

STATISTINIŲ KREDITO RIZIKOS VERTINIMO MODELIŲ EFEKTYVUMO ANALIZĖ

Ričardas Mileris

Kauno technologijos universitetas, Lietuva, ricardas.mileris@ktu.lt

Anotacija

Bankų vidaus kredito reitingų modeliais yra vertinama klientų kredito rizika. Norint nustatyti šių modelių efektyvumą turi būti skaičiuojami tam tikri kiekybiniai rodikliai. Straipsnyje pateikti kredito rizikos vertinimo modelių efektyvumą apibūdinantys rodikliai ir jų skaičiavimas: teisingo ir klaidingo klasifikavimo rodikliai, I ir II rūšies klaidos, modelių jautrumas ir specifiškumas, F -įvertis, ROC kreivė. Siekiant nustatyti dažniausiai naudojamų kredito rizikos vertinimo metodų efektyvumą buvo analizuojamos mokslinės publikacijos, kuriose aprašyti įvairių autorių sudaryti kredito rizikos vertinimo modeliai. Atlikus analizę nustatyta, kad tiksliausiai banko klientų kredito rizika įvertinama logistinės regresijos ir neuronų tinklų metodais. Mažiau patikimi metodai yra sprendimų medžiai ir diskriminantinė analizė. Taip pat buvo sudaryti 3 kredito rizikos vertinimo modeliai (dirbtinių neuronų tinklai), kuriais analizuoti Lietuvoje veikiančių įmonių duomenys. Suskaičiavus šių modelių efektyvumo rodiklius nustatyta, kad didžiausias įmonių klasifikavimo tikslumas pasiekiamas analizuojant 3 metų duomenis. Rodikliai leido įvertinti įmonių klaidingo klasifikavimo riziką bei kitas klasifikavimo charakteristikas. Remiantis šiais rodikliais galima objektyviai priimti sprendimą dėl bankų vidaus kredito reitingų modelių tinkamumo naudoti vertinant klientų kredito riziką.

Raktažodžiai: banko klientų klasifikavimas, kredito rizika, modelių efektyvumo rodikliai.

Išvadas

Kreditai klientams dažniausiai sudaro didelę bankų turto dalį ir yra viena rizikingiausių turto rūšių. Kreditų teikimas bankams yra ne tik pajamų šaltinis, tačiau taip pat gali būti ir pagrindinė nuostolių priežastis. Pagrindinė rizika, su kuria susiduria bankai skolindami pinigus, yra kredito rizika. Tai rizika, kad banko skolininkas dėl tam tikrų priežasčių negalės įvykdyti savo prisiimtų finansinių įsipareigojimų bankui. Bankai, norėdami sumažinti galimus nuostolius dėl kreditų suteikimo nepatikimiams klientams, turi sugebėti tinkamai įvertinti kiekvieno kliento kredito riziką. Vienas iš kredito rizikos vertinimo būdų yra bankų vidaus kredito reitingų modelių naudojimas. Sudarant bankų vidaus kredito reitingų modelius labai svarbu įvertinti modelių efektyvumą (Fernandes, 2005 ir kt.) apibūdinančius rodiklius. Modelio efektyvumas suprantamas kaip jo veiksmingumas, naudingumas. Tokių rodiklių skaičiavimas ir analizė moksliniuose darbuose dar vadinama klasifikavimo tikslumo (Golbraikh, 2005), modelių tinkamumo (Nagundkar, Priestley, 2006) nustatymu, klasifikavimo kokybės įvertinimu (Cortes, Mohri, 2004), klasifikatoriaus vertinimu (Orallo, 2004), klasifikatoriaus charakteristikų analize (Fawcett, 2006). Šie rodikliai leidžia tarpusavyje palyginti ir išrinkti geriausią modelį, kuriuo banko klientų patikimumas gali būti nustatytas tiksliausiai.

Tyrimo objektas – statistinių kredito rizikos vertinimo modelių efektyvumą apibūdinantys rodikliai.

Tyrimo tikslas – nustatyti statistinių kredito rizikos vertinimo modelių efektyvumą apibūdinančius rodiklius ir atlikti kredito rizikos vertinimo modelių efektyvumo analizę.

Tyrimo uždaviniai:

1. Nustatyti statistinių kredito rizikos vertinimo modelių efektyvumą apibūdinančius rodiklius.
2. Atlikti įvairių autorių sudarytų kredito rizikos vertinimo modelių efektyvumo analizę.
3. Sudaryti kredito rizikos vertinimo modelius ir įvertinti jų efektyvumą.

Tyrimo metodai:

1. Mokslinės literatūros ir mokslinių publikacijų, kuriose aprašyti kredito rizikos įvertinimo modeliai, analizė.
2. Kredito rizikos vertinimo modelių efektyvumo analizė.

Kredito rizikos vertinimo statistiniais modeliais esmė

Tam, kad būtų galima numatyti įmonių nemokumą ateityje, mokslininkai ir finansinių institucijų strategai rodo didelį susidomėjimą tyrinėjant įmonių nemokumo indikatorius bei kuriant ir vystant galimo nemokumo išankstinio perspėjimo sistemas. Banko vidaus kredito rizikos įvertinimo modeliais analizuojami dėl kredito besikreipiančių įmonių duomenys, o šios analizės rezultatai padeda priimti sprendimą dėl kredito suteikimo. Dažniausiai kiekybinių kredito rizikos vertinimo modelių paskirtis yra banko kliento finansinių

įsipareigojimų nevykdymo tikimybės nustatymas (Hamerle, Liebig, Rosch, 2003). Tam reikia turėti mokymo imtį, t.y. informaciją apie klientus ir jų patikimumą praeityje. Čia kredito prašanti įmonė turi būti apibūdinama tam tikru ją charakterizuojančių kintamųjų rinkiniu $x_1, x_2, \dots, x_n$. Klientus apibūdinantys požymiai turi leisti atskirti skirtingas klientų grupes, t.y. įvertinti jų kredito riziką. Mokymo imtį turi sudaryti patikimų ir nepatikimų klientų duomenys.

Ankstyvosios įmonių nemokumo prognozavimo studijos rodo, kad tyrinėtojai visų pirma kūrė ir vystė modelius, taikydami du statistinius duomenų analizės metodus – diskriminantinę analizę ir logistinę regresiją. Altman ir kt. (1968 ir 1977), Deakin (1972), Blum (1974) įmonių bankroto tikimybės nustatymui pritaikė diskriminantinę analizę, Ohlson (1980), Gentry (1985) – logistinę regresiją. Vėliau kredito rizikos vertinimui imta taikyti klasifikavimo medžius ir kitus metodus. Dabartiniu laiku kredito rizikos vertinimui plačiai taikomi dirbtinių neuronų tinklai (DNT). DNT teorija yra grindžiama biologinių neuronų tinklų veikimo principais ir leidžia atlikti sudėtingų netiesinių dinaminių sistemų analizę. Didžiojoje dalyje mokslinių straipsnių, kuriuose DNT lyginami su statistiniais duomenų analizės modeliais prognozuojant įmonių nemokumą teigiama, kad DNT yra pranašesni už statistinius modelius (Sookhnaphibarn (2007), Lai, Yn, Wang, Zhon (2006), Yim, Mitchel (2005) ir kt.). Tačiau kartais gaunami priešingi tyrimų rezultatai, kur DNT modeliais pasiekiamas mažesnis tikslumas nei analizuojant tuos pačius duomenis kitais statistiniais modeliais. Pvz., Bastos (2008), Yun, Jianying, Lin (2007) didesnis tikslumas pasiektas sprendimų medžiais. Altmanas ir kt. (1994) iškėlė idėją, kad DNT galėtų būti pagerinti integruojant į juos statistinius duomenų analizės metodus, taip optimizuojant DNT pateikiamus rezultatus. Hsieh, Liu, Hsieh (2007), Lai, Yn, Wang, Zhon (2006), Yim, Mitchel (2005), Yegorova (2001) sudarytuose hibridiniuose DNT (HDNT) dirbtinių neuronų tinklai derinami su kitais statistiniais duomenų analizės metodais. Jų tyrimų rezultatai parodė, kad HDNT modeliai yra labai veiksmingi prognozuojant įmonių bankrotą. Yim J., Mitchell H. (2005) teigė, jog DNT turi didelį potencialą kaip įrankis įmonių nemokumo prognozavimui, o jų tinkamas integravimas su kitais statistiniais duomenų analizės metodais gali dar labiau pagerinti įmonių klasifikavimo rezultatus.

Modelių efektyvumo įvertinimo rodikliai

Didelė kredito rizikos reikšmė finansų sektoriuje sąlygoja poreikį siekti kuo didesnio kredito rizikos vertinimo tikslumo, kuris įmanomas tik naudojant sudėtingus kiekybinius vertinimo metodus (Vasiliauskaitė, Cvilikas, 2008). Modelių kokybės įvertinimas galimas tik turint mokymo imtį, t.y. duomenis apie banko klientus ir jų finansinių įsipareigojimų vykdymą praeityje. Čia lyginami modelio analizės rezultatai su tikrosiomis (faktinėmis) klientų patikimumo reikšmėmis.

Vertinant modelių efektyvumą sudaroma banko klientų klasifikavimo matrica (1 lentelė). Matricoje informacija apie banko klientų patikimumą užkoduota taip: „0“ – įmonė savo finansinius įsipareigojimus įvykdė (klientas patikimas); „1“ – įmonė susidūrė su finansinių įsipareigojimų vykdymo problemomis (klientas nepatikimas).

Tikrinamos statistinės hipotezės:

1. H_0 : klientas nepatikimas (1).
2. H_1 : klientas patikimas (0).

1 lentelė. Banko klientų klasifikavimo matrica (Fawcett, 2006)

Modelio rezultatai	Tikrosios reikšmės	
	1	0
1	TP	FP (II rūšies klaida)
0	FN (I rūšies klaida)	TN

Dažniausiai banko klientų kredito rizikos įvertinimo modelių efektyvumui įvertinti skaičiuojamas teisingo klasifikavimo rodiklis (angl. *correct classification rate* (CCR) arba *hit rate*). Šis rodiklis parodo teisingai klasifikuotų įmonių dalį:

$$CCR = \frac{TP + TN}{N} \quad (1)$$

N – analizuojamų įmonių skaičius.

Klaidingo klasifikavimo rodiklis (angl. *misclassification rate*) parodo neteisingai klasifikuotų įmonių dalį:

$$MCR = \frac{FP + FN}{N} \quad (2)$$

Statistiniame teste klaida, kai teisinga nulinė hipotezė atmetama, laikoma I rūšies klaida (angl. *False negative rate*) (Weissstein, 2008). Ji parodo, kokia dalis nepatikimų klientų modeliui priskiriama patikimų klientų grupei:

$$\alpha = \frac{FN}{FN + TP} \quad (3)$$

Statistiniame teste klaida, kai klaidinga nulinė hipotezė priimama, laikoma II rūšies klaida (angl. *False positive rate*) (Weissstein, 2008). Ji parodo, kokia dalis patikimų klientų priskiriama nepatikimų klientų grupei:

$$\beta = \frac{FP}{FP + TN} \quad (4)$$

I ir II rūšies klaidos (α ir β) įvertina klaidingo klasifikavimo riziką. Galimi du klaidingo įmonių klasifikavimo atvejai:

1. Jei patikimas klientas priskiriamas nepatikimų klientų grupei ir jam nesuteikiamas kreditas, bankas netenka palūkanų, kurios būtų uždirbtos teisingo klasifikavimo atveju.
2. Jei nepatikimas klientas priskiriamas patikimų klientų grupei, tai bankui iškyla didelė tikimybė prarasti ne tik palūkanas, bet ir paskolintą sumą.

Taigi didesni nuostoliai galimi antruoju atveju (Mileris, 2008). Altman taip pat teigia, jog didesni bankų nuostoliai patiriami I rūšies klaidos atveju, todėl dažniausiai stengiamasi sumažinti šios klaidos tikimybę (Altman, 2002).

Modelio jautrumas (angl. *sensitivity*) yra išsipareigojimų neįvykdžiusių įmonių dalis, teisingai klasifikuota modeliui. Iš apibrėžimo matyti, kad jautrumas yra modelio statistinis klasifikavimo pajėgumas (t.y. $1 - \alpha$). Jis parodo, kokia dalis nepatikimų klientų buvo teisingai identifikuota modeliui:

$$Se = 1 - \alpha = \frac{TP}{TP + FN} \quad (5)$$

Modelio specifiškumas (angl. *specificity*) yra išsipareigojimus įvykdžiusių (patikimų) įmonių dalis, teisingai klasifikuota modeliui. Jis parodo, kokia dalis patikimų klientų buvo teisingai identifikuota modeliui (Fawcett, 2005):

$$Sp = 1 - \beta = \frac{TN}{TN + FP} \quad (6)$$

Didelio jautrumo modeliai dažniausiai linkę nustatyti analizuojamą ypatybę (finansinių išsipareigojimų nevykdymą). Didelio specifiškumo modeliai daugiau pateikia analizuojamos ypatybės neturinčių objektų. Kai analizuojamas įmonių finansinių išsipareigojimų nevykdymas, tai:

- Jautrūs modeliai labiau vengia neteisingai klasifikuoti nepatikimas įmones.
- Didelio specifiškumo modeliai pateikia labiau užtikrintą rezultatą įmonių finansinių išsipareigojimų nevykdymo atžvilgiu. Šiuo atveju nepatikimų įmonių grupei priskiriamos tik tos įmonės, kurios turi didžiausią riziką neįvykdyti finansinių išsipareigojimų.

Išsipareigojimų nevykdymo prognozės tikslumas (angl. *positive predictive value*) parodo, kokia dalis modeliui nustatytų nepatikimų klientų iš tikrųjų yra nepatikimi:

$$PPV = \frac{TP}{TP + FP} \quad (7)$$

Patikimų klientų klasifikavimo tikslumas (angl. *negative predictive value*): parodo, kokia dalis modeliui nustatytų patikimų klientų iš tikrųjų yra patikimi:

$$NPV = \frac{TN}{TN + FN} \quad (8)$$

F-įvertis (angl. *F-measure*):

$$F = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR} \quad (9)$$

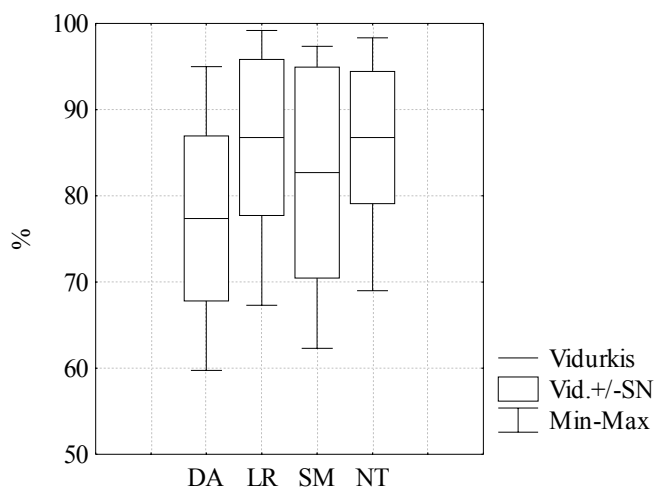
TPR šiuo atveju yra pažymėtas modelio jautrumas (*Se*). *F*-įvertis sujungia jautrumą ir išsipareigojimų nevykdymo prognozės tikslumą į vieną modelio efektyvumo matą. Matematiškai *F*-įvertis yra šių rodiklių harmoninis vidurkis (Hillenmeyer, 2005; Orallo, 2004).

Banko klientų kredito rizikos įvertinimo modelių tikslumui įvertinti taikoma ir grafinė analizė. ROC kreivė (angl. *receiver operator characteristics*) – tai kreivė, kuri dažniausiai naudojama rezultatų tikslumo įvertinimui sprendžiant klasifikavimo į dvi grupes uždavinius, kai turima mokymo imtis (2 pav.). Idealaus kredito rizikos vertinimo modelio atveju ROC kreivė turėtų pasiekti grafiko viršutinįjį kairinį kampą, kur teisingai nustatomos visos finansinių įsipareigojimų vykdymo problemų turinčios įmonės (100 proc.), t.y. modelio jautrumas būtų maksimalus. Šiuo atveju įsipareigojimų nevykdančių įmonių neteisingo klasifikavimo dalis lygi 0. Todėl kuo arčiau kreivė viršutiniojo kairiojo kampo, tuo didesnis modelio tikslumas. Ir priešingai, kuo mažiau išlenkta kreivė ir kuo ji yra arčiau įstrižinės, tuo modelis mažiau efektyvus. Grafiko įstrižinė vaizduoja bevertį modelį, t.y. visišką klasių neatskiriamumą. Nubraižius ROC grafikus galima vizualiai įvertinti ir palyginti kelių modelių efektyvumą. Kuo ROC kreivė yra arčiau kairiojo viršutiniojo kampo, tuo geresnis modelis. Tačiau jeigu kelių modelių ROC kreivės persikerta ir yra labai arti viena kitos, vizualiai įvertinti modelių efektyvumą sudėtinga. Tokiu atveju lyginamas plotas po ROC kreivėmis (AUC – angl. *area under curve*). Teoriškai jis yra intervale nuo 0 iki 1, tačiau kadangi modelis laikomas prasmingu tik tada, kai jo kreivė yra aukščiau $y = x$ įstrižinės, paprastai modelio efektyvumas vertinamas intervale nuo 0,5 (bevertis modelis) iki 1 (idealus modelis). Šios reikšmės gali būti gautos įvertinus plotą figūros, kuri apačioje ir dešinėje ribojasi su grafiko kraštinėmis, o viršuje ir kairėje pusėje – su ROC kreive (Falcon, 2007).

Modelių efektyvumo analizė

Siekiant nustatyti dažniausiai naudojamų kredito rizikos vertinimo metodų efektyvumą buvo analizuojama 30 mokslinių publikacijų, kuriose aprašyti 77 įvairių autorių sudaryti kredito rizikos vertinimo modeliai: Bastos (2008), Ciampi, Gordini (2008), Baxter, Gawlwer, Ang (2007), Hsieh, Liu, Hsieh (2007), Yun, Jianying, Lin (2007), Kim (2007), Altman, Sabato (2006), Lai, Yu, Wang, Zhou (2006), Nagundkar, Priestley (2006), Haselmann, Wachtel (2006), Keskinkilic, Sari (2006), Satchidananda, Simha (2006), Falavigna (2006), Huang, Chen, Wang (2006), Yim, Mitchell (2005) ir kt. Analizuotuose modeliuose dažniausiai kredito rizikos įvertinimui buvo pritaikyti dirbtinių neuronų tinklai (NT) - 63,3 proc., logistinė regresija (LR) - 53,3 proc., diskriminantinė analizė (DA) - 36,7 proc. ir sprendimų medžiai (SM) - 23,3 proc. Įvairūs kiti duomenų analizės metodai paminėti 33,3 proc. publikacijų.

Buvo vertinamas bendras modelių tikslumas, t.y. kokia dalis analizuotų klientų (proc.) buvo teisingai klasifikuota modeliais (CCR rodiklis). 1 pav. pavaizduotų objektų vidurinė linija žymi vidurkį, stačiakampių viršutinė ir apatinė kraštinės – tai reikšmių vidurkis $\pm$ standartinis nuokrypis ($Vid. \pm SN$), objektų viršūnės – tai didžiausios ir mažiausios reikšmės (Max ir Min).



1 pav. CCR rodiklio statistinės charakteristikos

Iš atliktos analizės rezultatų matyti, kad tiksliausiai banko klientų kredito rizika įvertinama logistinės regresijos ir neuronų tinklų metodais. Vidutinis šių metodų tikslumas yra beveik vienodas – 86,78 ir 86,77 proc. Mažiau patikimi metodai yra sprendimų medžiai ir diskriminantinė analizė, kurių vidutinis tikslumas atitinkamai yra 82,7 ir 77,38 proc. Labai svarbus rodiklis yra modelių tikslumo standartinis nuokrypis, atspindintis analizuotų reikšmių variaciją, t.y. vidutinį nuokrypį nuo vidurkio. Nors logistinės regresijos ir

neuronų tinklų tikslumo vidurkiai yra beveik vienodi (skirtumas tik 0,01 proc.), tačiau neuronų tinklų tikslumo variacija yra mažesnė. Tai rodo, kad neuronų tinklų vidurkis yra patikimesnė vidutinio modelių tikslumo charakteristika. Tačiau vertinant kiekvieną modelį atskirai, didžiausias tikslumas buvo pasiektas naudojant logistinę regresiją (99,2 proc.). Iš analizuotų modelių neuronų tinklais pasiektas maksimalus tikslumas – 98,35 proc. Taigi atlikta modelių lyginamoji analizė parodė, jog logistinės regresijos ir dirbtinių neuronų tinklų modeliai pateikia geriausius banko klientų kredito rizikos įvertinimo rezultatus.

Dirbtinių neuronų tinklų modelių efektyvumo analizė

Analizuojant 100 Lietuvos įmonių duomenis (50 iš jų bankrutavusių) buvo sudaryti 3 dirbtinių neuronų tinklai (neuronų tinklų tipas - daugiasluoksnis perceptronas). Modeliais analizuota 20 santykinų finansinių rodiklių. Esminis šių modelių skirtumas – analizuojamų rodiklių laikotarpis nuo 1 iki 3 metų (2 lentelė).

2 lentelė. Dirbtinių neuronų tinklų modelių charakteristikos

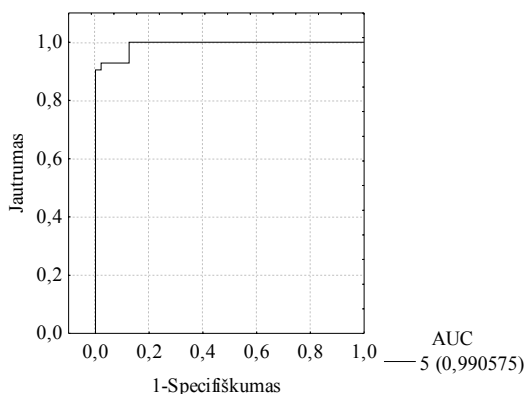
Modelis	Analizuojamų duomenų laikotarpis, metais	Analizei reikšmingų kintamųjų skaičius	Neuronų skaičius vidiniame sluoksnyje
1	1	15	13
2	2	37	16
3	3	50	23

Šių modelių efektyvumo rodikliai pateikti 3 lentelėje.

3 lentelė. Dirbtinių neuronų tinklų modelių efektyvumo rodikliai

Rodiklis	1 modelis	2 modelis	3 modelis
CCR	0,8587	0,9222	0,9551
MCR	0,1413	0,0778	0,0449
α	0,2143	0,1220	0,0976
β	0,0800	0,0408	0,0000
Se	0,7857	0,8780	0,9024
Sp	0,9200	0,9592	1,0000
F	0,8476	0,9168	0,9487
PPV	0,8919	0,9474	1,0000
NPV	0,8364	0,9038	0,9231
AUC	0,8876	0,9567	0,9906

Lentelėje matyti, kad didėjant analizuojamų metų skaičiui, teisingo klasifikavimo rodiklio (CCR) reikšmė didėja. Didžiausias įmonių klasifikavimo tikslumas pasiektas analizuojant 3 metų duomenis (95,5 proc.). Tai taip pat atspindi F-įverčio reikšmės. Didinant analizuojamų duomenų kiekį I ir II rūšies klaidos, parodančios, kokia dalis klientų modeliu klasifikuojama klaidingai, mažėja. Lentelėje matyti, kad 3-uoju modeliu teisingai klasifikuotos visos patikimos įmonės ($Sp = 1$) ir 90,2 proc. nepatikimų. 2 pav. pavaizduota 3-iojo modelio ROC kreivė, plotas po kuria (AUC) lygus 0,9906.



2 pav. 3-iojo modelio ROC kreivė

Suskaičiuoti kredito rizikos vertinimo modelių efektyvumo rodikliai leido įvertinti įmonių klaidingo klasifikavimo riziką bei kitas klasifikavimo charakteristikas. Remiantis šiais rodikliais galima objektyviai priimti sprendimą dėl bankų vidaus kredito reitingų modelių tinkamumo naudoti vertinant klientų kredito riziką.

Išvados

1. Sudarant bankų vidaus kredito reitingų modelius reikalingi jų efektyvumą apibūdinantys rodikliai. Jie leidžia įvertinti modelių tikslumo parametrus, tarpusavyje palyginti kelis modelius bei padeda priimti sprendimą dėl modelių tinkamumo naudoti bankų veikloje.
2. Pagrindinis modelių efektyvumą apibūdinantis rodiklis yra teisingo klasifikavimo rodiklis, kuris parodo, teisingai klasifikuotų banko klientų dalį. Vertinant modelio efektyvumą jis pateikiamas dažniausiai. Kiti rodikliai apibūdina įvairius modelių tikslumo parametrus.
3. Atlikus mokslinių publikacijų analizę nustatyta, kad iš dažniausiai naudojamų statistinių duomenų analizės metodų tiksliausiai banko klientų kredito riziką įvertinama pasitelkus logistinę regresiją ir dirbtinių neuronų tinklus.
4. Sudarius dirbtinių neuronų tinklų modelius nustatyta, jog modelių efektyvumo rodikliai gerėja didinant analizuojamų laikotarpių skaičių. Didžiausias įmonių klasifikavimo tikslumas pasiektas analizuojant 3 metų duomenis.

Literatūra

1. Altman E.I., 2002, Managing Credit Risk: The Challenge for the New Millennium // Economic Notes, Volume 31, Number 2, 1 July 2002, P. 201-214.
2. Altman, E.I., Sabato, G. (2006). Modeling Credit Risk for SMEs: Evidence from the US Market. Prieiga per internetą: www.stern.nyu.edu/~ealtman/ModelingCreditRiskforSMEs%20.pdf
3. Auria, L., Moro, R. (2007). Advantages and Disadvantages of Support Vector Machines // Credit Risk Assessment Revisited: Methodological Issues and Practical Implications. P. 49-68.
4. Bastos, J. (2008). Credit scoring with boosted decision trees // MPRA Paper No. 8156. 14 p.
5. Baxter, R.A., Gawlwer, M., Ang, R. (2007). Predictive Model of Insolvency Risk for Australian Corporations // Sixth Australasian Data Mining Conference (AusDM 2007). P. 19-25.
6. Chwee, P.G. (2004). Credit scoring using data mining techniques // Singapore Management Review. - Business Periodicals. Prieiga per internetą: <http://www.allbusiness.com/personal-finance/credit-cards/178130-10.html>
7. Ciampi, F., Gordini, N. (2008). Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: an Empirical Analysis. University of Florence, Italy. Prieiga per internetą: www.gcbe.us/2008_OBEC/data/Francesco%20Ciampi,%20Niccolo%20Gordini.doc
8. Cortes, C., Mohri M. (2004). AUC Optimization vs. Error Rate Minimization // Advances in Neural Information Processing Systems (NIPS 2003). volume 16, Vancouver, Canada, 2004. MIT Press. Prieiga per internetą: www.cs.nyu.edu/~mohri/postscript/auc.pdf
9. Falavigna, G. (2006). Models for Default Risk Analysis: Focus on Artificial Neural Networks, Model Comparisons, Hybrid Frameworks // Institute for Economic Research on Firms and Growth - CERIS Working Paper. Prieiga per internetą: <http://ideas.repec.org/p/csc/cerisp/200610.html>
10. Falcon, T. (2007). Logit Models to Assess Credit Risk // Credit Risk Assessment Revisited: Methodological Issues and Practical Implications. P. 28-47.
11. Fawcett, T. (2006). An Introduction to ROC Analysis // Science Direct Pattern Recognition Letters 27 (2006) P. 861-874.
12. Fernandes, J.E. (2005). Corporate Credit Risk Modelling: Quantitative Rating System And Probability Of Default Estimation. Prieiga per internetą: pwp.netcabo.pt/jed_fernandes/JEF_CorporateCreditRisk.pdf
13. Gallo, C., Letizia, C., Stasio, G. (2006). Artificial Neural Networks in Financial Modelling. Prieiga per internetą: <http://ideas.repec.org/n/nep-cba/2006-04-30.html>
14. Golbraikh, A. (2005). Classification Accuracy Criteria As Target Functions In QSAR. Prieiga per internetą: www.renci.org/focusareas/biosciences/ccega/meeting_notes/Golbraikh_4_26_05
15. Grabusts, P. (2004). Analysing Bankruptcy Data with Neural Network Method. 10-th International Conference on Soft Computing. Brno, June 16-18.
16. Hamerle, A., Liebig, T., Rosch, D. (2003). Credit Risk Factor Modeling and the Basel II IRB Approach // Deutsche Bundesbank Discussion Paper Series 2: Banking and Financial Supervision No 02/2003.

17. Haselmann, R., Wachtel, P. (2006). Bank Risk and Bank Management in Transition. A Progress Report on the EBRD Bankingenvironment and Performance Survey // EBRD Policy Conference Institutional Development, Market Integration and Growth in Emerging Markets. 27 p.
18. Hillenmeyer, M. (2005). Machine Learning. Stanford University. Prieiga per internetą: <http://www.stanford.edu/~maureenh/quals/html/ml>
19. Hsieh, W.K., Liu, S.M., Hsieh, S.Y. (2007). Hybrid Neural Network Bankruptcy Prediction: An Integration of Financial Ratios, Intellectual Capital Ratios, MDA, and Neural Network Learning // Journal of Nonlinear Mathematical Physics. ISSN: 1402-9251. Volume-issue: 14-2.
20. Huang, C.L., Chen, M.C., Wang, J.C. (2006). Credit Scoring With A Data Mining Approach Based On Support Vector Machines // Expert Systems with Applications. Prieiga per internetą: www.elsevier.com/locate/eswa
21. Yegorova I., Andrews B.H., Jensen J.B., Smoluk B.J., Walczak S., 2001, A Successful Neural Network-Based Methodology for Predicting Small Business Loan Default // Financial Services Industry Business Library. Prieiga per internetą: http://findarticles.com/p/articles/mi_qa3857/is_200110/ai_n8961377
22. Yim, J., Mitchell, H. (2005). A Comparison of Corporate Distress Prediction Models in Brazil: Hybrid Neural Networks, Logit Models and Discriminant Analysis // Nova Economia Belo Horizonte 15 (1). P. 73-93.
23. Yun, D., Jianying, Z., Lin, C. (2007). The Comparative Analysis of the Models in Default Warning of the Credit Clients in Commercial Banks // Journal of American Science, 3(2). P. 30-34.
24. Keskinilic, T., Sari, G. (2006). Probabilistic Prediction of Bankruptcy with Financial Ratios. - An empirical study on Swedish market. 40 p.
25. Kim, B.J. (2007). Bankruptcy Prediction: Book value or Market Value? APRIA Annual Meeting.
26. Lai, K.K., Yu, L., Wang, S., Zhou, L. (2006). Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model // Springer-Verlag Berlin Heidelberg. P. 682-690.
27. Mileris, R. (2008). Požymių reikšmių pasiskirstymo ir variacijos vertinimas sudarant banko vidaus kredito reitingų modelius // Ekonomika ir vadyba – 2008. Kaunas. P. 143-148.
28. Nagundkar, S., Priestley, J.L. (2006). Assessment of Evaluation Methods Forbinary Classification Modeling. Department of Management and Decision Sciences, J. Mack Robinson College of Business Administration, Georgia State University.
29. Satchidananda, S.S., Simha, J.B. (2006). Comparing Decision Trees With Logistic Regression For Redit Risk Analysis. International Institute of Information Technology, Bangalore, India. 7 p.
30. Sookhanaphibarn, K., Polsiri, P., Choensawat, W., Lin F.C. (2007). Application of Neural Networks to Business Bankruptcy Analysis in Thailand // International Journal of Computational Intelligence Research. Vol.3, No.1 (2007). P. 91–96.
31. Vasiliauskaitė, A., Cvilikas, A. (2008). Neapibrėžtų aibių teorijos taikymas kredito rizikos vertinimui // Apskaitos ir finansų mokslas ir studijos: problemos ir perspektyvos. Akademija. P. 161-165.
32. Vazza, D., Aurora, D., Kraemer, N. (2008). Default, Transition and Recovery: 2007 Annual Global Corporate Default study and Rating Transitions. „Standard & Poor’s“. 19 p.
33. Weisstein, E.W. (2008). Type II Error. MathWorld - A Wolfram Web Resource. <http://mathworld.wolfram.com>

ANALYSIS OF STATISTICAL CREDIT RISK ESTIMATION MODELS EFFICIENCY

Ricardas Mileris

Summary

Banks always seek to reduce potential loss due to crediting not reliable clients. So they must be able to estimate credit risk of each client properly. One of possible instruments for credit risk measure is use of internal credit rating models in banks. Bank must determine significant attributes and select methods for credit risk estimation so that possibility of false decisions would be minimized. In most cases more than one model can be constructed and there is a problem which model to use in bank's activity. In practice the best model must be used so for the comparison of different models certain rates of models efficiency must be calculated. The purpose of paper is to define the indicators of statistical credit risk estimation models efficiency and to analyze the efficiency of created credit risk estimation models. In this paper rates of credit risk estimation models were described: correct classification and misclassification rates, false negative and false positive rates, model sensivity and specificity, F-measure, ROC analysis. Analysis of scientific publications about credit risk estimation models has shown that the most efficient of the most commonly used methods are logistic regression and artificial neural networks. Also three artificial neural networks models (multilayer perceptrons) were constructed. The most efficient model analyses data about clients of 3 years. Models efficiency rates allowed to estimate risk of client misclassification and other characteristics. Also they help make a decision which model to use in practice.

Keywords: classification of banks clients, credit risk, model efficiency rates.

