. Modelis	coef	C	v	$\gamma$	SV sk.	Accuracy	Recall	FP	TN	FN
Altman	0	1	0,01	1	50	92,420%	0,51856	0,068457	0,931543	0,112929
Altman	0	1	0,01	0,1	50	92,356%	1	0,076442	0,923558	0
Altman	0	1	0,01	0,3	50	90,816%	0,140318	0,075065	0,924935	0,200131
Altman	0	1	0,01	0,001	50	90,508%	0,332729	0,061493	0,938507	0,387906
Altman	0	1	0,01	0,005	50	92,16%	0,460432	0,066606	0,933394	0,172106
Altman	0	1	0,01	2	50	92,354%	-	0,07646	0,92354	0

Šaltinis: sudaryta autoriaus.

Iš gautų rezultatų akivaizdu, kad išskirti kokią priklausomumo tendenciją yra praktiškai neįmanoma, todėl tinkamiausių parametų, nusakančių branduolio funkcijos transformaciją (coef0,  $\gamma$ ), radimui reikalingos kitos procedūros bei algoritmai. Pastebėtina, kad visi gauti rezultatai yra virš 90%; kad rezultatai tikslūs, parodo ir aukštos TN reikšmės.

**3.2.2. Gautų rezultatų apibendrinimas**

Galutiniai atlikto tyrimo rezultatai apibendrinami 20 lentelėje, kurioje pateikiami geriausi rezultatai, gauti naudojant kiekvieną iš tirtų diskriminantinių modelių.

Siekiant suvienodinti tyrimo sąlygas, priimta, kad visais atvejais naudota 5000 įrašų apmokymui ir sekantys 50000 įrašų testavimui. Čia pateikiami pilni tyrimo rezultatai – visi abiejų klasių maišos matricos parametrai.

20 lentelė. Tyrimo metu gauti rezultatai

Diskr. modelis	Altman	Springate	Zmijewski
Branduolio funkcija	RBF	RBF	RBF
$\nu$	0,01	0,05	0,05
$\gamma$	1	1	1
SV skaičius	50	250	250
Teisingumo reikšmė	92,420%	57,380%	59,152%
Recall	Klasė 0	0.5185601799775028	0.6022727272727273
	Klasė 1	0.9315428315448677	0.5734961190168176
FP	Klasė 0	0.06845716845513225	0.4265038809831824
	Klasė 1	0.4814398200224972	0.3977272727272727
TN	Klasė 0	0.9315428315448677	0.5734961190168176
	Klasė 1	0.5185601799775028	0.6022727272727273
FN	Klasė 0	0.11292875989445911	0.0098545283904270
	Klasė 1	0.887071240105541	0.990145471609573
P	Klasė 0	0.12058592728223907	0.0148473246801755
	Klasė 1	0.9907313164562445	0.9926527184941571

Šaltinis: sudaryta autoriaus.

Akivaizdu, kad šiuo atveju teisingusiai klasifikavo SVM su Altman modelis, tačiau reikia atsižvelgti ir į dar vieną svarbų parametą – tikslumą (P), kuris nusako prognozavimo naudojant šį modelį tikslumą. Galima pastebėti, kad visi trys modeliai priklausymą klasei 1 (t.y., įmonės, turinčios mažą bankroto riziką) prognozuoja su labai dideliu tikslumu, tuo tarpu priklausymą pirmajai klasei geriausiai prognozuoja vėlgi SVM su Altman modeliu. Tuo remiantis, galima teigti, kad geriausi rezultatai gaunami suhibridinius SVM su Altman diskriminantiniu modeliu

### 3.3. Sukurtos sistemos realizacijos aprašymas

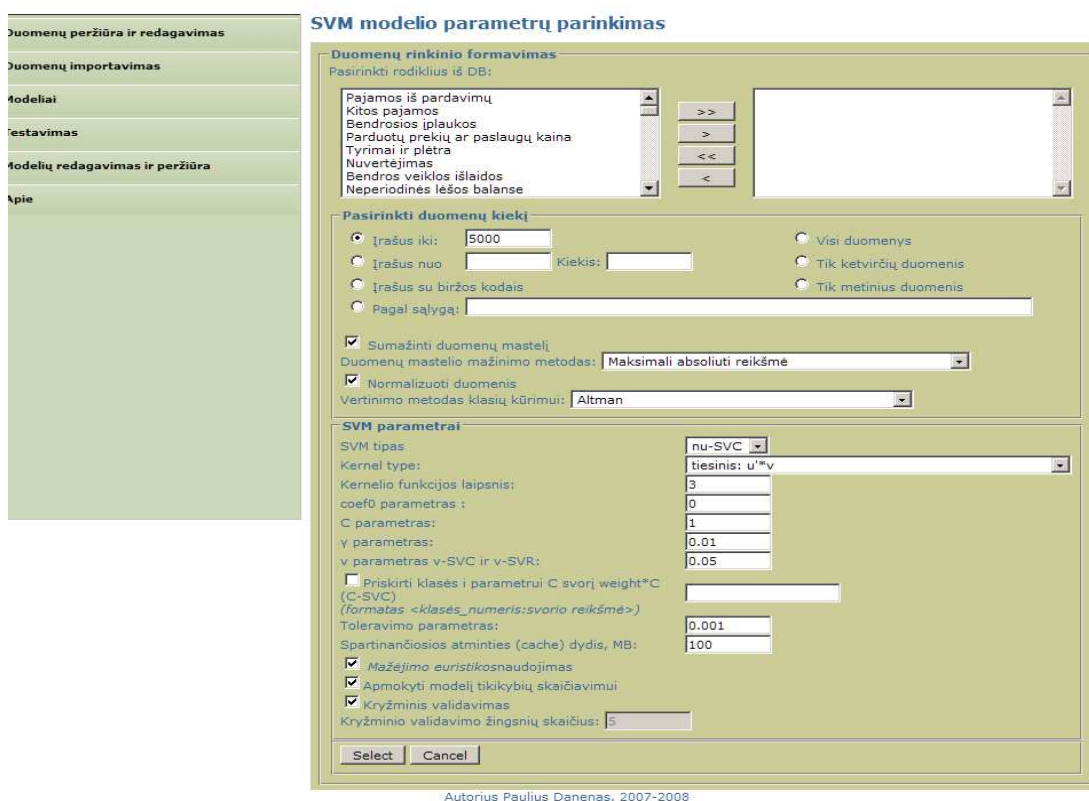
Eksperimentui buvo naudojama autoriaus sukurta eksperimentinė „CreditRiskEval“ sistema, skirta būtent modelių kredito rizikos vertinimui kūrime ir testavimui. Jos struktūra, diagramos bei specifikavimas pateikiami 2 šio darbo skyriuje. Šiuo metu sukurtoje sistemoje realizuotos tokios galimybės:

- Duomenų (bendrovių bei pramonės šakų biržos kodų bei finansinių rodiklių) įvedimas, redagavimas ir šalinimas;
- Duomenų (tiek (bendrovių bei pramonės šakų biržos kodų, tiek ir finansinių rodiklių) importavimas iš CSV failo;
- $\nu$ -SVM metodų paremtų modelių kūrimas, testavimas;
- Modelių bazė, kurioje saugojami sukurti modeliai ir testavimo parametrai, joje esančių modelių duomenų (pavadinimo, aprašymo) redagavimas bei pačių modelių šalinimas iš modelių DB;
- Modelio testavimas, naudojant maišos matricą ir jos parametrus.

Ši sistema ateityje gali būti išplėsta naujomis funkcijomis, aprašytomis šiame darbe. Sistema

<sup>14</sup> NaN – Not A Number; paprastai toks rezultatas gaunamas, kai yra dalyba iš 0.

realizuota JAVA kalba, kas užtikrina jos nepriklausomumą nuo naudojamos platformos (t.y., sistema gali veikti tiek Windows, tiek Linux, tiek BSD ar Macintosh platformose). Duomenų saugojimui naudojama MySQL duomenų bazė, taip pat turinti versijas įvairioms platformoms. Pačios sistemos veikimui reikalingas atvirojo kodo Apache Tomcat serveris; taip sudaroma galimybė sistemą naudoti vidiniame organizacijos (banko, akademinės įstaigos) intranete arba ektranete. Ateityje sistemą planuojama pritaikyti ir kitoms DBVS bei Java aplikacijų serveriams, tokiems, kaip Sun Java Application Server. Sistema įdiegiama nesudėtingai – į *webapps* direktoriją, esančią Tomcat instaliaciniame kataloge, užtenka nukopijuoti *CreditRiskEval.war* failą, tačiau į *common/lib* direktoriją turi būti nukopijuotas *mysql-connector-java-5.0.4-bin.jar* failas, reikalingas prisijungimui bei darbui su MySQL duomenų baze.



Autorius Paulius Danenas, 2007-2008

Šaltinis: sudaryta autoriaus.

## 25 pav. Sukurtos sistemos vartotojo sąsajos pavyzdys

Programos vartotojo sąsaja nesudėtinga ir perprantama intuityviai; ateityje planuojama sukurti ir pagalbos vartotojui sistemą, pateikiančią ir parametrų aprašymus, kas vartotojui palengvintų darbą renkantis tinkamus parametrus. 25 pav. pateikiamas vartotojo sąsajos pavyzdys; 9 priede pateikiami kiti jos pavyzdžiai.

#### 4. TOLIMESNIO TYRIMO GALIMYBĖS IR PERSPEKTYVA

Atliktas eksperimentas parodė, kad atraminių vektorių mašinų metodo taikymas kredito rizikos vertinime turi gana neblogas galimybes. Tyrimai, atlikti šiame darbe, parodė, kad klasifikavimo tikslumas, naudojant šį metodą, gali siekti 90-92% (geriausias rezultatas pasiektas, kai naudojama RBF branduolio funkcija,  $\nu=0,01$ ,  $\gamma=0,01$ ), tačiau tikėtina, kad šie rezultatai gali būti dar geresni, aprašomame metode naudojant tokius metodus:

- Požymių išskyrimas (feature extraction), leidžiantis iš viso naudojamų rodiklių sąrašo atrinkti esminius kintamuosius, ir taip sumažinti modelyje naudojamų rodiklių, o tuo pačiu ir skaičiavimų, apimtį;

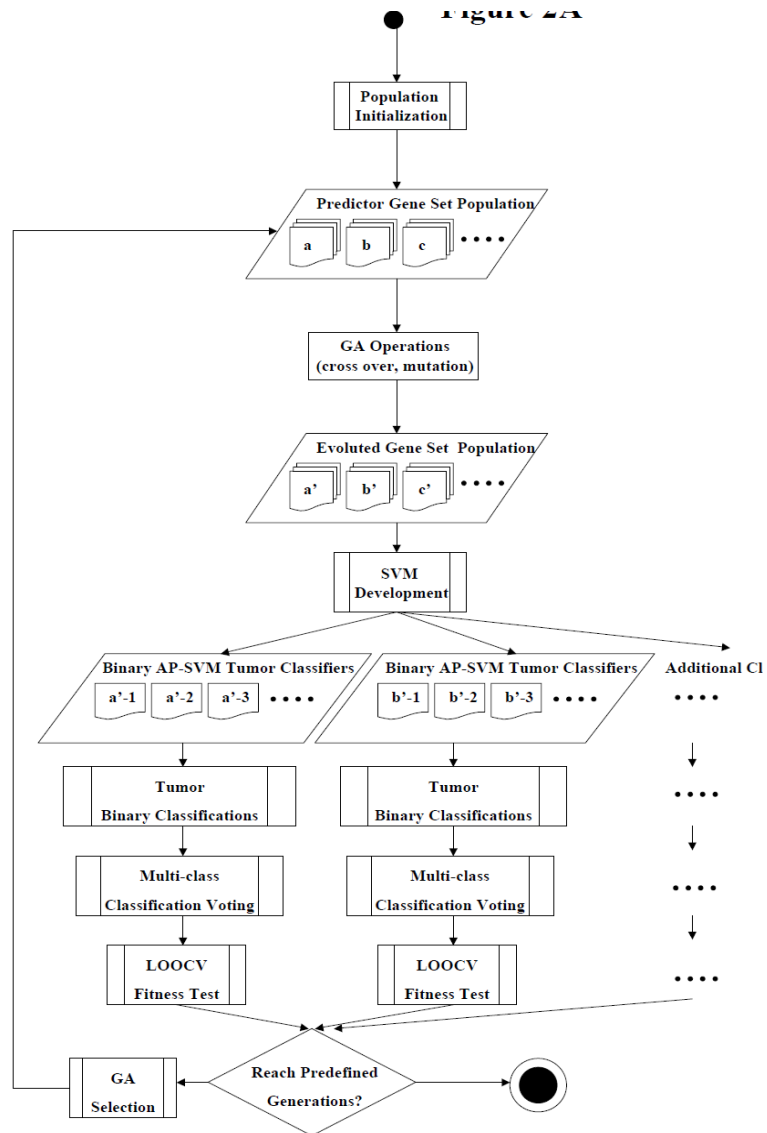
- Rasti SVM parametrus, kurie leistų pasiekti dar geresnį klasifikavimo tikslumą;

- Naudojant evoliucinius skaičiavimus, sukurti hibridinį metodą, kuris leistų automatiškai rasti geriausias SVM parametrų reikšmes, t.y., reikšmes, su kuriomis pasiekiami geriausi rezultatai. Vienas iš tokių būdų būtų metodas, kuomet apmokomas didelis skaičius SVM ir jų efektyvumas vertinamas pagal tinkamumo funkciją, įvertinančių klasifikavimo tikslumą ir efektyvumą. Šiems skaičiavimams realizuoti galima naudoti vieną iš GA (tęstinį, hibridinį ar kt.) ar kitų evoliucinių skaičiavimų (tabu paieška, spiečiaus optimizacija ar skruzdžių kolonijos optimizacija) metodus (jų aprašymai, privalumai ir trūkumai pateikiami 6 priede). Genetinio algoritmo atveju realizacija būtų paprastesnė, o naudojant elitizmo parinkimo metodą, tikėtinas ir geras našumas bei tikslumas. Spiečiaus optimizacijos ar skruzdžių kolonijos optimizacijos atveju realizacija sudėtingesnė, tačiau šie metodai leistų išspręsti sudėtingas problemas, ypač turinčias daug lokalių minimumų;

- Galimas ir kitas hibridinimo variantas – genetinis algoritmas būtų naudojamas atrinkti geriausiems atvejams, o šie atvejai būtų pateikiami atraminių vektorių mašinai, kuri vėliau vėl atliktų klasifikavimą. Kaip tokio algoritmo pavyzdį galima pateikti Jane Jijun Liu, Gene Cutler, Wuxiong Li, Zheng Pan, Sihua Peng, Tim, Hoey, Liangbiao Chen and Xuefeng Bruce Ling sukurtą GA\SVM algoritimą (26 pav.).

- Panaudojant kurią kitą SVM variaciją (kai kurios jų pateikiamos 11 lentelėje).

Sukurtas metodas gali būti išplėstas ir kokybinių (ekspertinių) parametrų panaudojimu, leidžiančiu sistemoje panaudoti ir ekspertines žinias (*fuzzy SVM*). Tokių žinių vertinimas aprašytas 1.5.2 skyrelyje, o pavyzdiniai ekspertiniai įvertinimai – 1.5.3 skyrelyje bei 5 priede pateikta jų realizacija Jess aplinkoje, kas jas leidžia integruoti į dabartinę sistemą. Be to, taikant SVM regresijos metodus ( $\nu$ -SVR, C-SVR), galima apskaičiuoti modelį kaip regresinę lygtį.



Šaltinis: Jane Jijun Liu, Gene Cutler, Wuxiong Li, Zheng Pan, Sihua Peng, Tim, Hoey, Liangbiao Chen and Xuefeng Bruce Ling. Multiclass Cancer Classification and Biomarker Discovery Using GA-based Algorithms.

**26 pav. Galimas GA/SVM hibridinis algoritmas**

## IŠVADOS

1. Vertinant tiek banko, tiek ir bet kokios kitos finansinės ar kito tipo institucijos finansinį pajėgumą, svarbu atsižvelgti ne tik į finansinius rodiklius, bet ir kitus aspektus, t.y., vertinant kredito riziką, svarbu įvertinti ir kitas rizikas. Kai kurios iš šių rizikų gali būti vertinamos ekspertiniu būdu, t.y., jų įvertinimui gali būti naudojami kokybiniai įverčiai; taip pat daugumoje jų naudojama techninė analizė bei matematiniai/ekonometriniai metodai. Tai sudaro prielaidas taikyti DI modelius ir ekspertines sistemas.
2. Kredito rizika gali būti analizuojama fundamentinės analizės principu, pradedant analizuoti nuo makroekonominės aplinkos analizės ir baigiant skolininko veiklos rodiklių analize, arba atvirkščiai, pradedant nuo skolininko analizės. Ir vienu, ir kitu atveju tiriami rodikliai gali apimti tiek kiekybinius, tiek kokybinius rodiklius, kurie gali būti analizuojami atskirai, taikant skirtingus modelius, tačiau naudojant tą patį principą (klasifikavimą ar pan.)
3. Iš klasikinių rizikos vertinimo metodų galima išskirti rizikos vertinimą balais (reitingavimą), naudojant klasikinius diskriminantinius bei modernius rizikos vertinimo modelius. Ir vieni, ir kiti turi savo privalumų ir trūkumų, tačiau daugumos jų tikslas yra suskirstymas į grupes arba reitingų priskyrimas; taigi galimas klasifikavimo ir klasterizavimo metodų taikymas kartu su jais arba integravimas į juos.
4. Dirbtinis intelektas nuo pat atsiradimo buvo tiriamas ir bandomas taikyti įvairių sudėtingų problemų, tarp jų ir susijusių su ekonomika bei finansais, sprendimui, taigi yra galimybė jį taikyti ir kredito rizikoje. Jo pagrindu kuriamos ekspertinės arba sprendimų paramos sistemos, skirtos pagelbėti individui priimti sprendimus (bet jų nepriimančios pačios). Galimas ir tokių sistemų panaudojimas kredito rizikos vertinimo procese, kaip pavyzdys pateiktas tam skirtos galimos sprendimų paramos sistemos pavyzdys ir jo veikimo principas.
5. Šiuo metu žinoma daug dirbtinio intelekto (mašininio mokymo, duomenų gavybos ir ekspertinių žinių panaudojimo) metodų; patys populiariausi ir jų ankstesnis taikymas kredito rizikos vertinime apžvelgti šiame darbe. Atlikus šių metodų analizę, nuspręsta orientuotis į atraminių vektorių mašinų (angl. Support Vector Machines, sutr. SVM) metodą ir jo taikymo galimybes, motyvuojant tuo, kad jo taikymas kredito rizikoje mažai tirtas anksčiau, be to, jis leidžia aproksimuoti bet kokio sudėtingumo funkciją bei naudoti bet kokį rodiklių skaičių .
6. Pasiūlyta koncepcinė schema (angl. framework), pagal kurią ateityje galėtų būti sukurta sistema, skirta dirbtiniu intelektu paremtų modelių kūrimui; ši sistema galėtų būti sukurta kaip atskiras modulis arba integruota į jau esamą. Šiuo metu yra sukurtas tik SVM (v-SVM) metodą realizuojantis šios sistemos prototipas, tačiau ateityje jis galėtų būti išplėstas tiek kitais klasifikavimo,

ties ir klasterizavimo metodais ir galbūt jų rezultatų palyginimo galimybėmis.

7.  $\nu$ -SVM metodo tyrimas parodė, kad geriausi rezultatai gaunami jį suhibridinus su Altman diskriminantiniu modeliu bei naudojant tiesinę ir RBF kernelio funkcijas. Tyrime naudoti visi turimi pradiniai ir išvestiniai rodikliai, iš viso jų 56. Didžiausias pasiektas tikslumas yra 92,420% (RBF branduolio funkcija,  $\nu=0,01$ ,  $\gamma=0,01$ ), tačiau tikėtina, kad gali būti pasiektas ir dar didesnis tikslumas, tinkamai parinkus įvairius SVM parametrus.
8. Pateikti pasiūlymai, kaip būtų galima bandyti gauti dar geresnius rezultatus, o taip pat aptartos galimybės sukurti hibridinius algoritmus, SVM metodą naudojant kartu su evoliuciniais algoritmais. Taip pat aptartos ir ekspertinių žinių panaudojimo galimybės, kaip pavyzdį pateikiant realios kredito rizikos vertinimo ekspertinės sistemos įvertinimus ir pagal juos sumodeliuojant pavyzdines ekspertines taisykles.

## LITERATŪRA

1. VAŠKELAITIS, Vytautas. Pinigai: komerciniai bankai ir jų rizikos valdymas. Vilnius, 2003.
2. BOGUSLAUSKAS, V. Ekonometrika. Kaunas: Technologija, 2007.
3. KENDALL, Graham. G5AIAI : History of AI : History. 2001.09.11. Adresas Internetete: <http://www.cs.nott.ac.uk/~gjk/courses/g5ai/002history/history.htm>. Žiūrėta 2005.11.25
4. HENKE, Stottler. Artificial Intelligence History. 2002-2004. Adresas Internetete: [http://www.stottlerhenke.com/ai\\_general/history.htm](http://www.stottlerhenke.com/ai_general/history.htm). Žiūrėta 2005.11.25
5. RUSSELL, Stuart J.; NORVIG, Peter. Artificial Intelligence: A Modern Approach. Prentice-Hall, 1995.
6. SCHLOTTMANN, Frank; SEESE, Detlef. A Hybrid Genetic-Quantitative Method For Risk-Return Optimisation Of Credit Portfolios (2001). Proceedings of the Conference of Quantitative Methods in Finance, University of Technology, Sydney, Australia, p. 55, 2001. Adresas Internetete: <http://citeseer.ist.psu.edu/schlottmann01hybrid.html>. Žiūrėta 2006.12.11
7. SCHLOTTMANN, Frank; SEESE, Detlef. Hybrid Multi-Objective Evolutionary Computation of Constrained Downside Risk-Return Efficient Sets for Credit Portfolios (2002). No 78, Computing in Economics and Finance 2002 from Society for Computational Economics Adresas Internetete: <http://citeseer.ist.psu.edu/schlottmann02hybrid.html>. Žiūrėta 2006.12.11
8. APOTEKER, Thierry; BARTHELEMY, Sylvain. "Genetic Algorithms and Financial Crises in Emerging Markets" (2000 gegužė). CEFI International Conference Proceedings. Adresas Internetete: <http://www.sylbarth.com/pdf/cefi-sienne2000.pdf>. Žiūrėta 2006.12.11
9. VELLIDO, A.; LISBOA, P.J.G.; VAUGHAN, J. Neural networks in business: a survey of applications (1992–1998). Expert Systems with Applications 17 (1999) 51–70, 1999. Adresas Internetete: <http://www.personal.psu.edu/faculty/p/x/pxp19/INFSY554/surveyNN.pdf>. Žiūrėta 2006.12.11
10. LAI, Kin Keung; YU, Lean; WANG, Shouyang; ZHOU, Ligang. Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model. 2006.07.19. Adresas Internetete: <http://madis1.iss.ac.cn/madis.files/pub-papers/2006/41320682.pdf>. Žiūrėta 2006.12.11
11. LAI, Kin Keung; YU, Lean; WANG, Shouyang; ZHOU, Ligang. Neural Network Metalearning for Credit Scoring. 2006.06.23. Adresas Internetete: <http://madis1.iss.ac.cn/madis.files/pub-papers/2006/41130403.pdf>. Žiūrėta 2006.12.11
12. TOLLO, Giacomo di. Credit Risk: A Neural Net Approach. Electronic Proceedings of RCRA 2006 (Udine (I), 23.06.2006). Adresas Internetete: <http://www.dim.uniud.it/bortolus/rcra06/files/ditollo.pdf>. Žiūrėta 2006.12.11
13. BAESSENS, Bart; VAN GESTEL, Tony; STEPANOVA, Maria; VAN DEN POEL, Dirk. Neural Network Survival Analysis for Personal Loan Data. Journal of the Operational Research Society, 59, (9), 1089-1098. Adresas Internetete: [http://www.feb.ugent.be/fac/research/WP/Papers/wp\\_04\\_281.pdf](http://www.feb.ugent.be/fac/research/WP/Papers/wp_04_281.pdf). Žiūrėta 2006.12.11
14. HANDZIC, Meliha; TJANDRAWIBAWA, Felix; YEO, Julia. How Neural Networks Can Help Loan Officers to Make Better Informed Application Decisions. Informing Science, June 2003. Adresas Internetete: <http://proceedings.informingscience.org/IS2003Proceedings/docs/024Handz.pdf>. Žiūrėta 2006.12.11
15. Britannica Concise. Adresas Internetete: <http://concise.britannica.com/>. Žiūrėta 2006.12.11
16. VAN DEN BERG, Jan. Credit Rating Prediction with Self-Organizing Maps. Expert Systems with Applications, Volume 30, Issue 3, April 2006, Pages 479-487. Adresas Internetete: [http://www.abo.fi/instut/iamsr/eunite/reports/CS\\_CreditRating\\_SOM.pdf](http://www.abo.fi/instut/iamsr/eunite/reports/CS_CreditRating_SOM.pdf). Žiūrėta 2006.12.11
17. DEBOECK, G. 1998. "Financial Applications of Self-Organizing Maps". Neural Network World, 8(2):213--241. Adresas Internetete: <http://citeseer.ist.psu.edu/deboeck98financial.html>. Žiūrėta 2006.12.11
18. KASKI, S.; SINKKONEN, J.; PELTONEN, J. Bankruptcy Analysis with Self-Organizing Maps in Learning Metrics. IEEE Transactions on Neural Networks, 2001. Adresas Internetete: <http://citeseer.ist.psu.edu/kaski01bankruptcy.html>. Žiūrėta 2006.12.11



19. MERKEVIČIUS, Egidijus; GARŠVA, Gintautas; SIMUTIS, Rimvydas. Forecasting Of Credit Classes With The Self-Organizing Maps. ISSN 1392 – 124X Informacinės Technologijos Ir Valdymas, 2004, Nr.4(33). Adresas Internete: <http://itc.ktu.lt/itc33/Simut33.pdf>. Žiūrėta 2006.12.11
20. ZUMBACH, Gilles; PICTET, Olivier V.; MASUTI, Oliver. Genetic Programming with Syntactic Restrictions applied to Financial Volatility Forecasting. 2001.04.20. Adresas Internete: <http://www.olsen.ch/research/workingpapers/gpForVolatility.pdf>. Žiūrėta 2006.12.11
21. AT&T Speech and Images Processing Services Research Lab, Computer Learning Research Centre Web Page (interaktyvus). Adresas Internete: <http://svm.dcs.rhbnc.ac.uk/>. Žiūrėta 2006.12.11
22. HUANGA, Zan; CHENA, Hsinchun; HSUA, Chia-Jung; CHENB, Wun-Hwa; WU, Soushan. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems 37 (2004), p. 543– 558, 2003 liepos 4 d. Adresas Internete: <http://www.personal.psu.edu/faculty/h/u/huz2/Zan/papers/credit.dss.pdf>. Žiūrėta 2006.12.11
23. WANG, Yongqiao; WANG, Shouyang; LAI, Kin Keung. A New Fuzzy Support Vector Machine to Evaluate Credit Risk. IEEE Transactions On Fuzzy Systems, Vol. 13, NO. 6, 2005 gruodis. Adresas Internete: <http://madis1.iss.ac.cn/madis.files/01556587.pdf>. Žiūrėta 2006.12.11
24. LAI, Kin Keung; YU, Lean; WANG, Shouyang; ZHOU, Ligang. Credit Risk Evaluation with Least Square Support Vector Machine. 2005.06.26. Adresas Internete: <http://madis1.iss.ac.cn/madis.files/pub-papers/2006/40620490.pdf>. Žiūrėta 2006.12.11
25. SHIN, Kyung-Shik; LEE, Taik Soo; KIM, Hyun-jung. An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications Volume 28, Issue 1 , January 2005, Pages 127-135.
26. MIN, Jae H., LEE, Young-Chan. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications Volume 28, Issue 4 , May 2005, Pages 603-614.
27. riskglossary.com (interaktyvus). Adresas Internete: <http://www.riskglossary.com/>. Žiūrėta 2006.12.03
28. LAI, Kin Keung; YU, Lean; WANG, Shouyang; ZHOU, Ligang. Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model. Lecture Notes in Computer Science, LNCS 4132, edited by S. Kollias, Springer, 2006, pp682-690. Adresas Internete: <http://madis1.iss.ac.cn/madis.files/pub-papers/2006/41320682.pdf>. Žiūrėta 2006.12.11
29. OBITKO, Marek. Introduction to genetic algorithms with interactive Java applets. 1998. Adresas Internete: <http://cs.felk.cvut.cz/~xobitko/ga/index.html>. Žiūrėta 2006.12.11
30. ALTMAN, Edward I. Financial Ratios, Discriminant Analysis And The Prediction of Corporate Bankruptcy. The Journal of Finance, Vol. 23, No. 4 (Sep., 1968), 589-609.
31. GARŠKAITĖ K.; GARŠKIENĖ A.. Įmonių bankroto diagnostikos sistema // Verslas: teorija ir praktika, 2003 , IV tomas, Nr.4
32. ALLEN L. Credit Risk Modeling of Middle Markets. 2002. Adresas Internete: <http://pages.stern.nyu.edu/~lallen/whartonmidmarket.pdf>. Žiūrėta 2007.06.13
33. KANCEREVYČIUS, Gitanas. Finansai ir investicijos. Kaunas, Smaltija, 2004.
34. KAASTRA, Ieabeling; BOYD, Milton (1996). Designing a neural network for forecasting financial and economic time series. Neurocomputing 10(3): 215-236. Adresas Internete: <http://citeseer.ist.psu.edu/kaastra96designing.html>. Paskutinį kartą žiūrėta: 2006.12.11
35. JANTZEN, Jan. Tutorial On Fuzzy Logic. 1998.08.19. Adresas Internete: [www.iau.dtu.dk/~jj/pubs/logic.pdf](http://www.iau.dtu.dk/~jj/pubs/logic.pdf). Paskutinį kartą žiūrėta: 2007.06.25
36. ABNER, Sundance L. Fuzzy Logic in Expert Systems. Adresas Internete: <http://www.cs.rockhurst.edu/seminars/CS2002/Sundance>. Paskutinį kartą žiūrėta: 2007.06.25

37. ELIZALDE, A. Credit Risk Models II: Structural Models. MSc in Financial Mathematics, King's College London, 2003. Adresas Internetė: <http://citeseer.ist.psu.edu/elizalde03credit.html>. Paskutinį kartą žiūrėta: 2007.06.25
38. ELIZALDE, A. Credit Risk Models III: Reconciliation Reduced - Structural Models. MSc in Financial Mathematics, King's College London, 2003. Adresas Internetė: <http://citeseer.ist.psu.edu/article/elizalde03credit.html>. Paskutinį kartą žiūrėta: 2007.06.25
39. MAHOTRA, Rashmi, MAHOTRA, D.K.. Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research 136 (2002), Pages 190-211.
40. PIRAMUTHU, Selwyn. Financial credit-risk evaluation with neural and neurofuzzy systems. European Journal of Operational Research 112 (1999), Pages 310-321
41. KHAN, Haider A.. Can banks learn to be rational? University of Denver, 2002 kovas.
42. VAPNIK, V. The Nature of Statistical Learning Theory. Springer-Verlag, 2000.
43. HAUPT, Randy L., HAUPT, Sue Ellen. Practical Genetic Algorithms, 2nd Ed. Wiley, 2004.
44. CANTÚ-PAZ, E. (1998). A survey of parallel genetic algorithms. Calculateurs Paralleles, Reseaux et Systems Repartis, 10(2), 141-171. Adresas Internetė: <http://www.illgal.uiuc.edu/web/advancedgec/wp-content/files/cparalleles98-survey.ps>.
45. MERKEVIČIUS, E., GARŠVA, G., CEPKOVATAJA, O. Intelektualios sprendimų paramos sistemos kredito rizikos vertinimui struktūra. Informacinės technologijos'2006, p.725-733. Adresas Internetė: [http://www.ktu.lt/lt/apie\\_renginius/konferencijos/2006/k6\\_02/IT2005/Sekc13.pdf](http://www.ktu.lt/lt/apie_renginius/konferencijos/2006/k6_02/IT2005/Sekc13.pdf).
46. SIMUTIS, Rimvydas. Neuroniniai tinklai ir neuroskaičiavimai. Paskaitų konspektai, Kaunas, VU, 2006.
47. DUNHAM, Margaret H. Data Mining: Introductory and Advanced Topics. Prentice-Hall, Pearson Education Inc., 2003.
48. DANĖNAS P., MERKEVIČIUS E., GARŠVA G. Sistemos modulis, skirtas intelektualių modelių kredito rizikos vertinimui kūrimui, koncepcinė struktūra. Informacinės technologijos 2008, p.62-68.
49. jFreeChart [interaktyvus]. Adresas Internetė: <http://www.jfree.org/jfreechart/>. Žiūrėta 2008.03.28.
50. jMatLink [interaktyvus]. Adresas Internetė: <http://jmatlink.sourceforge.net/>. Žiūrėta 2008.03.28.
51. SourceForge. Adresas Internetė: <http://sourceforge.net/>. Žiūrėta 2008.03.28.
52. Weka 3: Data Mining Software in Java. Adresas Internetė: <http://www.cs.waikato.ac.nz/ml/weka/>. Žiūrėta 2008.03.28.
53. Rapid-I [interaktyvus]. Adresas Internetė: <http://rapid-i.com/content/blogcategory/10/69/lang,en/>. Žiūrėta 2008.03.28
54. JSR-000073 Data Mining API (Final Release)[interaktyvus]. Adresas Internetė: <http://jcp.org/aboutjava/communityprocess/final/jsr073/index.html> Žiūrėta 2008.03.28
55. jFuzzyLogic: Open Source Fuzzy Logic (Java)[interaktyvus]. Adresas Internetė: <http://jfuzzylogic.sourceforge.net/html/index.html>. Žiūrėta 2008.04.10
56. Jess, the Rule Engine for the Java Platform [interaktyvus]. Adresas Internetė; <http://herzberg.ca.sandia.gov/> Žiūrėta 2008.04.10
57. Quang Nhat Nguyen. Machine Learning: Algorithms and Applications. Teaching material. Faculty of Computer Science, Free University of Bozen-Bolzano, October, 2007.
58. LIBSVM FAQ [interaktyvus]. Adresas Internetė: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/faq.html#f411>. Žiūrėta 2008.03.28
59. CHANG, Chih-Chung, LIN, Chih-Jen. LIBSVM: a Library for Support Vector Machines. Adresas Internetė: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>. Žiūrėta 2008.03.28

60. Confusion Matrix [interaktyvus]. Adresas Internetė :  
[http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion\\_matrix/confusion\\_matrix.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html). Žiūrėta 2008.03.28
61. LIBSVM -- A Library for Support Vector Machines [interaktyvus]. Adresas Internetė :  
<http://www.csie.ntu.edu.tw/~cjlin/libsvm/> Žiūrėta 2008.05.14
62. PELCKMANS, K., SUYKENS Johan A.K., VAN GESTEL T., DE BRABANTER, LUKAS, J., L., HAMERS B., DE MOOR, B., VANDEWALLE J.. LS-SVMLab: a MATLAB/C toolbox for Least Squares Support Vector Machines. Adresas Internetė :  
[http://www.esat.kuleuven.ac.be/sista/lssvmlab/tutorial/lssvmlab\\_paper0.pdf](http://www.esat.kuleuven.ac.be/sista/lssvmlab/tutorial/lssvmlab_paper0.pdf). Žiūrėta 2008.05.14
63. MANGASARIAN, O. L., MUSICANT, David R. Lagrangian Support Vector Machine Classification. Data Mining Institute, Computer Sciences Department, University of Wisconsin, 2000 birželis. Adresas Internetė:  
<ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-06.ps>. Žiūrėta 2008.05.14
64. MANGASARIAN, O. L., MUSICANT, David R. Active Support Vector Machine Classification. Data Mining Institute, Computer Sciences Department, University of Wisconsin, 2000 balandis. Adresas Internetė:  
<ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-04.ps>. Žiūrėta 2008.05.14
65. LEE, Yuh-Jye, MANGASARIAN, O. L. A Smooth Support Vector Machine. Data Mining Institute, University of Wisconsin, Technical Report 99-03. Computational Optimization and Applications, 2001, 20 tomas, p.5-22. Adresas Internetė: <ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-03.ps>. Žiūrėta 2008.05.14
66. FUNG, G., MANGASARIAN, O. L.. A Feature Selection Newton Method for Support Vector Machine Classification. Data Mining Institute, Computer Sciences Department, University of Wisconsin, 2002 rugsėjis. Adresas Internetė: <ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/02-01.ps>. Žiūrėta 2008.05.14
67. FUNG, G., MANGASARIAN, O. L.. Proximal Support Vector Machine Classifiers. Data Mining Institute, Computer Sciences Department, University of Wisconsin. Proceedings KDD-2001: Knowledge Discovery and Data Mining, August 26-29, 2001, p. 77-86; San Francisco, CA, leidėjas - Association for Computing Machinery, New York. Adresas Internetė: <ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-02.ps>. Žiūrėta 2008.05.14
68. JOACHIMS, Thorsten. SVMlight - Support Vector Machine [interaktyvus]. Paskutinį kartą redaguotas 2007.11.07. Adresas Internetė: [http://www.cs.cornell.edu/people/tj/svm\\_light/](http://www.cs.cornell.edu/people/tj/svm_light/). Žiūrėta 2008.05.14
69. NEDOVIĆ Ljubica, DEVEDŽIĆ Vladan. Expert systems in finance – a cross-section of the field. Expert Systems With Applications, Vol.23, No.1, 2002, pp. 49-66. Adresas Internetė:  
<http://fon.fon.bg.ac.yu/~devedzic/ESwA2002-1.pdf>. Žiūrėta 2008.05.14
70. OMG SysML v. 1.0 (Available Specification) [September 2007]. Adresas Internetė:  
<http://www.sysml.org/docs/specs/OMGSysML-v1.0-07-09-01.pdf>. Žiūrėta 2008.05.14
71. Jane Jijun Liu, Gene Cutler, Wuxiong Li, Zheng Pan, Sihua Peng, Tim, Hoey, Liangbiao Chen and Xuefeng Bruce Ling. Multiclass Cancer Classification and Biomarker Discovery Using GA-based Algorithms [interaktyvus]. Adresas Internetė; <http://www.fishgenome.org/publication/Liu/bioinformatics/>. Žiūrėta 2008.05.14

## PRIEDAI

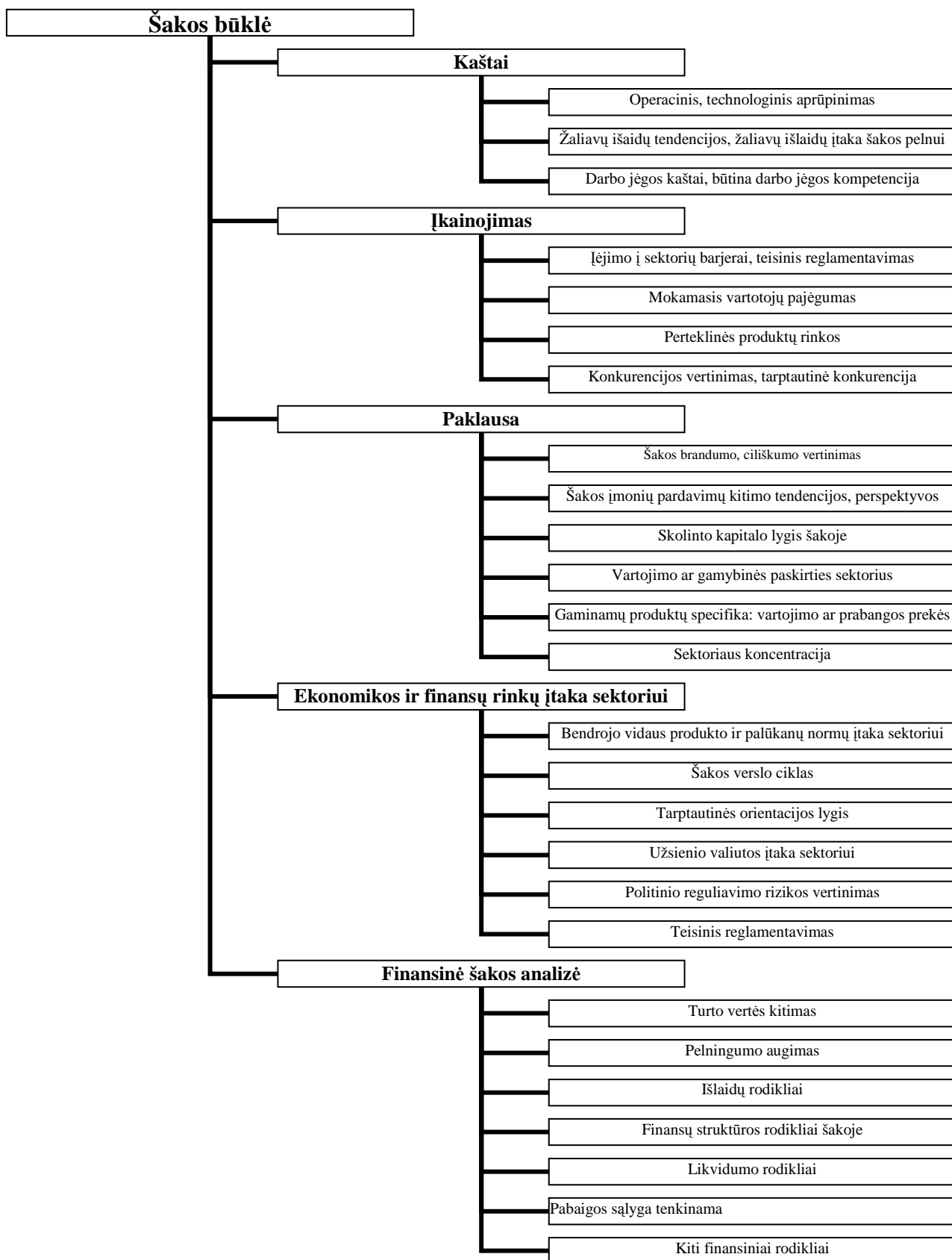
1 priedas. Galimi kredito rizikos vertinimo kintamieji ir kriterijai .....	90
2 priedas. Darbe naudojamų diskriminantinių rizikos vertinimo modelių sudėtis ir išraiškos.....	96
3 priedas. Neuroninių tinklų taikymas bankroto prognozavime bei kredito rizikos vertinime 1992-1998 m. ....	98
4 priedas. Fuzzy logikos operacijos ir išvadų darymo metodai.....	99
5 priedas. Ekspertinių taisyklių realizacija CLIPS/JESS ekspertinėse aplinkose .....	100
6 priedas. Evoliucinių metodų strategijos ir algoritmai.....	103
7 priedas. Sistemoje naudojamų duomenų aprašymas ir duomenų bazės schema .....	107
8 priedas. Pagrindinių šiuolaikinių DBVS privalumai ir trūkumai .....	111
9 priedas. Sukurtos sistemos realizacijos vartotojo sąsajos pavyzdžiai .....	112

## Galimi kredito rizikos vertinimo kintamieji ir kriterijai

**Bendrovės būklės vertinimas (Dzidzevičiūtė, 2005)**

<b>Vertinimo kriterijus</b>	<b>Vertinimas</b>
<b>Veiklos efektyvumas</b>	
	Prekių (paslaugų) vertinimas
	Naudojamų technologijų charakteristikos
	Realizavimo tinklas, produktų ir rinkos plėtros strategija
	Finansinės prognozės, planavimo lygis
	Personalas, darbuotojų kvalifikacija
	Įmonės darbuotojai, personalas
<b>Produkcijos konkurencingumas</b>	
	Įmonės užimama rinkos dalis
	Prekių (paslaugų) paklausa, konkurencingumo lygis
	Realizavimo galimybės
	Prekės (paslaugos) gyvavimo ciklo fazė
<b>Vadovavimo profesionalumas, kvalifikacija</b>	
	Vadovo vadovavimo patirtis
	Darbo laikas šioje konkrečioje įmonėje
	Įmonės strategijos aiškumas
	Informacijos patikimumas
<b>Apskaitos kokybė</b>	
	Audituotos finansinės ataskaitos
	Audito išvadose esančių pastabų vertinimas
	Bendrovės apskaitos politikos vertinimas
	Informacijos apie dukterines įmones atskleidimas
<b>Įmonės verslo planas</b>	
	Investicijų atsipirkimo periodas
	Grynoji dabartinė vertė
	Investicijų pelningumas
	Alternatyviųjų investicijų grąža
<b>Bendrovės rizikų įvertinimas</b>	
	Papildomų finansavimo šaltinių pritraukimo galimybės
	Savininkų/ vadovybės konfliktų galimybės
	Gamybinės ir technologinės veiklos rizikos
	Kitos rizikos
<b>Stabilumas</b>	
	Pirkėjų ir tiekėjų stabilumas
	Veiklos diversifikacija
	Sąnaudų lankstumas
<b>Įmonės finansinė būklė</b>	
	Pajamų augimo tempas, pajamų augimo standartinis nuokrypis
	Paskolos vieta kliento balanso struktūroje, jos gražinimo galimybės
	Finansinių ataskaitų prognozės
	Pelno augimo tempas, pelno augimo tempo svyravimai
	Išlaidų struktūra
	Bendrasis, veiklos pelningumas
	Pinigų srautų analizė
	Apyvartumo rodikliai

## Šakos būklės vertinimas (Dzidzevičiūtė, 2005)



## Svarbiausi verslo šakos vertinimo kriterijai (Dzidzevičiūtė, 2005)

<b>Įmonės tipas</b>	<b>Svarbiausi rodikliai</b>
<b>Apdirbamosios pramonės įmonės</b>	Finansų struktūros rodikliai, prekybos skolų apyvartumas, turto vertės kitimas
<b>Statybos įmonės</b>	Prekybos skolų bei atsargų apyvartumas
<b>Žuvininkystės įmonės</b>	Bendrojo pelningumo rodiklis, likvidumo, finansų struktūros, atsargų, gautinų sumų ir grynojo apyvartinio kapitalo apyvartumo rodikliai
<b>Miškininkystės įmonės</b>	ROA, ROE, veiklos, grynasis pelningumas, nepaskirstyto pelno ir turto santykis, gamybos išlaidų ir pardavimų santykis, finansų struktūros, apyvartumo rodikliai
<b>Elektros, dujų ir vandens tiekimo sektorius</b>	Likvidumo, finansų struktūros, apyvartumo rodikliai, ROA, ROE
<b>Didmeninės ir mažmeninės prekybos įmonės</b>	Bendrosios finansų struktūros, likvidumo, pelningumo rodikliai
<b>Viešbučių ir restoranų įmonės</b>	Išlaidų rodikliai, finansų struktūros, likvidumo, veiklos pelno bei bendrojo pelno ir turto santykiai
<b>Transporto ir sandėliavimo įmonės</b>	Išlaidų rodikliai, apyvartumo rodikliai, veiklos pelningumo rodikliui
<b>Pašto ir telekomunikacijų įmonės</b>	ROA, finansų struktūros ir likvidumo rodikliai
<b>Nekilnojamo turto įmonės</b>	ROA, bendrojo pelno ir turto santykis, finansų struktūros, likvidumo rodikliai, grynojo apyvartinio kapitalo apyvartumas
<b>Švietimo sektoriaus įmonės</b>	Apyvartumo rodikliai, nepaskirstytojo pelno ir turto santykis, likvidumas, bendrasis pelningumas
<b>Sveikatos priežiūros ir socialinio darbo įmonės</b>	ROA, finansų struktūros, likvidumo, grynojo pelningumo, apyvartumo rodikliai

### Įmonės bankroto vertinimo kriterijai (K. Garškaitė, A. Garškienė)

<b>Įmonės stebėjimo objektai</b>	<b>Absolūtieji rodikliai</b>	<b>Santykiniai rodikliai</b>	<b>Lengva finansinė krizė</b>	<b>Gili finansinė krizė</b>	<b>Finansinė katastrofa</b>
<b>1. Grynieji pinigų srautai</b>	A. Visų grynųjų pinigų srautų suma B. Atskirų operacijų grynieji piniginiai srautai	A. Grynųjų pinigų srautų įplaukos / išplaukos B. Grynųjų pinigų srautų suma / įsipareigojimai	Grynųjų pinigų srautų likvidumas mažėja	Neigiama grynųjų pinigų srautų vertė	Itin didelė neigiama grynųjų pinigų srautų reikšmė
<b>2. Rinkos vertė</b>	Įmonės rinkos vertė	Turto apyvartumo koeficientas, naudojant turto rinkos vertę	Įmonės rinkos vertė stabilizuojasi	Įmonės rinkos vertė mažėja	Įmonės rinkos vertė sparčiai mažėja
<b>3. Kapitalo struktūra</b>	A. Nuosavas kapitalas B. Skolintas kapitalas	A. Finansinės nepriklausomybės koeficientas B. Nuosavas kapitalas/skolintas C. Ilgalaiškės finansinės nepriklausomybės koeficientas	Finansinės nepriklausomybės koeficientas mažėja	Finansinio sverto koeficientas didėja, o jo efektyvumas mažėja	Labai didelis finansinio sverto koeficientas
<b>4. Sugebėjimas padengti įsipareigojimus</b>	A. Ilgalaikiai įsipareigojimai B. Trumpalaikiai įsipareigojimai C. Finansinis kreditas D. Prekinis (komercinis) kreditas E. Kreditinio įsiskolinimo suma	A. Ilgalaikiai įsipareigojimai/trumpalaikiai B. Finansinis kreditas/prekinis C. Finansiniai įsipareigojimai, kurių atidėti negalima/visi įsipareigojimai D. Kreditorinių įsiskolinimų apyvartos periodas	Trumpalaikiai įsipareigojimai ir jų dalis visuose įsipareigojimuose didėja	Didelis finansinių įsipareigojimų, kurių atidėti negalima, ir visų įsipareigojimų santykio koeficientas	Itin didelis pastarasis koeficientas

Įmonės stebėjimo objektai	Absolūtieji rodikliai	Santykiniai rodikliai	Lengva finansinė krizė	Gili finansinė krizė	Finansinė katastrofa
<b>5.Turto sudėtis</b>	A.Neapyvartinis turtas B.Apyvartinis turtas C.Debitorinis įsiskolinimas (įskaitant uždelstus) D.Piniginės lėšos	A.Turto manevringumas B.Apsirūpinimo didelio likvidumo aktyvais koeficientas C.Einamasis mokumas D.Absolūtus mokumas E.Debitorinis įsiskolinimas/kreditorinis F.Debitorinio įsiskolinimo apyvartos periodas	Absoliutaus mokumo koeficientas mažėja	Absoliutaus ir einamojo mokumo koeficientai sparčiai mažėja	Absolūtus nemokumas, kadangi nėra piniginių aktyvų
<b>6.Einamųjų išlaidų sudėtis</b>	A.Visos einamosios išlaidos B.Pastovios einamosios išlaidos	A. Einamųjų išlaidų suma/realizuota produkcija B.Kintamosios einamosios išlaidos/ realizuota produkcija C.Visos išlaidos/pajamos	Kintamosios išlaidos didėja	Didelis kintamųjų išlaidų ir pajamų koeficientas	Itin didelis visų išlaidų ir pajamų koeficientas
<b>7.Finansinių operacijų koncentracija padidintos rizikos zonose</b>	A.Kapitalo dalis, įdėta kritinės rizikos zonose B.Kapitalo dalis, įdėta katastrofinės rizikos zonose	Didėja kapitalo dalis, įdėta krizinės rizikos zonoje	Didėja kapitalo dalis, įdėta krizinės rizikos zonose	Itin didėja kapitalo dalis, įdėta katastrofinės rizikos zonoje	

Šaltinis: K. Garškaitė, A. Garškienė. Įmonių bankroto diagnostikos sistema

### Standard & Poor's vertinimo kriterijai

Vertinimo kriterijus	Vertinimas
<b>Pramonės šakos rizika – pramonės šakos stiprumas ekonomikos kontekste</b>	
	Ekonomikos ciklo svarba;
	Verslo šakos cikliškumas – trukmė, pajamų svyravimas;
	Įvairių ekonomikos jėgų įtaka
	Socialiniai ir politiniai veiksniai
	Žaliavų kainų kaštai ir prieinamumas
	Tarptautinė konkurencinė pozicija
	Infliacijos
	Paklausos veiksniai – augimas lyginant su BVP, rinkos brandumas;
	Pagrindinės finansinės charakteristikos – kapitalo intensyvumas, kredito panaudojimas;
	Pasiūlos veiksniai – žaliavos, darbas, pajėgumų išnaudojimas;
	Teisinis reguliavimas;
	Verslo koncentracija;
	Įžengimo barjerai.
<b>Įmonės pozicija pramonės šakoje – rinka.</b>	
	Įmonės pardavimų pozicija pagrindinėse jos veiklos srityse ir galimybė apsaugoti turimą poziciją ateityje
	Sugebėjimas generuoti pardavimus;
	Dominuojančios ir stabilios rinkos dalys;
	Rinkodaros ir paskirstymo poreikiai – stiprumas, silpnumas, vidaus ir tarptautiniai;
	Tyrimai ir plėtra, jo svarba, produkto gyvavimo ciklas;
	Aptarnavimo/aprūpinimo organizavimas;
	Priklausomybė nuo pagrindinių klientų;
	Ilgalaikiai pardavimo kontraktai;
	Produkto diversifikacija.



<b>Vertinimo kriterijus</b>	<b>Vertinimas</b>
<b>Firmos pozicija pramonės šakoje – veiklos efektyvumas</b>	
	Istorinės veiklos maržos ir sugebėjimas jas išlaikyti;
	Sugebėjimas išlaikyti ir pagerinti pelno maržas;
	Kainų lyderiavimas;
	Gamybos operacijų integralumas;
	Pastatai ir įranga – nauja ir moderni ar sena, aukštų ar žemų kaštų;
	Žaliavų tiekimas;
	Kapitalo ir darbuotojų produktyvumo lygis;
	Darbo jėgos kaštai, santykiai su profsajungomis;
	Aplinkosaugos reikalavimai ir jų įtaka kaštams;
	Energetiniai kaštai.
<b>Vadovybės įvertinimas</b>	
	Veiklos ir finansinių rezultatų pasiekimai per laiką;
	Planavimas, tiek strateginis, tiek finansinis, augimo planas;
	Kontrolė – valdymo, finansinis ir vidinis auditas;
	Finansavimo politika ir praktika;
	Bendra valdymo kokybė;
	Susijungimų planai.
<b>Apskaitos įvertinimas</b>	
	Auditorių kvalifikacija;
	LIFO ar FIFO atsargų metodas;
	Prestižas ir nematerialus turtas;
	Pajamų pripažinimas;
	Nusidėvėjimo apskaita;
	Nekonsoliduojamos dukterinės firmos;
	Apskaitos metodai;
	Pensijų atidėjimai;
	Nepakankamai įvertintas turtas, pavyzdžiui, LIFO rezervas
<b>Pajamų apsauga – ilgalaikė pajamų uždirbimo galia</b>	
	Kapitalo pelningumas;
	Ikimokestinio padengimo koeficientai;
	Pelno maržos;
	Turto pelningumas pagal segmentus;
	Ateities pajamų augimo šaltiniai;
	Pajėgumas finansuoti augimą iš vidaus
<b>Finansinis svertas ir turto apsauga</b>	
	Ilgalaikė skola ir skolos/kapitalo santykis;
	Privilegiuotos akcijos;
	Svertas nebalansiniuose straipsniuose;
	Turto kilmė;
	Apyvartinio kapitalo valdymas
<b>Pinigų srautai</b>	
	Kapitalo poreikis;
	Pinigų srautų nepastovumas;
	Laisvas pinigų srautas;
	Pinigų srautas skolos aptarnavimui

### Šalies vertinimo kriterijai (Standard & Poor's)

- santaupos ir investicijos kaip procentas nuo BVP (matuojamas santaupų-investicijų normų skirtumas);

- užsienio valiutos skola metų gale kaip procentas nuo eksporto;
- biudžeto deficitas, kaip procentas nuo BVP;
- realus BVP augimas;
- BVP defliatorius, matuojantis infliaciją;
- nedarbas;
- investicijų procentas nuo BVP;
- trumpalaikės palūkanų normos, lyginamos su JAV 3 mėn. išdo vekselių norma;
- skola užsienio valiuta, milijardais USD;
- einamosios sąskaitos balansas/BVP;
- eksporto augimas;
- prekybos balansas/BVP.

## Darbe naudojamų diskriminantinių rizikos vertinimo modelių sudėtis ir išraiškos

Modelio pavadinimas, autorius	Modelį sudarantys kintamieji, rodikliai	Matematinė išraiška	Įverčių reikšmės
<b>Altmano Z-Score (Altman, 1968)</b>	Apyvartinis kapitalas, turtas, nepaskirstytas pelnas, EBIT, nuosavybės rinkos vertė, visi įsipareigojimai, pardavimai	$Z = 1,2 * (\text{apyvartinis kapitalas} / \text{turtas}) + 1,4 * (\text{nepaskirstytas pelnas} / \text{turtas}) + 3,3 * (\text{EBIT} / \text{turtas}) + 0,6 * (\text{nuosavybės rinkos vertė} / \text{visi įsipareigojimai}) + 0,999 * (\text{pardavimai} / \text{turtas})$	$Z < 1,21$ - tai didelė bankroto tikimybė $1,21 \leq Z \leq 2,9$ - neapibrėžtumo zona $Z > 2,9$ - maža bankroto bankroto tikimybė
<b>Altmano vertinimo balais modelis įmonėms, kurių akcijomis nėra prekiaujama biržoje (Altman E.I., 2000, Altman E.I. 2002)</b>	Apyvartinis kapitalas, turtas, nepaskirstytas pelnas, EBIT, nuosavybės buhalterinė vertė, turtas, pardavimai	$Z = 0,717 * (\text{apyvartinis kapitalas} / \text{turtas}) + 0,847 * (\text{nepaskirstytas pelnas} / \text{turtas}) + 3,107 * (\text{EBIT} / \text{turtas}) + 0,420 * (\text{nuosavybės buhalterinė vertė} / \text{visi įsipareigojimai}) + 0,998 * (\text{pardavimai} / \text{turtas})$	
<b>Altmano modelis negamybinėms įmonėms</b>	Apyvartinis kapitalas, turtas, nepaskirstytas pelnas, EBIT, savininkų nuosavybė, visi įsipareigojimai	$Z = 6,56 * (\text{apyvartinis kapitalas} / \text{turtas}) + 3,26 * (\text{nepaskirstytas pelnas} / \text{turtas}) + 6,72 * (\text{EBIT} / \text{turtas}) + 1,05 * (\text{savininkų nuosavybė} / \text{visi įsipareigojimai})$	
<b>Altmano modelis kylančioms įmonėms (ZETA<sup>R</sup>) (Altman, 1977)</b>	Pajamų stabilumas, ROA, skolos aptarnavimas, besikaupiantis pelningumas, likvidumas, kapitalizacija, firmos turto dydis		
<b>Springate modelis (Springate, 1978)</b>	Apyvartinis kapitalas, turtas, EBIT, trumpalaikiai įsipareigojimai, pardavimai	$Z'' = 1,03 * (\text{apyvartinis kapitalas} / \text{turtas}) + 3,07 * (\text{EBIT} / \text{turtas}) + 0,66 * (\text{EBIT} / \text{trumpalaikiai įsipareigojimai}) + 0,4 * (\text{pardavimai} / \text{turtas})$	$Z'' < 0,862$ – įmonė yra bankroto zonoje
<b>Fulmer modelis</b>	Nepaskirstytas pelnas, turtas, pardavimai, EBIT, savininkų nuosavybė, grynujų pinigų srautas, įsipareigojimai, trumpalaikiai įsipareigojimai, materialus turtas, apyvartinis kapitalas, mokamos palūkanos	$H = 5,528 * (\text{nep. pelnas} / \text{turtas}) + 0,212 * (\text{pardavimai} / \text{turtas}) + 0,073 * (\text{EBIT} / \text{savininkų nuosavybė}) + 1,270 * (\text{grynujų pinigų srautas} / \text{įsipareigojimai}) - 0,12 * (\text{įsipareigojimai} / \text{turtas}) + 2,335 * (\text{trumpalaikiai įsipareigojimai} / \text{turtas}) + 0,575 * (\log(\text{materialus turtas})) + 1,083 * (\text{apyvartinis kapitalas} / \text{įsipareigojimai}) + 0,894 * (\log(\text{EBIT} / \text{mokamos palūkanos})) - 6,075$	$H < 0$ – įmonė yra bankroto zonoje
<b>Shumway modelis</b>	Grynasis pelnas, turtas, įsipareigojimai, trumpalaikis turtas, trumpalaikiai įsipareigojimai	$H = -7,811 - 6,307 * (\text{grynasis pelnas} / \text{turtas}) + 4,068 * (\text{įsipareigojimai} / \text{turtas}) - 0,158 * (\text{trumpalaikis turtas} / \text{trumpalaikiai įsipareigojimai})$	
<b>Zmijewski modelis (Zmijewski,</b>	Grynasis pelnas, turtas, įsipareigojimai, trum-	$Z = -4,336 - 4,513 * (\text{grynasis pelnas} / \text{turtas}) + 5,679 * (\text{įsipareigoji-}$	$Z < 0$ – įmonė yra bankroto zonoje

Modelio pavadinimas, autorius	Modelį sudarantys kintamieji, rodikliai	Matematinė išraiška	Įverčių reikšmės
<b>1984)</b>	palaikis turtas, trumpalaikiai įsipareigojimai	mai/turtas) + 0,004 * (trumpalaikis turtas/ trumpalaikiai įsipareigojimai)	
<b>Ohlson modelis</b>	Turtas, infliacijos lygis, įsipareigojimai, apyvartinis kapitalas, trumpalaikiai įsipareigojimai, grynasis pelnas, pinigų srautai iš pagrindinės veiklos, bendras indeksas	$Y = -1,3 - 0,4 * (\text{turtas} / \text{infliacijos lygis}) + 6 * (\text{įsipareigojimai} / \text{turtas}) - 1,4 * (\text{apyvartinis kapitalas} / \text{turtas} + 0,1 * (\text{trumpalaikiai įsipareigojimai} / \text{turtas}) - 2,4 * (1, \text{ jei trumpalaikiai įsipareigojimai} > \text{turtas}; 0 \text{ kitu atveju}) - 1,8 * (\text{grynasis pelnas} / \text{turtas}) + 0,3 * (\text{pinigų srautai iš pagrindinės veiklos}) - 1,7 * (1, \text{ jei pelnas} < 0 \text{ du metus iš eilės}) - 0,5 * (\text{bendras indeksas})$	
<b>S. Grigaravičiaus modelis (Grigaravičius, 2003)</b>	Trumpalaikis turtas, trumpalaikiai įsipareigojimai, grynasis apyvartinis kapitalas, turtas, nuosavybė, skola, veiklos pelnas, palūkanų išlaidos, EBIT, ROA, grynojo apyvartinio kapitalo apyvartumas	$Z = -0,762 + 0,003 * (\text{trumpalaikis turtas} / \text{trumpalaikiai įsipareigojimai}) - 0,424 * (\text{grynasis apyvartinis kapitalas} / \text{turtas}) - 0,06 * (\text{turtas} / \text{nuosavybė}) + 0,22 * (\text{nuosavybė} / \text{skola}) - 0,774 * (\text{veiklos pelnas} / \text{palūkanų išlaidos}) - 0,189 * (\text{EBIT} / \text{turtas}) + 6,842 * (\text{ROA}) - 12,262 * (\text{grynojo apyvartinio kapitalo apyvartumas})$	

Šaltinis: sudaryta autoriaus pagal (Dzidzevičiūtė, 2005)

## Neuroninių tinklų taikymas bankroto prognozavime bei kredito rizikos vertinime 1992-1998 m.

Reference	NN model	Validation	Data size	Benchmark methods	MC
<i>Bankruptcy prediction</i>					
<i>Banks</i>					
Tam and Xiang, 1992	BPGD	TR-TS/LOO	202s/19v	LDA,QDA,LogR,kNN,ID3	1
Salchenberger et al., 1992	BPGD	TR-TS	458-604s/29dt5v	Log	1,6
Martin-del-Brio and Serrano-Cinca, 1993	SOFM (um)	NA	66s/9v	No	7
Markham and Ragsdale, 1995	BPGD	TR-TS	202s/19v	LDA	5
Serrano-Cinca, 1996	SOFM (um)	NA	66s/9v	No	5,7
Olmeda and Fernandez, 1997	BPGD	TR-TS/6FCV	66-129s/9-5v	DA,Log,MARS,C4.5	1,5
Serrano-Cinca, 1997	BPGD	LOO	66s/9v	LDA,LogR	1
<i>Firms/general</i>					
Fletcher and Goss, 1993	BPGD	18FCV	36s/3v	Log	1
Udo, 1993	BPGD	TR-TS	400s/16v	MLR	1,6
Altman et al., 1994	BPGD	TR-TS	1108s/NA	LDA	2,6,8
Wilson and Sharda, 1994	BPGD	MCRT	129s/5v	LDA	1,6
Chen et al., 1995	SOFM (um)	NA	46s/4v	(1)	1,7
Boritz and Kennedy, 1995	BPGD/OET	TR-TS	342s/VARv	LDA,Log,Prob	2,4
Boritz et al., 1995	BPGD/OET	CV	VARs/VARv	LDA,QDA,kNN,Log,Prob	2
Back, Latinen, and Sere, 1996	BPGD	TR-TS	74s/31dt2-15v	LDA,Log	1,4
Lee et al., 1996	BPGD	TR-TS	166s/57dt7-18v	No	4,5
Ignizio and Soltys, 1996	BPGD/ONN	TR-TS	129s/5v	LDA	4,5,7
Jo and Han, 1996	BPGD	TR-VA-TS	542s/20v	LDA,CBFS	1,5
Lee et al., 1996	BPGD	TR-TS	88s/70dt5-41v	LogR,LDA	1
Jo, Han and Lee, 1997	BPGD	TR-VA-TS	542s/61dt9-23v	LDA,CBFS	1
Morris, 1997	NA	NA	NA	NA	NA
van Bussel and Veelenurf, 1997	BPGD	TR-TS	356s/20v(dr)	No	4,5
Pirramuthu, Ravagan, and Shaw, 1998	BPGD	TR-TS	48-118-182s/4-19v	No	5
Kiviluoto, 1998	SOFM, RBF-SOFM,LVQ	5FCV	1137s/4v	LDA, QDA, k-NN	2,4,8
<i>Credit</i>					
<i>Credit card performance prediction</i>					
Jagielska and Jaworski, 1996	BPGD	TR-VA-TS	2430-6000s/23v	(*)	1,8
<i>Credit evaluation/consumers</i>					
Jensen, 1992	BPGD	TR-TS	125s/24v	(*)	2
Robins, 1993b	NA	NA	NA	NA	NA
Richeson, Zimmermann, and Barnett, 1994	BPGD	TR-TS	300s/7v	LDA	2,6
Desay et al., 1996	BPGD-MNN	CV	1962s/18v	LDA,LogR	1,5,7,8
Torsun, 1996	BPGD	TR-TS	310 000s/18v	GMLC,FuzC,(*)	1,8
Arminger, Enache, and Bonne, 1997	BPGD	TR-VA-TS	8163s/6v	LogD,CART	3,4
Williamson, 1995	BPGD	TR-TS	2403s/27dt20v	No	4,5
<i>Credit evaluation/firms</i>					
Deng, 1993	BPGD	NA	NA	NA	7
Pirramuthu, Shaw and Gentry, 1994	BPGD-NR	TR-TS	32-100s/18-15v	Prob,ID3,NEWQ	1,5
Glorfeld and Hardgrave, 1996	BPGD/ADA(N)PAWR	10FCV	40s/12dt5v	No	1
Glorfeld, 1996	BPGD	10FCV	42s/17dt5v	LDA,LogR,CART	7
<i>General</i>					
Hand and Henley, 1997	NA	NA	NA	NA	NA
Borowsky, 1995	NA	NA	NA	NA	NA

Šaltinis: A. Vellido, P.J.G. Lisboa, J. Vaughan. Neural networks in business: a survey of applications

Fuzzy logikos operacijos ir išvadų darymo metodai

<i>p</i>	<i>q</i>	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

$\vee$	0	1	<i>q</i>
0	0	1	
1	1	1	
<i>p</i>			

Bulio logika ir jos operacija or

Or  
 $p \vee q$

0	0.5	1
0.5	0.5	1
1	1	1

Nor  
 $\neg(p \vee q)$

1	0.5	0
0.5	0.5	0
0	0	0

Implication  
 $(p \leq q) \vee q$

1	1	1
0	1	1
0	0.5	1

Equivalence  
 $(p \Rightarrow q) \wedge (q \Rightarrow p)$

1	0	0
0	1	0.5
0	0.5	1

Nand  
 $(\neg p) \vee (\neg q)$

1	1	1
1	0.5	0.5
1	0.5	0

And  
 $\neg((\neg p) \vee (\neg q))$

0	0	0
0	0.5	0.5
0	0.5	1

Fuzzy logika ir jos pagrindinės operacijos

Šaltinis: Jan Jantzen. Tutorial On Fuzzy Logic

Operacijos su fuzzy aibėmis

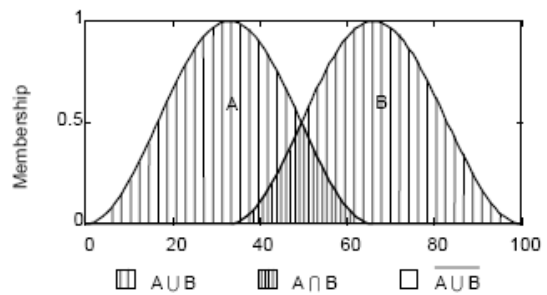
- A ir B sankirta (intersection):

$A \cap B = A \min B$

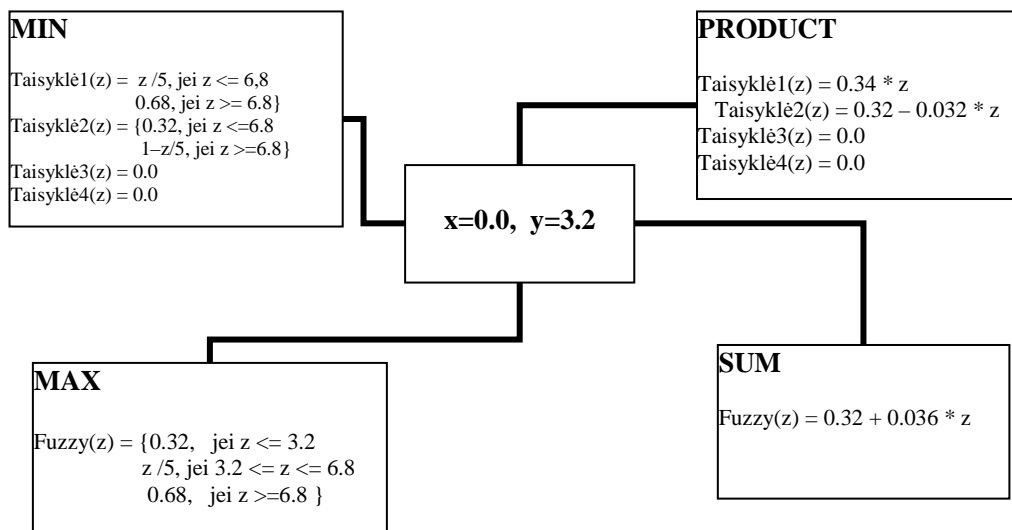
- A ir B sąjunga (union):

$A \cup B = A \max B$

- A papildymas (complement):



Šaltinis: Jan Jantzen. Tutorial On Fuzzy Logic



Šaltinis: sudaryta autoriaus.

## Ekspertinių taisyklių realizacija CLIPS/JESS ekspertinėse aplinkose

```

;-----
;Pramoninis pelningumas A1
;-----
(defrule a1rez1
  (A1 ?x &: (<= ?x 10)) => (assert (pram_pelningumas nepa-
tenkinamas))
)
(defrule a1rez2
  (A1 ?x &: (> ?x 10) &: (<= ?x 20)) => (assert
(pram_pelningumas vidutinis))
)
(defrule a1rez3
  (A1 ?x &: (> ?x 20) &: (<= ?x 30)) =>(assert
(pram_pelningumas patenkinamas))
)
(defrule a1rez4
  (A1 ?x &: (> ?x 30)) => (assert (pram_pelningumas "labai
geras"))
)
;-----
;Finansinis pelningumas A2
;-----
(defrule a2rez1
  (A2 ?x &: (<= ?x 17.5)) => (assert (fin_pelningumas nepa-
tenkinamas))
)
(defrule a2rez2
  (A2 ?x &: (> ?x 17.5) &: (<= ?x 20)) => (assert
(fin_pelningumas vidutinis))
)
(defrule a2rez3
  (A2 ?x &: (> ?x 20) &: (<= ?x 23)) =>(assert
(fin_pelningumas patenkinamas))
)
(defrule a2rez4
  (A2 ?x &: (> ?x 23)) => (assert (fin_pelningumas "labai
geras"))
)
;-----
;Bendrasis pelningumas/bendrieji aktyvai A3
;gross profit/gross assets
;-----
(defrule a3rez1
  (A3 ?x &: (<= ?x 0)) => (assert (bendras_pelningumas nepa-
tenkinamas))
)
(defrule a3rez2
  (A3 ?x &: (> ?x 0) &: (<= ?x 50)) => (assert (ben-
dras_pelningumas vidutinis))
)
(defrule a3rez3
  (A3 ?x &: (> ?x 50) &: (<= ?x 75)) =>(assert (ben-
dras_pelningumas patenkinamas))
)
(defrule a3rez4
  (A3 ?x &: (> ?x 75)) => (assert (bendras_pelningumas "labai
geras"))
)
;-----
;Pelno marža (profit margin) A4
;-----
(defrule a4rez1
  (A4 ?x &: (<= ?x 0)) => (assert (marza nepatenkinamas))
)
)
(defrule a4rez2
  (A4 ?x &: (> ?x 0) &: (<= ?x 50)) => (assert (marza viduti-
nis))
)
(defrule a4rez3
  (A4 ?x &: (> ?x 50) &: (<= ?x 100)) =>(assert (marza paten-
kinamas))
)
(defrule a4rez4
  (A4 ?x &: (> ?x 100)) => (assert (marza "labai geras"))
)
;-----
;Trumpalaikiu skolu galimybes B1
;-----
(defrule b1rez1
  (B1 ?x &: (<= ?x 25)) => (assert (short_debt nepatenkinamas))
)
(defrule b1rez2
  (B1 ?x &: (> ?x 25) &: (<= ?x 50)) => (assert (short_debt
vidutinis))
)
(defrule b1rez3
  (B1 ?x &: (> ?x 50) &: (<= ?x 75)) =>(assert (short_debt
patenkinamas))
)
(defrule b1rez4
  (B1 ?x &: (> ?x 75) &: (<= ?x 100)) => (assert (short_debt
"labai geras"))
)
;-----
;Bendros skolu galimybes B2
;-----
(defrule b2rez1
  (B2 ?x &: (>= ?x 80)) => (assert (global_debt nepatenkina-
mas))
)
(defrule b2rez2
  (B2 ?x &: (> ?x 60) &: (<= ?x 80)) => (assert (global_debt
vidutinis))
)
(defrule b2rez3
  (B2 ?x &: (> ?x 40) &: (<= ?x 60)) =>(assert (global_debt
patenkinamas))
)
(defrule b2rez4
  (B2 ?x &: (<= ?x 40)) => (assert (global_debt "labai geras"))
)
;-----
;Ilgalaikiu skolu galimybes B3
;-----
(defrule b3rez1
  (B3 ?x &: (> ?x 0.5)) => (assert (long_debt nepatenkinamas))
)
(defrule b3rez2
  (B3 ?x &: (<= ?x 0.5)) => (assert (long_debt patenkinamas))
)
;-----
;Likvidumas B4
;-----
(defrule b4rez1
  (B4 ?x &: (<= ?x 2)) => (assert (likvidumas nepatenkinamas))
)
)

```

```

(defrule b4rez2
  (B4 ?x &: (> ?x 2)) => (assert (likvidumas patenkinamas))
)
;-----
;Tiesioginis likvidumas B5
;-----
(defrule b5rez1
  (B5 ?x &: (<= ?x 1)) => (assert (ties_likvidumas nepatenkinamas))
)
(defrule b5rez2
  (B5 ?x &: (> ?x 1) &: (<= ?x 1.5)) => (assert
(ties_likvidumas patenkinamas))
)
(defrule b5rez3
  (B5 ?x &: (>= ?x 1.5)) => (assert (ties_likvidumas "labai
geras"))
)
;-----
;Finansines islaidos C1
;-----
(defrule c1rez1
  (C1 ?x &: (>= ?x 5)) => (assert (fin_islaidos nepatenkinamas))
)
(defrule c1rez2
  (C1 ?x &: (> ?x 3) &: (<= ?x 5)) => (assert (fin_islaidos
vidutinis))
)
(defrule c1rez3
  (C1 ?x &: (> ?x 2) &: (<= ?x 3)) => (assert (fin_islaidos pa-
tenkinamas))
)
(defrule c1rez4
  (C1 ?x &: (< ?x 2)) => (assert (fin_islaidos "labai geras"))
)
;-----
;Bendru ir administraciniu islaidu pasikeitimas C2
;-----
(defrule c2rez1
  (C2 ?x &: (>= ?x 8)) => (assert (bendros_islaidos nepatenki-
namas))
)
(defrule c2rez2
  (C2 ?x &: (> ?x 6) &: (<= ?x 8)) => (assert (bendros_islaidos
vidutinis))
)
(defrule c2rez3
  (C2 ?x &: (> ?x 4) &: (<= ?x 6)) => (assert (bendros_islaidos
patenkinamas))
)
(defrule c2rez4
  (C2 ?x &: (> ?x 2) &: (<= ?x 4)) => (assert (bendros_islaidos
"labai geras"))
)
(defrule c2rez5
  (C2 ?x &: (< ?x 2)) => (assert (bendros_islaidos puikus))
)
;-----
;Prekiu atsargu kitimas C3
;-----
(defrule c3rez1
  (C3 didejantis) => (assert (prekiu_atpargos nepatenkinamas))
)
(defrule c3rez2
  (C3 nedidejantis) => (assert (prekiu_atpargos patenkinamas))
)
;-----
;Vadovo darbo patirtis Q1
;-----
(defrule q1rez1
  (Q1 0) => (assert (patirtis nepatenkinamas))
)
(defrule q1rez2
  (Q1 1) => (assert (patirtis vidutinis))
)
(defrule q1rez3
  (Q1 2) => (assert (patirtis patenkinamas))
)
(defrule q1rez4
  (Q1 3) => (assert (patirtis "labai geras"))
)
(defrule q1rez5
  (Q1 4) => (assert (patirtis puikus))
)
;-----
;Imones rinkos padetis Q2
;-----
(defrule q2rez1
  (Q2 0) => (assert (rinkos_padetis nepatenkinamas))
)
(defrule q2rez2
  (Q2 1) => (assert (rinkos_padetis vidutinis))
)
(defrule q2rez3
  (Q2 2) => (assert (rinkos_padetis patenkinamas))
)
(defrule q2rez4
  (Q2 3) => (assert (rinkos_padetis "labai geras"))
)
(defrule q2rez5
  (Q2 4) => (assert (rinkos_padetis puikus))
)
;-----
;Irenginiu padetis Q3
;-----
(defrule q3rez1
  (Q3 0) => (assert (irengimai nepatenkinamas))
)
(defrule q3rez2
  (Q3 1) => (assert (irengimai vidutinis))
)
(defrule q3rez3
  (Q3 2) => (assert (irengimai patenkinamas))
)
(defrule q3rez4
  (Q3 3) => (assert (irengimai "labai geras"))
)
(defrule q3rez5
  (Q3 4) => (assert (irengimai puikus))
)
;-----
;Personalas Q4
;-----
(defrule q4rez1
  (Q4 0) => (assert (personalas nepatenkinamas))
)
(defrule q4rez2
  (Q4 1) => (assert (personalas vidutinis))
)
(defrule q4rez3
  (Q4 2) => (assert (personalas patenkinamas))
)
(defrule q4rez4

```



```

(Q4 3) => (assert (personalas "labai geras"))
)
(defrule q4rez5
(Q4 4) => (assert (personalas puikus))
)
;-----
;Konkurenciniai privalumai Q5
;-----
(defrule q5rez1
(Q5 0) => (assert (konkurencija nepatenkinamas))
)
(defrule q5rez2
(Q5 1) => (assert (konkurencija vidutinis))
)
(defrule q5rez3
(Q5 2) =>(assert (konkurencija patenkinamas))
)
(defrule q5rez4
(Q5 3) => (assert (konkurencija "labai geras"))
)
(defrule q5rez4
(Q5 4) => (assert (konkurencija puikus))
)
;-----
;Rinkos lankstumas Q6
;-----
(defrule q6rez1
(Q6 0) => (assert (rinka nepatenkinamas))
)
(defrule q6rez2
(Q6 1) => (assert (rinka vidutinis))
)
(defrule q6rez3
(Q6 2) =>(assert (rinka patenkinamas))
)
(defrule q6rez4
(Q6 3) => (assert (rinka "labai geras"))
)
(defrule q6rez5
(Q6 4) => (assert (rinka puikus))
)
;-----
;Ivertinimu skaiciavimas
; (tai tik pavyzdines taisykles; jos realiai turetu buti daug ilges-
nes ir apimti daugybe variantu)
;-----
(defrule kiekyb1
(pram_pelningumas "labai geras")(fin_pelningumas "labai
geras")
(bendras_pelningumas "labai geras")(marza "labai geras")
(short_debt "labai geras")(global_debt "labai geras")
(long_debt patenkinamas)(likvidumas patenkinamas)
(ties_likvidumas patenkinamas)
(fin_islaidos "labai geras")
(or (bendros_islaidos "labai geras")(bendros_islaidos pui-
kus))
(prekiu_atsargos patenkinamas)
=>
(assert (kiek_klase "labai geras")))
)
(defrule kiekyb2
(pram_pelningumas patenkinamas)(fin_pelningumas paten-
kinamas)
(bendras_pelningumas patenkinamas)(marza patenkinamas)
(short_debt patenkinamas)(global_debt patenkinamas)
(long_debt patenkinamas)(likvidumas patenkinamas)
(ties_likvidumas patenkinamas)
(fin_islaidos patenkinamas)
(bendros_islaidos patenkinamas)
(prekiu_atsargos patenkinamas)
=>
(assert (kiek_klase patenkinamas)))
)
(defrule kiekyb3
(pram_pelningumas vidutinis)(fin_pelningumas vidutinis)
(bendras_pelningumas vidutinis)(marza vidutinis)
(short_debt vidutinis)(global_debt vidutinis)
(long_debt patenkinamas)(likvidumas patenkinamas)
(ties_likvidumas patenkinamas)
(fin_islaidos vidutinis)
(bendros_islaidos vidutinis)
(prekiu_atsargos patenkinamas)
=>
(assert (kiek_klase vidutinis)))
)
(defrule kiekyb4
(pram_pelningumas nepatenkinamas)(fin_pelningumas nepa-
tenkinamas)
(bendras_pelningumas nepatenkinamas)(marza nepatenkina-
mas)
(short_debt nepatenkinamas)(global_debt nepatenkinamas)
(long_debt nepatenkinamas)(likvidumas nepatenkinamas)
(ties_likvidumas nepatenkinamas)(pram_pelningumas nepa-
tenkinamas)
(fin_islaidos nepatenkinamas)
(bendros_islaidos nepatenkinamas)
(prekiu_atsargos nepatenkinamas)
=>
(assert (kiek_klase nepatenkinamas)))
)
(defrule kokyb1
(patirtis "labai geras")(rinkos_padetis "labai geras")
(irengimai "labai geras")(personalas "labai geras")
(konkurencija "labai geras")(rinka "labai geras")
=>
(assert (kok_klase "labai geras")))
)
(defrule kokyb2
(patirtis puikus)(rinkos_padetis puikus)
(irengimai puikus)(personalas puikus)
(konkurencija puikus)(rinka puikus)
=>
(assert (kok_klase puikus)))
)
(defrule kokyb3
(patirtis nepatenkinamas)(rinkos_padetis nepatenkinamas)
(irengimai nepatenkinamas)(personalas nepatenkinamas)
(konkurencija nepatenkinamas)(rinka nepatenkinamas)
=>
(assert (kok_klase nepatenkinamas)))
)
(defrule kokyb4
(patirtis patenkinamas)(rinkos_padetis patenkinamas)
(irengimai patenkinamas)(personalas patenkinamas)
(konkurencija patenkinamas)(rinka patenkinamas)
=>
(assert (kok_klase patenkinamas)))
)
(defrule kokyb5
(patirtis vidutinis)(rinkos_padetis vidutinis)
(irengimai vidutinis)(personalas vidutinis)
(konkurencija vidutinis)(rinka vidutinis)
=>
(assert (kok_klase vidutinis)))
)

```

## Evoliucinių metodų strategijos ir algoritmai

Metodo pavadinimas	Aprašymas	Privalumai	Trūkumai
<i>Genetinių algoritmų variantai ir strategijos</i>			
<b>Binarinis GA</b>	Pasirinkus vertės funkciją bei kintamuosius bei GA parametrus, sugeneruojama pradinė populiacija. Po to randama kiekvienos chromosomos vertė (tinkamumo įvertis). Naudojant tam tikrus kryžminimosi ir mutacijos principus, pasirenkamos tam tikros chromosomos ir atliekamas jų kryžminimas ir mutacija. Po to atliekamas konvergencijos tikrinimas, t.y., jei pabaigos sąlyga tenkinama, algoritmo vykdymas nutraukiamas, ir gražinamas geriausias esamos populiacijos sprendimas; kitu atveju algoritmas kartojamas su naujai sugeneruotomis populiacijomis tol, kol tenkinama pabaigos sąlyga.	Tai pats paprasčiausias ir lengviausiai realizuojamas algoritmas	Binarinis kintamųjų atvaizdavimas riboja tikslumą, todėl binariniai GA nėra gerai tinkami tada, kai kintamieji yra tęstiniai, t.y., kai jų skaičiavimui reikalingas didelis tikslumas. Nėra ypatingai greitas tada, kai ieškomas minimumas lokaliai kvadratiname regione
<b>Tęstinis GA</b>	Algoritmas beveik toks pat, kaip ir binarinio GA, tačiau kintamieji atvaizduojami ne 0 ir 1 bitais, o realiaisiais norimo tikslumo skaičiais	Didesnis nei binarinio GA tikslumas Kintamųjų ir rezultatų saugojimui reikalinga mažiau vietos atmintyje, kadangi vietoje N bitų sveikojo skaičiaus naudojamas vienas realusis skaičius Greitesnis nei binarinis, kadangi chromosomos neturi būti išskoduojamos įvertinimui pagal vertės funkciją	
<b>Hibridinis GA</b>	Genetinis algoritmas randa optimalios reikšmės regioną, o lokalus optimizatorius ieško minimalios reikšmės pagal vieną iš šių principų: 1. Naudojamas GA tol, kol jis sulėtėja, tada pradedamas naudoti lokalus optimizatorius. Tikimasi, kad GA tuo metu yra labai arti globalaus minimumo. 2. GA populiacijai suteikiant tam tikrą lokalų minimumą, rastą pagal atsitiktinius pradinius populiacijos taškus. 3. Per kiekvieną iš GA iteracijų naudojant lokalų optimizatorių geriausiam vienam ar keliems sprendimams ir pridėdant gautas chromosomas į naują populiaciją. Kaip lokalus optimizatorius gali būti naudojama svorinė vertės reikšmių	Genetinio algoritmo našumas apjungiamas su lokalinio optimizatoriaus greičiu	

Metodo pavadinimas	Aprašymas	Privalumai	Trūkumai
	funkcija ar Pareto optimizatorius		
„Netvarkingi“ GA (Messy GA, sutr. mGA)	Šis GA tipas turi skirtingo dydžio chromosomas bei nepriklausančius nuo pozicijos genus. Pirmos algoritmo fazės metu naudojama pradinė populiacija, turinti visus galimus kūrimo blokus apibrėžto ilgio populiacijai. Po to keletos populiacijų metu naudojamas tik kryžminimas, kiekvienos jų metu atmetant pusę populiacijų. Sugretinimo fazės metu naudojamos kitos genetinės operacijos, tokios, kaip pjaustymas ( <i>cutting</i> ) ir sujungimas ( <i>splicing</i> ), tol, kol randamas priimtinas sprendimas	Greitesnė konvergencija nei įprastuose GA Geri našumo rezultatai	Sudėtingumas
Paraleliniai GA	Taikomi, kai reikia naudoti sudėtingas vertės funkcijas, apimančias sudėtingas ir didelio techninio pajėgumo skaičiavimus. Pagal kilmę ir veikimo pobūdį gali būti išskirti keletas tipų: globalūs vienos populiacijos pavaldumo režimo GA ( <i>global single-population master-slave GA</i> ), vienos populiacijos smulkių vienetų ( <i>single-population fine-grained GA</i> ) ir daugybės populiacijų stambių subjektų ( <i>multiple-population coarse-grained GA</i> ) (Cantú-Paz, 1998)	Galimybė naudoti lygiagrečius skaičiavimus Didesnis skaičiavimų greitis, našumas Chromosomos gali būti išskirstytos į subpopuliacijas atskiruose kompiuteriuose, taip išbandant įvairias kombinacijas ir apsaugant nuo vieno ar kelių individų dominavimo Galimybė rasti kelis sprendimus	Sudėtinga realizacija Gali būti reikalingi dideli techniniai išteklių
<b>Kiti evoliuciniai algoritmai</b>			
Modeliuojamas atkaitinimas ( <i>Simulated annealing</i> )	Tai atsitiktiniu ieškojimu paremtas procesas, kurio idėja paimta iš skysčių užšalimo ir kristalizavimosi procesų. Šiuo atveju šildymas reiškia kintamųjų reikšmių atsitiktinį keitimą; kuo didesnis šilumos kitimas, tuo didesni atsitiktiniai svyravimai.	Algoritmas sėkmingai pritaikytas daugelio problemų sprendimui Geresnis našumas kombinuotų vertės funkcijų atžvilgiu, lyginant su lokaliais optimizatoriais Galimybė išvengti sustojimo lokaliame minimume Praktiniai rezultatai rodo, kad jau per baigtinį iteracijų skaičių pasiekiami neblogi rezultatai, o labai dažnai į globalinį minimumą papuolama gana greitai (Simutis, 2005)	Sudėtingas iteracijų skaičiaus skaičiavimas kiekvieno temperatūros kitimo atveju bei aukščiausios temperatūros nustatymas
Tabu paieška ( <i>Tabu search</i> )	Kaip ir modeliuojamo atkaitinimo algoritmas, pereina skersai visą sprendimų erdvę tikrindamas visus einamojo sprendimo „kaimynus“. Šiuo atveju buvę sprendimai įrašomi į „tabu“ sąrašą ir sprendimo eigoje daugiau nenagrinėjami.	Našesnis nei SA metodas, kadangi reikalingas mažesnis iteracijų skaičius	
Spiečiaus	Algoritmas pagrįstas paukščių būriavi-	Lengvai realizuojamas	

Metodo pavadinimas	Aprašymas	Privalumai	Trūkumai
<b>optimizacija (Particle swarm optimization)</b>	mosi principu. Kaip ir tęstinis GA, šis metodas pradedamas nuo pradinės populiacijos matricos inicializavimo, tačiau jis neturi kryžminimosi ir mutacijos operacijų. Matricos eilutės vadinamos dalelėmis; jose yra ne binariškai koduotos kintamųjų reikšmės. Kiekviena dalelė juda vertės paviršiumi greičiu, kuris, kaip ir dalelės pozicija, keičiama pagal geriausią lokalų ir globalų sprendimą. PSO algoritmas keičia greičių vektorių kiekvienai dalelei, po to prideda greitį prie jos pozicijos ar vertės. Jei geriausias lokalus sprendimas turi vertę, mažesnę nei globalus sprendimas tuo metu, geriausias globalus sprendimas keičiamas šia lokalaus sprendimo reikšme.	Mažai parametrų, kuriuos reikia tinkamai nustatyti Tinkamas spręsti uždavinius su sudėtingomis vertės funkcijomis, turinčiomis daug lokalių minimumų	
<b>Skruzdžių kolonijos optimizacija (Ant colony optimization)</b>	Šis metodas paremtas tuo, kad skruzdėlės randa trumpiausią kelią iki maisto palikdamos feromonų pėdsaką, kuriuo iki maisto seka ir kitos skruzdėlės. Skruzdėlė, randanti trumpiausią kelią, sukurs stipresnį feromonų pėdsaką, kuriuo seks kitos skruzdėlės.	Galimybė išspręsti sudėtingas problemas	Algoritmas efektyvus tik su lokaliu optimizatoriumi Pirmalaikė konvergencija į neoptimalų sprendimą

Šaltinis: sudaryta autoriaus pagal (Haupt R., Haupt, S.E. 2004), Simutis (2005)

### Genetinio algoritmo kryžminimosi ir mutacijos strategijos ir metodai

Metodo pavadinimas	Aprašymas	Privalumai	Trūkumai
<b>„Ruletės rato“ principas</b>	Apskaičiuojama visos populiacijos chromosomų tinkamumo įverčių (angl. <i>fitnesses</i> ) suma $S$ , po to iš intervalo $[0; S]$ išrenkamas atsitiktinis skaičius $r$ . Po to skaičiuojama populiacijos chromosomų tinkamumo įverčių suma $s$ iki tol, kol ji neviršija $r$ ; jai viršijus $r$ , gražinama tuo metu esanti chromosoma	Nesudėtingas realizavimas	Chromosomų įverčių sklaida, t.y., dideli skirtumai tarp reikšmių. Jei geriausios chromosomos rodiklis sudaro didelę dalį sumos, yra labai mažai galimybių, kad bus parinktos kitos chromosomos
<b>Parinkimas pagal reitingą (rank selection)</b>	Visos populiacijos chromosomos pirma „sureitinguojamos“ ir jos gauna įverčius pagal savo reitingą (blogiausios reitingas lygus 1, geriausios – chromosomų populiacijoje skaičiui)	Visos chromosomos gali būti parinktos	Lėtas artėjimas prie sprendimo (konvergencija)
<b>Ramios būsenos pasirinkimas (Steady-State Selection)</b>	Kiekvienoje generacijoje keletas „gerų“ (su aukštais tinkamumo įverčiais) chromosomų parenkamos naujos chromosomos formavimui. Tuomet	Didelė dalis chromosomų gali pereiti į kitą kartą, kas reiškia greitesnius skaičiavimus	Galimybė prarasti chromosomas su geriausiais įvertinimais

<b>Metodo pavadinimas</b>	<b>Aprašymas</b>	<b>Privalumai</b>	<b>Trūkumai</b>
	kai kurios chromosomos su žemais tinkamumo įverčiais pašalinamos ir naujai sukurta chromosoma įkeliama į jų vietą.		
<b>Elitizmas</b>	Kai kurios chromosomos su aukštais tinkamumo įverčiais pirma kopijuojamos į naują populiaciją, o likusi jos dalis kuriama aukščiau aprašytais metodais	Ženkliai didesnis genetinio algoritmo našumas, kadangi geriausias rastas sprendimas apsaugotas nuo praradimo	

Šaltinis: sudaryta autoriaus pagal (Obitko, 1998).

## Sistemoje naudojamų duomenų aprašymas ir duomenų bazės schema

<b>Ataskaita</b>	<b>Santrumpa</b>	<b>Aprašymas</b>
<b>Pelno ataskaita</b>	salesinc	Pajamos iš pardavimų Sales income
	otherinc	Kitos pajamos Other income
	grossinc	Bendrosios įplaukos Gross income
	costofg	Parduotų prekių ar paslaugų kaina Cost of goods sold
	randd	Tyrimai ir plėtra Research and development
	deprec	Nuvertėjimas Depreciation
	totalop	Bendros veiklos išlaidos Total operation expenses
	nonrecc	Neperiodinės lėšos balanse Nonrecurring items
	interest	Išlaidos palūkanoms Interest expenses
	totalint	Bendros išlaidos palūkanoms Total interest expenses
	grossop	Bendrosios veiklos pajamos Gross operating expenses
	unusual	Netikėtos pajamos Unusual income
	pretax	Pajamos prieš mokesčius Pre-tax income
	adjust	Pajamų patikslinimai Adjustments to income
	inctax	Pajamų mokestis Income tax
	netincome	Grynasis pelnas Net income
	sharesav	Vidutinė akcijos kaina Shares average
	eps	EPS EPS
	epscont	Dabartinis (continued) EPS Continued EPS
	epsdilut	EPS Diluted DilutedEPS
dividend	Dividendai Dividend	
<b>Balansas</b>	cash	Grynieji pinigai Cash
	shorti	Trumpalaikės investicijos Short-term investments
	receive	Debitorinės sąskaitos Receivables

<b>Ataskaita</b>	<b>Santrumpa</b>	<b>Aprašymas</b>
	inventory	Prekių atsargos Inventory
	assetcur	Trumpalaikis turtas Other Current Assets
	assettot	Bendras dabartinis turtas Total Current Assets
	netprop	Grynoji gamybinių fondų nuosavybė Net property plant and equipment
	longinv	Ilgalaikės investicijos Long-term investments
	assetlong	Kiti ilgalaikiai aktyvai Other long-term assets
	goodwill	Prestižo vertė ir nematerialiosios vertybės Goodwill and intangibles
	totassets	Bendrieji aktyvai Total assets
	accpay	Tiekėjų įsiskolinimas Accounts payable
	debshort	Trumpalaikiai įsiskolinimai Short-term debt
	liabcurr	Kiti dabartiniai įsipareigojimai Other current liabilities
	debtlong	Ilgalaikiai įsiskolinimai Long-term debt
	liablong	Kiti ilgalaikiai įsipareigojimai Other long-term liabilities
	liabtotal	Bendri įsipareigojimai Total liabilities
	stockpref	Privilegiuotosios akcijos Stocks preferred
	equity	Bendrasis kapitalas Common equity
	liabshare	Bendri įsipareigojimai ir akcininkų kapitalas Total liabilities and shareholder's equity
	average	Vidutiniškai emituotų akcijų Average shares outstanding
	marketcap	Rinkos kapitalizacija Market capitalization
	pricetoshare	Kainos ir akcijų skaičiaus santykis periodo gale Price to share (end of period)
<b>Nežinomi pa- rametrai (tyri- muose nenau- dojami)</b>	Whoknows_1	Whoknows_1
	Whoknows_2	Whoknows_2
	whoknows_3	Whoknows_3
	whoknows_4	Whoknows_4
<b>Išvestiniai ro- dikliai</b>	likvid	Trumpalaikis/likvidumo koeficientas Current ratio Trumpalaikis turtas / Trumpalaikiai įsipareigojimai assetcur/debshort
	liabinventory	Trumpalaikių įsipareigojimų ir inventoriaus santykis Current liabilities to inventory ratio Trumpalaikės skolos/inventorius liabcurr/inventory

Ataskaita	Santrumpa	Aprašymas
	totalliab	Visi išsiskolinimai grynai vertei Total liabilities to net worth ratio Visi išsiskolinimai / gryna vertė $Liabtotal/assetlong$
	collection	Surenkimo periodo koeficientas Collection period ratio Gaunamos pajamos / pardavimai x 365 dienos $salesinc/ costofg*365$
	salesinventory	Pardavimų ir inventoriaus koeficientas Sales to inventory ratio Grynieji metiniai pardavimai /inventorius $Sales/inventory$
	assetssales	Turto ir pardavimų koeficientas Assets to sales ratio Bendras turtas / Grynieji pardavimai $assetstot/salesinc$
	salescapital	Pardavimai ir grynasis įstatinis kapitalas Sales to net working capital Pardavimai / grynasis įstatinis kapitalas $salesinc/equity$
	accountssales	Mokėjimų ir pardavimų koeficientas Accounts payable to sales ratio Mokėjimai / Grynieji pardavimai $accpay/salesinc$
	quickratio	Greitas koeficientas Quick ratio (Gryni pinigai + gautinos pajamos) / trumpalaikiai įsipareigojimai $(cash+salesinc)/debshort$
	ros	Pardavimų grąža Return on sales ratio Grynasis pelnas atskaičius mokesčius / Grynieji pardavimai $netincome/salesinc$
	roa	Turto grąža Return on assets (ROA) ratio Grynasis pelnas atskaičius mokesčius / Visas turtas $netincome/assetstot$
	ronw	Grynosios vertės grąža Return on net worth ratio Grynasis pelnas atskaičius mokesčius / Grynoji vertė $netincome/equity$



tickers		
PK,FK1	<u>id</u>	INTEGER
I1	ticker	VARCHAR(5)
I2	company	VARCHAR(100)
I3	siccode	VARCHAR(10)
I4	industry	TINYINT
	lith	TINYINT

models		
PK	<u>id</u>	INTEGER
I1	date	DATETIME
	mid	TINYINT
	mtext	LONGVARCHAR
	name	VARCHAR(255)
	description	LONGVARCHAR
	attributes	LONGVARCHAR
	rmodel	VARCHAR(20)

10q_data		
PK	<u>id</u>	INTEGER
I2	tickerID	INTEGER
I1	quarter	TINYINT
	salesinc	DECIMAL(8,2)
	otherinc	DECIMAL(8,2)
	grossinc	DECIMAL(8,2)
	costofg	DECIMAL(8,2)
	randd	DECIMAL(8,2)
	deprec	DECIMAL(8,2)
	totalop	DECIMAL(8,2)
	nonrecc	DECIMAL(8,2)
	interest	DECIMAL(8,2)
	totalint	DECIMAL(8,2)
	grossop	DECIMAL(8,2)
	unusual	DECIMAL(8,2)
	pretax	DECIMAL(8,2)
	adjust	DECIMAL(8,2)
	inctax	DECIMAL(8,2)
	netincome	DECIMAL(8,2)
	sharesav	DECIMAL(8,2)
	eps	DECIMAL(8,2)
	epscont	DECIMAL(8,2)
	epsdilut	DECIMAL(8,2)
	dividend	DECIMAL(8,2)
	cash	DECIMAL(8,2)
	shorti	DECIMAL(8,2)
	receive	DECIMAL(8,2)
	inventory	DECIMAL(8,2)
	assetcur	DECIMAL(8,2)
	assettot	DECIMAL(8,2)
	netprop	DECIMAL(8,2)
	longinv	DECIMAL(8,2)
	assetlong	DECIMAL(8,2)
	goodwill	DECIMAL(8,2)
	totassets	DECIMAL(8,2)
	accpay	DECIMAL(8,2)
	debshort	DECIMAL(8,2)
	liabcurr	DECIMAL(8,2)
	Whoknows_1	DECIMAL(8,2)
	debtlong	DECIMAL(8,2)
	liablong	DECIMAL(8,2)
	liabtotal	DECIMAL(8,2)
	stockpref	DECIMAL(8,2)
	equity	DECIMAL(8,2)
	liabshare	DECIMAL(8,2)
	average	DECIMAL(8,2)
	Whoknows_2	DECIMAL(8,2)
	marketcap	DECIMAL(8,2)
	pricetoshare	DECIMAL(8,2)
I3	balancedata	DATE
I4	incomedata	DATE
	whoknows_3	DECIMAL(8,2)
	whoknows_4	DECIMAL(8,2)
I5	year_qtr	TINYINT

## Pagrindinių šiuolaikinių DBVS privalumai ir trūkumai

	MS AC- CESS	VISUAL FOXPRO 6.0	MYSQL 4.1	POSTGRES QL 7.2	FIREBIRD 2.0	MS SQL SERVER 2000	ORACLE 9
<b>Priemonės IS kūrimui</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Platforma</b>	Windows	Windows	Windows, Linux, BSD, kitos OS	Windows, Linux, BSD, kitos OS	Windows, Linux, kitos OS	Windows	Windows
<b>Sparta</b>	Vidutinė	Greita pa- ieška	Viena spar- čiausių	Gera	Gera	Didelė	Didelė
<b>Kaina<sup>15</sup></b>	Vidutinė <sup>16</sup>	Vidutinė	Nemokama, Open Sour- ce	Nemokama, Open Sour- ce	Nemokama, Open Sour- ce	Didelė	Didelė
<b>SQL standarto palaikymas<sup>17</sup></b>	Specifinis MS Access	Specifinis VFP	Nepilnas <sup>18</sup>	Pilnas	Specifinis Firebird	Pilnas	Pilnas
<b>Patikimumas</b>	Stabili	Nestabili	Stabili	Stabili	Stabili	Labai stabi- li	Labai stabi- li
<b>Darbas su di- deliu duomenų kiekiu</b>	Nelabai tinkamas	Vidutiniš- kas	Geras	Labai geras	Geras	Labai geras	Labai geras
<b>Daugelio varto- tojų palaiky- mas</b>	Nelabai tinkamas	Nelabai tinkamas	Labai geras	Labai geras	Geras	Labai geras	Labai geras
<b>DB serverio būtinybė</b>	Nėra, bet darbui rei- kalingas MS Access	Nebūtinus <sup>19</sup>	Nėra, yra embedded <sup>20</sup> galimybė	Būtinus	Nėra, yra embedded galimybė	Būtinus	Būtinus
<b>Administravi- mas</b>	Nesudėtin- gas	Nesudėtin- gas	Nesudėtin- gas	Sudėtingas	Nesudėtin- gas	Sudėtingas	Sudėtingas
<b>Kiti aspektai</b>						Papildomos SQL gali- mybės	Papildomos SQL gali- mybės

Šaltinis: sudaryta autoriaus.

<sup>15</sup> Jei kaina iki 1000 Lt, laikoma vidutinė, jei daugiau - didelė

<sup>16</sup> Parduodamas tik kartu su MS Office

<sup>17</sup> Turimas galvoje vidinis SQL dialektas

<sup>18</sup> Serverio procedūras ir vartotojo funkcijas (*User Defined Functions*) ir užklausas užklausoje (subqueries) galima kurti tik nuo 5 versijos.

<sup>19</sup> Jei MS FoxPro naudojama tik kaip aplikacijų kūrimo aplinka, tada priklauso nuo serverio

<sup>20</sup> Galimybė platinti kartu su aplikacija kaip jos dalį, kai nereikalingas serveris

Sukurtos sistemos realizacijos vartotojo sąsajos pavyzdžiai

**Duomenų redagavimo ir šalinimo modulio langai:**

**Duomenų peržiūra ir redagavimas**

Biržos kodų duomenų peržiūra

Finansinių duomenų peržiūra

**Duomenų importavimas**

Modeliai

Testavimas

Modelių redagavimas ir peržiūra

Apie

**Kompanijų ir biržos kodų sąrašas**

Pasirinkite puslapį: 1

Biržos kodas	Kompanija	SIC kodas	Pramonės šaka		
	AGRICULTURE, FORESTRY, AND FISHING	0	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	AGRICULTURAL PRODUCTION - CROPS	100	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Cash Grains	110	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Wheat	111	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Rice	112	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Corn	115	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Soybeans	116	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Cash grains (not elsewhere classified)	119	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Field Crops (Except Cash Grains)	130	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Cotton	131	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Tobacco	132	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Sugarcane and sugar beets	133	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Irish potatoes	134	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Field crops, except cash grains (not elsewhere classified)	139	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Vegetables and Melons	160	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Vegetables and melons	161	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Fruits and Tree Nuts	170	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>
	Berry crops	171	<input checked="" type="checkbox"/>	<a href="#">Redaguoti</a>	<a href="#">Šalinti</a>

**Duomenų peržiūra ir redagavimas**

Biržos kodų duomenų peržiūra

Finansinių duomenų peržiūra

**Duomenų importavimas**

Modeliai

Testavimas

Modelių redagavimas ir peržiūra

Apie

**Finansiniai ketvirčių/metų duomenys**

Filtruoti pagal biržos kodą: AGRICULTURAL PRODUCTION - CROPS

Pasirinkite puslapį: 1

Biržos kodo ID	Ketvirtis	Pajamos iš pardavimų	Kitos pajamos	Bendrosios įplaukos	Parduotų prekių ar paslaugų kaina	Tyrimai ir plėtra	Nuvertėjimas	Bendros veiklos išlaidos	Neperiodinės iššos balanse	Išlaidos palūkanoms	Bendr. išlaido palūka
1339	4	1822.00	-17.00	780.00	1042.00	232.00	0.00	232.00	0.00	0.00	10.00
1340	3	5818.00	0.00	1154.00	4664.00	43.00	294.00	43.00	-9.00	66.00	66.00
1341	3	978.40	-3.80	640.90	337.50	148.50	0.00	148.50	0.00	0.00	0.00
1342	3	1.30	0.00	0.30	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1343	3	193.20	-10.30	67.30	125.90	0.00	4.10	0.00	0.00	0.00	14.10
1344	2	56.00	0.10	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.20
1345	3	62.00	-0.10	20.10	41.90	0.00	0.70	0.00	0.00	0.60	0.60
1339	1	1212.00	-26.00	591.00	621.00	175.00	0.00	175.00	52.00	0.00	6.00
1347	2	0.80	0.10	0.30	0.50	0.00	0.00	0.00	0.00	0.00	0.00
1348	3	245.60	-1.10	21.80	223.80	0.00	3.70	0.00	0.00	0.00	4.60
1349	4	3.40	0.00	-0.50	3.90	0.00	0.00	0.00	0.00	0.00	1.00
1350	2	0.90	0.00	0.30	0.60	0.00	0.00	0.00	0.00	0.00	0.00
1351	3	49.60	0.00	3.90	45.70	0.00	0.00	0.00	0.00	0.70	0.70
1352	3	3.50	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50
1353	3	0.00	0.00	0.00	0.00	1.10	0.00	1.10	0.00	0.00	0.00
1354	3	47.40	-0.10	6.00	41.40	0.00	0.00	0.00	0.00	0.00	0.00
1355	3	890.20	0.00	416.60	473.60	0.00	0.00	0.00	0.00	0.00	4.30
1356	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1357	3	2350.00	-22.00	634.00	1716.00	122.00	0.00	122.00	0.00	0.00	3.00
1358	4	0.70	0.10	0.50	0.20	0.00	0.00	0.00	0.00	0.00	0.10

**Duomenų importavimo iš CSV failo modulio langų pavyzdžiai:**

**Duomenų peržiūra ir redagavimas**

**Duomenų importavimas**

Pasirinkite CSV failą su duomenimis:

Autorius Paulius Danėnas, 2007-2008

<b>Duomenų peržiūra ir redagavimas</b>	Pasirinkite atitinkamus duomenų stulpelius:
<b>Duomenų importavimas</b>	
Biržos kodų duomenų importavimas	
Finansinių duomenų įrašų importavimas	
<b>Modeliai</b>	
<b>Testavimas</b>	

**Bendri parametrai**

Biržos kodas:

Kompanijos/pramonės šakos pavadinimas:

SIC kodas:

Pramonės šakos žymė:

Autorius Paulius Danenas, 2007-2008

Įmonių duomenų importavimo dialogo pavyzdys

<b>Duomenų peržiūra ir redagavimas</b>	Pasirinkite atitinkamus duomenų stulpelius:
<b>Duomenų importavimas</b>	
<b>Modeliai</b>	
<b>Testavimas</b>	
<b>Modelių redagavimas ir peržiūra</b>	
<b>Apie</b>	

**Bendri parametrai**

Biržos kodas:

Kompanijos/pramonės šakos pavadinimas:

SIC kodas:

Pramonės šakos žymė:

---

**Pelno ataskaitos duomenys**

Ketvirtis:

Pelno ataskaitos pateikimo data:

Pajamos iš pardavimų:

Kitos pajamos:

Bendrosios įplaukos:

Parduotų prekių ar paslaugų kaina:

Tyrimai ir pletra:

Nuvertėjimas:

Bendros veiklos išlaidos:

Neperiodinės lešos balanse:

Išlaidos palūkanoms:

Bendros išlaidos palūkanoms:

Bendrosios veiklos pajamos:

Netikėtos pajamos:

Pajamos prieš mokesčius:

Pajamų patikslinimai:

Pajamų mokestis:

Grynasis pelnas:

Vidutinė akcijos kaina:

EPS:

Dabartinis (continued) EPS:

EPS Diluted:

Dividendai:

---

**Balanso duomenys**

Balanso pateikimo data:

Grynieji pinigai:

Trumpalaikės investicijos:

Debitorinės sąskaitos:

Prekių atsargos:

Kitas dabartinis turtas:

Bendras dabartinis turtas:

Grynoji gamybinių fondų nuosavybė:

Ilgalaikės investicijos:

Kiti ilgalaikiai aktyvai:

Prestižo vertė ir nematerialiosios vertybės:

Bendrieji aktyvai:

Tiekėjų įsiskolinimas:

Trumpalaikiai įsiskolinimai:

Kiti dabartiniai įsipareigojimai:

Whoknows\_1:

Ilgalaikiai įsiskolinimai:

Kiti ilgalaikiai įsipareigojimai:

Bendri įsipareigojimai:

Privilegiuotosios akcijos:

Bendrasis kapitalas:

Bendri įsipareigojimai ir akcininkų kapitalas:

Vidutiniškai emituotų akcijų:

Whoknows\_2:

Rinkos kapitalizacija:

Kainos ir akcijų skaičiaus santykis:

Whoknows\_3:

Whoknows\_4:

Autorius Paulius Danenas, 2007-2008

Finansinių duomenų importavimo dialogo pavyzdys

## SVM paremtų modelių kūrimo, testavimo, peržiūros ir redagavimo modulių langų pavyzdžiai

Sukurto modelio pateikimas:

Rezultatai | Atraminų vektorių sąrašas | Apmokymo proceso istorija

SVM tipas: nu\_svc  
 Kernelio tipas: rbf  
 $\gamma$  parametras: 0.01  
 Klasių skaičius: 2  
 Klasių žymės: 0 1  
 Bendras atraminų vektorių skaičius: 26  
 rho: 125,56  
 Tikimybė A: 0.12117419879796246  
 Tikimybė B: -2.155866562222182  
 Atraminų vektorių skaičius:  
 0: 13 1: 13

Išsaugoti modelį

Modelio pavadinimas:  
 Altman, RBF kernel, nu=0.005

Aprašymas:  
 Altman diskriminantinis modelis, RBF kernel, nu=0.005,  
 gamma=0.01, kiti standartiniai parametrai

Išsaugoti

Rezultatai | Atraminų vektorių sąrašas | Apmokymo proceso istorija

**Atraminio vektoriaus koeficientas: 447392.426666667**

Pajamos iš pardavimų	0.004670716887376996
Kitos pajamos	-0.002133633729138938
Bendrosios įplaukos	0.006150215499331573
Parduotų prekių ar paslaugų kaina	0.003723780289023118
Tyrimai ir plėtra	0.043574124219460594
Nuvertėjimas	-0.0012637087757223805
Bendros veiklos išlaidos	0.04462575194100743
Neperiodinės lėšos balanse	-5.002763079036964E-4
Išlaidos palūkanoms	-0.0010048092902181893
Bendros išlaidos palūkanoms	-7.806890837035548E-4
Bendrosios veiklos pajamos	4.2210654177919383E-4
Netikėtos pajamos	-6.824456018974936E-4
Pajamos prieš mokesčius	9.343264279229324E-4
Pajamų patikslinimai	6.08071866248414E-4
Pajamų mokestis	1.0357928464645415E-4
Grynasis pelnas	0.0011898344818144415
Vidutinė akcijos kaina	0.0011983871811030733
EPS	3.4588629896914914E-4
Dabartinis (continued) EPS	3.8354344344578093E-4
EPS Diluted	3.8445743785362526E-4
Dividendai	-8.429897448570982E-4
Grynieji pinigai	0.015382317874169502
Trumpalaikės investicijos	-4.5857434262070503E-4
Debitorinės sąskaitos	6.127099802000374E-4
Prekių atsargos	0.011936904375497347

Rezultatai | Atraminų vektorių sąrašas | Apmokymo proceso istorija

**Modelio apmokymo log'as**

```
<p>Warning: wrong number of cross validation, non-using any by default</p>
57
*
optimization finished, #iter = 0
C = -1230.632729406587
obj = 977.1583140780682, rho = 0.7063742389789481
nSV = 20, nBSV = 20
Total nSV = 20
*
optimization finished, #iter = 0
C = -6601.954156419085
obj = 3883.8844914300175, rho = 0.22144613871126415
nSV = 20, nBSV = 20
Total nSV = 20
```

## Sukurtų modelių sąrašas

Pasirinkite puslapį:

ID	Sukūrimo data	Pavadinimas	Aprašymas	Redaguoti	Šalinti
10	2008-01-21	Altman, 1000	Altman, first 1000 records	Redaguoti	Šalinti
11	2008-01-21	Zmijewski, all attributes	Zmijewski, 500 records, all attributes	Redaguoti	Šalinti
14	2008-01-21	Altman allattr	Altman, all attributes, 5000 records for training	Redaguoti	Šalinti
15	2008-01-21	Altman full-10000	Altman, used all attributes and 10000 records for training	Redaguoti	Šalinti
16	2008-01-21	Springate full-20000	Springate model, all attributes, used 20000 records for training	Redaguoti	Šalinti
17	2008-01-21	Altman all rbf	Used: Altman model, RBF for kernel, all attributes, 20000 instances for training	Redaguoti	Šalinti
19	2008-01-22	Altman full polynomial	Used: Altman model, polynomial kernel function, 10000 instances from database, all attributes	Redaguoti	Šalinti
20	2008-01-22	Altman full sigmoid	Used: Altman model, sigmoid kernel function, 10000 instances for training, all attributes	Redaguoti	Šalinti
21	2008-01-22	Springate full polynomial	Used: Springate model, all attributes, 10000 instances used for training, polynomial kernel	Redaguoti	Šalinti
22	2008-01-22	Springate 10 attributes	Springate model, used 10 attributes, which also include the ones that are used in the original Springate model. 20000 instances are used for training	Redaguoti	Šalinti
23	2008-01-26	Altman noprob rbf	Used: Altman model, 5000 instances, no probabilities option, RBF kernel	Redaguoti	Šalinti
24	2008-01-26	Altman noprob sigmoid	Altman, used 5000 instances, sigmoid kernel function, no probabilities option	Redaguoti	Šalinti
25	2008-01-26	Altman noprob polynomial	Altman, first 5000 instances, polynomial kernel function, no probabilities option	Redaguoti	Šalinti
26	2008-01-26	Altman polynomial 5	Altman model, polynomial kernel, all instances, degree of kernel equal 5	Redaguoti	Šalinti
29	2008-04-03	Altman, 5000, normalized	Altmano modelis, pirmi 5000 įrašų, visi rodikliai, normalizuoti duomenys, tiesinis kernelis	Redaguoti	Šalinti
31	2008-04-08	Altman, all ratios, normalized, 5000 records	Altman rizikos modelis, 5000 įrašų apmokymui, standartiniai SVM parametrai, normalizuoti duomenys	Redaguoti	Šalinti
32	2008-04-08	Altman, 9364 records, normalized, scaled, all ratios	Altman, 9364 records, normalized, scaled, all ratios , default SVM parameters	Redaguoti	Šalinti
34	2008-04-26	Springate, 5000 records, all ratios, standartized data	Springate diskriminantinis modelis, 5000 A rašų, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
35	2008-04-26	Zmijewski, all ratios, 5000 records, standartized data	Zmijewski diskriminantinis modelis, 5000 A rašų, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
36	2008-04-26	Altman, 5000 records, all ratios, polynomial	Altmano diskriminantinis modelis, 5000 irasu, polinominis kernelis, standartizuoti duomenys, standartiniai SVM parametrai, visi rodikliai	Redaguoti	Šalinti
37	2008-04-26	Springate, polynomial kernel, 5000 records, all ratios	Springate diskriminantinis modelis, polinominis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
39	2008-04-26	Zmijewski, polynomial kernel, 5000 records, all ratios	Zmijewski diskriminantinis modelis, polinominis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
40	2008-04-26	Altman, RBF kernel, 5000 records, all ratios	Altman diskriminantinis modelis, RBF kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
41	2008-04-26	Springate, RBF kernel, 5000 records, all ratios	Springate diskriminantinis modelis, RBF kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
42	2008-04-26	Zmijewski, RBF kernel, 5000 records, all ratios	Zmijewski diskriminantinis modelis, RBF kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
43	2008-04-26	Altman, sigmoid kernel, 5000 records, all ratios	Altman diskriminantinis modelis, sigmoidinis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
44	2008-04-26	Springate, sigmoid kernel, 5000 records, all ratios	Springate diskriminantinis modelis, sigmoidinis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
45	2008-04-26	Zmijewski, sigmoid kernel, 5000 records, all ratios	Zmijewski diskriminantinis modelis, sigmoidinis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai	Redaguoti	Šalinti
46	2008-05-16	Springate, 9364 records, linear	Springate diskriminantinis modelis, 9364 irasai apmokymui, standartiniai SVM parametrai, tiesine kernelio funkcija	Redaguoti	Šalinti
47	2008-05-19	Altman , linear kernel, c=2	Altman, tiesinis kernelis, 5000 irasu apmokymui, c=2	Redaguoti	Šalinti

Autorius Paulius Danenas, 2007-2008

## Modelio 45 duomenų redagavimas

Modelio ID:	<input type="text" value="45"/>
Sukūrimo data:	<input type="text" value="2008-04-26"/>
Modelio pavadinimas:	<input type="text" value="Zmijewski, sigmoid kernel, 5000 records, all ratios"/>
Aprašymas:	<input type="text" value="Zmijewski diskriminantinis modelis, sigmoidinis kernelis, 5000 irasu apmokymui, visi rodikliai, standartizuoti duomenys, standartiniai SVM parametrai"/>
Naudoti rodikliai:	<input type="text" value="salesinc, otherinc, grossinc, costofg, randd, deprec, totalop, nonrecc, interest, totalint, grossop,"/>
Rizikos modelis klasių formavimui:	<input type="text" value="Zmijewski"/>
Modelio tekstas:	<pre>svm_type nu_svc kernel_type sigmoid gamma 0.01 coef0 0.0 nr_class 2 total_sv 250 rho -1.2106109596662142 label 0 1 probA 0.0253575817431556 probB -0.3535464381168423 nr_sv 125 125 SV 460.7164211683756 0:0.005856883771905628 1:-0.002192111299439739 2:0.009512914601505275 3:0.00429746233549926 4:0.038084792688879474 5:-0.0014667627097960584 6:0.038084792688879474 7:-1.6006982067854247E-4 8:-0.0010102013166527809 9:-7.636628084617858E-4 10:5.980426571464932E-4 11:-6.326235285239582E-4 12:0.0012390486102804375 13:9.184513396904739E-4 14:3.449386104888713E-4 15:0.001471636333590401 16:0.001175456103869793 17:2.2926626366013636E-4 18:2.76799413168967E-4 19:2.7166460116120146E-4 20:-8.015164499665003E-4 21:0.011433490536809534 22:-4.970605628635411E-4 23:7.021720337009635E-4 24:0.012096087493924783 25:0.012006139171342103 26:0.016841869663105028</pre>
<input type="button" value="Sąrašas"/> <input type="button" value="Rodyti modelio duomenis"/> <input type="button" value="Pakeisti"/> <input type="button" value="Atšaukti pakeitimus"/>	

Autorius Paulius Danenas, 2007-2008

Modelio testavimo rezultatų pavyzdys

