

Doc. dr. sc. Ljiljanka Kvesić
ljkvesic@gmail.com
Sveučilište u Mostaru
Fakultet prirodoslovno - matematičkih i odgojnih znanosti
Matice Hrvatske bb, 88000 Mostar
Bosna i Hercegovina
tel.: +38736355455

UIDK 336.77
Izvorni znanstveni članak

PRIMJENA STABLA ODLUČIVANJA U KREDITNOM SKORINGU

Sažetak

Banke su u svomu poslovanju posebno izložene kreditnome riziku. Pogrešna procjena zajmotražitelja ima za izravnu posljedicu gubitke. Financijska kriza kroz koju svjetsko gospodarstvo još uvijek prolazi nedvosmisleno pokazuje kakvi problemi mogu proisteći iz pogrešno vođene kreditne politike. Stoga je primarni zadatak bankarskih menadžera reducirati kreditni rizik na najmanju moguću mjeru. Kao podrška menadžerima u procjeni kreditne sposobnosti zajmotražitelja razvijeni su modeli kreditnoga skoringa. U ovom je radu kao takav prezentiran model stabla odlučivanja temeljen na iscrpnom CHAID algoritmu. Budući da primjena modela kreditnog skoringa nije na zadovoljavajući način istražena u hrvatskoj bankarskoj teoriji i praksi, ovim se radom nastojalo, ne samo determinirati karakteristike koje su ključne za predikciju statusa neispunjavaanja obveza, već i aktualizirati važnost kvantitativnog pristupa pri procjeni kreditne sposobnosti zajmotražitelja.

Ključne riječi: kreditni skoring, model, stabla odlučivanja, iscrpni CHAID algoritam

1. Uvod

Kreditni skoring je sustav dodjeljivanja bodova za zajmotražitelju pri čemu se primjenjuje statistička analiza ključnih karakteristika zajmotražitelja s ciljem numeričke kvantifikacije rizika koja pokazuje vjerojatnost da će klijent doći u status neispunjavaanja ugovorene obveze. Kreditni skoring sustav dodjeljuje jednu kvantitativnu mjeru, takozvani skor, potencijalnom klijentu predstavljajući njegovo buduće ponašanje u otplati dodijeljenoga kredita. Ukupan broj bodova (skor) dobiven sumom bodova po određenim karakteristikama uspoređuje se s najniže prihvatljivim brojem bodova definiranim kreditnom politikom banke na temelju čega se kredit odobrava ili ne odobrava.

Skoring modeli mogu se podijeliti na aplikativni skoring (odnosi se na nove klijente) i bihevioralni skoring (odnosi se na postojeće klijente). Bihevioralni skoring je nastavak razvoja kreditnog scoringa. Kod bihevioralnog scoringa upotrebljavaju se informacije o ponašanju klijenta pri otplati kreditnih obveza u prošlom periodu, koji se zove period promatranja. Mogu se koristiti i podaci trenutno dostupni iz kreditnoga registra kao i oni prikupljeni pri podnošenju zahtjeva za kredit. Svi se ovi podaci koriste za predviđanje rizika neplaćanja klijenta u sljedećih dvanaest mjeseci ili bilo kojem budućem fiksnom periodu. Zajmodavatelji ažuriraju bihevioralni skor na mjesečnoj razini i u većini slučajeva informacije o povijesnim performansama i tekuće informacije kreditnoga registra mnogo su kvalitet-

nije nego informacije iz zahtjeva za kredit.

Financijske institucije u razvijenim zemljama primjenjuju različite modele kreditnoga skoringa razvijene na temelju vlastitih podataka ili podataka drugih financijskih institucija. Poseban problem nastaje kada kreditni odjel banke ne raspolaže podacima te kada unutar bankarskoga sektora ne postoje prethodne spoznaje i iskustva vezana uz razvoj i primjenu modela kreditnoga skoringa, kao što je to slučaj u Hrvatskoj. Cilj je ovoga rada razvijanje kreditnog skoring modela, utemeljenoga na stablu odlučivanja, koji bi bio primjenjiv u poslovanju hrvatskih financijskih institucija te na taj način popunio postojeću prazninu u teoriji i praksi. Dakle, istraživanjem se nastojalo otkriti koja su karakteristike ključne za predikciju statusa neispunjavanja obveza (eng. *default*). Pri izgradnji modela korišteno je nekoliko varijabli demografskoga i financijskoga karaktera.

2. Prethodna istraživanja

Stabla odlučivanja značajna su podrška menadžerima u području kreditnog skoringa. U nastavku će, ukratko, kronološkim redoslijedom, biti prezentirano nekoliko radova u kojima su analizirane mogućnosti korištenja stabala odlučivanja pri procjeni kreditne sposobnosti zajmotražitelja.

Joos et al. usporedili su rezultate logit modela i stabla odlučivanja. S tom su svrhom koristili opsežnu bazu podataka jedne od belgijskih najvećih banaka. Prema autorima, globalna točnost logit modela pokazala se veća u slučaju cjelovite sheme skupa podataka dok su za kvalitativne podatke i kraće sheme, važnija bila stabla odlučivanja.

Zurada i Zurada istražili su podatke o kreditima koje je odobrila jedna financijska institucija. Podaci su se sastojali od financijskih obilježja svakoga klijenta. Autor su u radu primijenili tri različite tehnike rudarenja podataka (stabla odlučivanja, neuronske mreže i logit regresiju) te model koji kombinira sve tri metode. S ciljem usporedbe učinkovitosti svakoga od primijenjenih modela razvili su dva scenarija. Zurada i Zurada zaključuju da su najbolje rezultate klasifikacije dobili primjenom neuronskih mreža i modela koji je kombinirao tri analizirane tehnike.

U svom su radu Xiao, Zhao i Fei ispitali performanse različitih modela kreditnoga skoringa i pripadajuće troškove kreditnoga rizika. U istraživanju su koristi-

li podatke koji se odnose na njemačke i australske zajmotražitelje, pri čemu su definirali 24 varijable. Kao klasifikacijske algoritme primijenili su linearnu diskriminacijsku analizu, logističku regresiju, neuronske mreže, k-najbližih susjeda, metodu potpornih vektora, klasifikacijska i regresijska stabla te metodu MARS (Multivariate Adaptive Regression

Splines). Na temelju dobivenih rezultata autori su zaključili da metoda potpornih vektora, MARS, logistička regresija i neuronske mreže imaju vrlo dobre performanse, poglavito u smislu klasifikacije dobrih klijenata. No, linearna diskriminacijska analiza i klasifikacijska i regresijska stabla pokazala su se značajno preciznijim u identificiranju loših zajmotražitelja.

Satchidananda i Simha koristili su podatke dvaju indijskih banaka kako bi usporedili učinkovitost stabala odlučivanja i logističke regresije u predikciji dobrih i loših zajmotražitelja. Autori su se fokusirali na kredite vezane uz poljoprivrednu proizvodnju. Njihov je zaključak da stabla odlučivanja bolje klasificiraju dobre i loše klijente nego model logističke regresije.

Zurada je primijenio pravilo metode indukcije (stabla odlučivanja) kako bi predvidio hoće li ili ne zajmotražitelj biti u nemogućnosti izvršiti ugovornu obvezu. Njegov je uzorak brojao 5690 podnositelja zahtjeva za kredit. Autor je proveo računalnu simulaciju tri različite metode stabla odlučivanja (hikvadrat, entropijsku redukciju i Ginijevu redukciju) Na temelju rezultata analize svih dobivenih klasifikacijskih stopa Zurada preporučuje upotrebu hikvadrat metode.

U svojoj su se studiji Pang i Gong fokusirali na procjenu kreditne sposobnosti pojedinaca pomoću klasifikacijskog algoritma stabla odlučivanja C5.0. Model su evaluirali pomoću podataka o zajmotražiteljima jedne njemačke banke te su usporedili dobivene rezultate s originalnim modelom. Istraživanje je pokazalo da prilagođeni model stabla odlučivanja bolje klasificira zajmotražitelje kao dobre i loše.

Kao podršku menadžerima iz područja kreditnog skoringa, Zhang et al. su razvili poboljšani model vertikalnih "bagging" stabala odlučivanja. U testiranju modela koristili su dvije baze podataka. Zaključak je autora da predloženi model pokazuje jako dobre karakteristike s obzirom na korektnost procjene zajmotražitelja.

Yap, Ong i Husain prikazali su mogućnosti korištenja tehnika rudarenja podataka kao podrške u procesu procjene kreditne sposobnosti zajmotražitelja. U radu su usporedili tradicionalni model bodovanja, logističku regresiju i stablo odlučivanja, pri čemu su dobivene stope pogreške 27.9%, 28.8% i 28.1%, respektivno.

3. Stabla odlučivanja

Stabla odlučivanja predstavljaju neparаметarsku tehniku klasifikacije klijenata u skupine, gdje je svaka skupina homogena u odnosu na rizik neispunjenja obveza i maksimalno diferencirana od rizika neispunjenja obveza ostalih skupina. Stabla odlučivanja vrlo su moćne i popularne tehnike za klasifikacijske i predikcijske probleme.

Stablo odlučivanja predstavljeno je klasifikacijskim algoritmom u obliku razgranatog stabla s čvorovima. Kod stabla odlučivanja postoje dvije vrste čvorova:

- krajnji čvor (njime završava određena grana stabla);
- čvor odluke (definira određeni kriterij u obliku vrijednosti atributa iz kojeg izlaze grane koje zadovoljavaju određene vrijednosti tog atributa).

Kod tehnike klasifikacijskoga stabla skup odgovora iz zahtjeva za kredit dijeli se na dva podskupa. Fokusiranjem na odgovor na jedno pitanje, opredjeljuje se za podjelu mogućih odgovora na dva podskupa, gdje je razlika u prosječnom riziku neispunjenja obveza između dva podskupa što je moguće veća.

Pretpostavka primjene stabla odlučivanja je posjedovanje baze podataka aplikantata za kredit, koji su

opisani s n atributa x_1, x_2, \dots, x_n . Podnositelji zahtjeva pripadaju dvjema skupinama te su shodno tome označeni kao dobri (nerizični) i loši (rizični) klijenti. Cilj modela kreditnog skoringa jest pronalazak klasifikatora (atributa) koji najbolje razdvaja uzorak dobrih klijenata od uzorka loših klijenata. Algoritam počinje čvorom koji sadrži uzorke dobrih i loših klijenata, nakon čega se pronalaze svi mogući ishodi s ciljem dobivanja najkorisnijeg atributa x i odgovarajuće granične vrijednosti c koja najbolje vrši razdvajanje uzoraka dobrih i loših klijenata. Podaci se dijele prema svim mogućim kriterijima u dvije grane. Pri tome se odabire kriterij koji podatke dijeli u skupine koje su više homogene

od početne skupine podatka. Procedura se ponavlja sve dok podatke nije moguće dalje dijeliti u skupine koje su homogenije od početnih podataka.

Prema Buću proces konstruiranja stabla odlučivanja može se opisati u tri osnovna koraka:

1. Izgradnja logičkog modela - stabla odlučivanja s ulaznim podacima:

- svim čvorovima odluka, čvorovima posljedica, granama alternativnih akcija i granama posljedičnih stanja, i to u kronološkom redoslijedu; granama se pridružuju vjerojatnosti pojave pojedine posljedice te parcijalan tijek novca (potreban priljev ili odljev sredstava kako bi se stablo nastavilo granati).

2. Računanje očekivanih vrijednosti odluka postupkom računanja unatrag:

- računanje unatrag započinje na krajnjim čvorovima stabla i kreće prema unatrag, do početnog čvora odluke. Svakom čvoru pridodaje se ekvivalentna očekivana vrijednost, i to:
 - na završnom čvoru izračunata je konačna vrijednost te alternative,
 - čvoru posljedica pridružuju se očekivane vrijednosti izračunate kao:

$$EV_{i-1} = \sum_j p_j EV_j, \quad i \in \{1, 2, \dots, n\}, \quad j \in \{1, 2, \dots, m\},$$

gdje je:

EV_{i-1} - očekivana vrijednost u čvoru $i-1$,

p_j - vjerovatnost grane j koja izlazi iz čvora posljedica $i-1$

- na čvoru odluka očekivana pridružena vrijednost jednaka je najvećoj od algoritmom prethodno izračunatih očekivanih vrijednosti neposrednih sljedećih čvorova u stablu odlučivanja.

3. Pronalaženje optimalnoga puta postupkom računanja prema naprijed:

Nakon što su se u prethodnom koraku izračunale očekivane vrijednosti za svaki čvor, može se prepoznati optimalan put u stablu odlučivanja, računanjem prema naprijed, od početnoga čvora odluke.

Njemu pridružena vrijednost je jednaka očekivanoj vrijednosti grane koja se nalazi na optimalnom putu. Analogno se promatra sljedeći čvor, sve do završnog čvora.

Za testiranje svakog pojedinog razdvajanja i mjerenje homogenosti podataka koriste se Kolmogorov-Smirnovljeva statistika, indeks diverzifikacije, Ginijev koeficijent i indeks entropije.

Osnovne prednosti metode stabla odlučivanja su: mogućnost generiranja razumljivih modela, relativno mali zahtjevi za računalnim izvorima (vrijeme i memorija), jasna važnost pojedinih atributa za konkretni problem te široka dostupnost softverskih rješenja. Nedostatak stabala odlučivanja je njihova nestabilnost jer male fluktuacije u uzorku podataka mogu rezultirati velikim varijacijama u dodijeljenim klasifikacijama.

Stabla odlučivanja tehnika su modeliranja koja se koristi u slučajevima klasifikacijskih i predikcijskih problema. Njihova prikladnost u predviđanju statusa neispunjavanja kreditnih obveza proizlazi iz činjenice da je model dobio primjenom stabla odlučivanja pregledan i vrlo razumljiv, odnosno prezentiran je u obliku pravila.

4. Rezultati empirijskoga istraživanja

4.1. Uzorak

Istraživanjem je bilo obuhvaćeno 200 klijenata jedne poslovne banke. Među njima je bilo 100 dobrih i 100 loših klijenata. Tablica 1. prikazuje razdiobu klijenata prema spolu.

Tablica 1. Razdioba klijenata prema spolu

Spol	Broj klijenata	Postotak
muški	114	57.0
ženski	86	43.0
Ukupno	200	100.0

Prosječna starost klijenata izabраниh u uzorak iznosi je 56,54 godine, sa standardnom devijacijom 13,74 godine. Vrijednost medijana bila je 60 godina. Najmlađi klijent imao je 25, a najstariji 82 godine, iz čega proizlazi da je raspon varijacije 57 godina. Na temelju izračunatog koeficijenta varijacije, koji je u ovom slučaju iznosio 24,3%, može se zaključiti da razdiobu klijenata prema starosti karakterizira stupanj varijabilnosti podataka koji se ne može smatrati malim.

Tablicom 2. prikazana je razdioba klijenata prema njihovom statusu.

Tablica 2. Razdioba klijenata prema statusu

Status	Broj klijenata	Postotak
zaposlen	31	15.5
nezaposlen	4	2.0
umirovljenik	82	41.0
nepoznato	83	41.5
Ukupno	200	100.0

Za najveći broj klijenata nije bilo moguće utvrditi status, što predstavlja ograničenje u istraživanju. Nakon te skupine, najbrojniji su bili umirovljenici. Zaposlenih je bilo samo 15.5%, dok su 4 osobe bile prijavljene na Zavodu za zapošljavanje.

S obzirom na starost računa klijenti su raspodijeljeni u tri skupine (tablica 3.).

Tablica 3. Razdioba klijenata prema starosti računa

Starost računa	Broj klijenata	Postotak
3-9	108	54.0
10-17	68	34.0
18-27	24	12.0
Ukupno	200	100.0

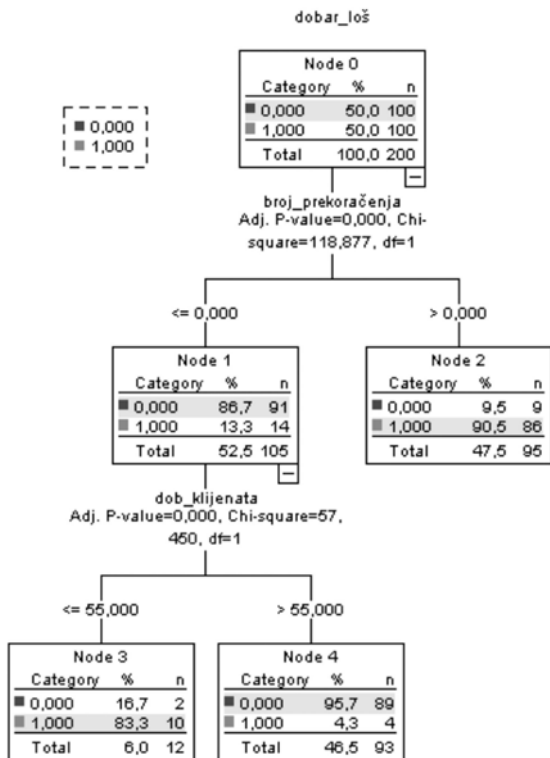
Na temelju formirane razdiobe može se zaključiti da je najveći broj računa bio star između 3 i 9 godina. Samo 12% klijenata imalo je u banci otvoren račun prije 18 i više godina.

4.2. Primjena CHAID algoritma u kreditnom skoringu

Razvijeno je više algoritama za konstruiranje stabla odlučivanja. U ovom istraživanju primijenjena je iscrpna hi-kvadrat automatizirana interakcijska detekcija (eng. *exhaustive chi-squared automatic interaction detection* (CHAID)).

Kod CHAID algoritma u svakom se koraku izabire nezavisna varijabla koja ima najjaču interakciju sa zavisnom varijablom. Ako se s obzirom na zavisnu varijablu značajno ne razlikuju, modaliteti svakog prediktora se spajaju. Iscrpni CHAID algoritam predstavlja modifikaciju izvorne metode u kojoj se ispituju sva moguća razdvajanja svakoga prediktora.

Slika 1. Stablo odlučivanja dobiveno primjenom iscrpnog CHAID algoritma



Varijabla definirana kao dobar/loš klijent predstavljala je zavisnu varijablu u modelu stabla odlučivanja, dok su kao nezavisne (prediktorske) definirane sljedeće varijable:

- spol;
- dob;
- status klijenata;
- minimalno dozvoljeno prekoračenje tijekom 6 mjeseci;
- prosječan iznos svih plaćanja čekovima;
- prosječan iznos plaćanja karticom;
- prosječan saldo za svih 6 mjeseci (prosječno stanje na računu);
- prosječan iznos sredstava koji klijent još smije potrošiti;
- broj nedozvoljenih prekoračenja (koliko je puta klijent bio u nedozvoljenom prekoračenju).

Slika 1 prikazuje stablo odlučivanja dobiveno primjenom iscrpnog CHAID algoritma pomoću statističkog paketa SPSS.

Na temelju dobivenog stabla odlučivanja može se zaključiti da je među klijentima koji su se najmanje jedanput nalazili u nedozvoljenom prekoračenju bilo 90,5% loših klijenata. Među klijentima koji nisu bili niti jednom u nedozvoljenom prekoračenju, a imaju 55 ili manje godina bilo je 83,3% loših klijenata, dok je među klijentima koji nisu bili niti jednom u nedozvoljenom prekoračenju, a imaju više od 55 godina, bilo 95,7% dobrih klijenata.

Tablica 4. Klasifikacijska tablica determiniranog modela

		Predviđen		
		dobar klijent	loš klijent	Postotak korektno predviđenih
Opažen	dobar klijent	89	11	89.0
	loš klijent	4	96	96.0
Ukupno				92.5

P(loši)	% dobri	% loši	Kumulativno % dobri	Kumulativno % loši	Apsolutna razlika kumulativno % dobri - kumulativno % loši
0.04301	89	5	89	4	85
0.83333	2	10	91	14	77
0.90526	9	86	100	100	0

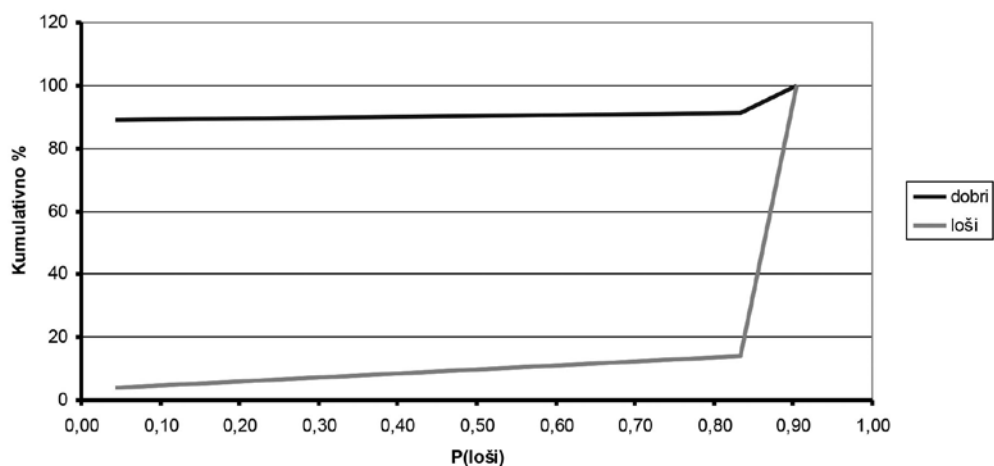
Tablica 5. Distribucija frekvencija dobrih i loših klijenata prema skor vrijednostima, odnosno vjerojatnostima

Stablom odlučivanja, kreiranom pomoću CHAID algoritma, ispravno je klasificirano 89 dobrih i 96 loših klijenata. Dakle, 11 dobrih i 4 loša klijenta nisu bila ispravno klasificirana. Ukupno je 92,5% klijenata korektno klasificirano pa se model može smatrati zadovoljavajućim.

Tablica 5. prikazuje distribuciju frekvencija dobrih i loših klijenata prema score vrijednostima, odnosno vjerojatnostima.

Slikom 2. grafički je prikazana kumulativna distribucija dobrih i loših klijenata dobivena na temelju modela stabla odlučivanja uz primjenu iscrpnog CHAID algoritma.

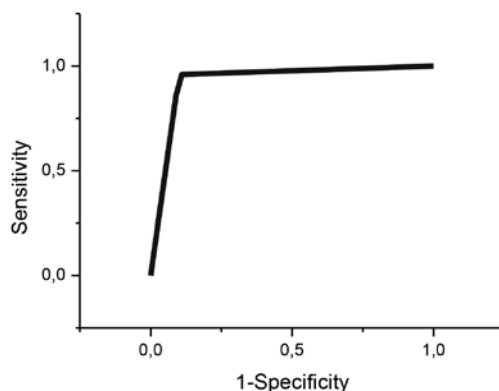
Slika 2. Kumulativna distribucija dobrih i loših klijenata dobivena na temelju modela stabla odlučivanja uz primjenu iscrpnog CHAID algoritma



Slika 3. prikazuje ROC krivulju modela dobivenog primjenom stabla odlučivanja (CHAID algoritam).

I ROC krivulja, koja je brzo dostigla razinu potpune senzitivnosti, ukazuje na relativno veliku predikcijsku snagu modela stabla odlučivanja temeljenog na iscrpnom CHAID algoritmu.

Slika 3. ROC krivulja



5. Zaključak

Pogrešna procjena kreditne sposobnosti zajmotražitelja ima za posljedicu gubitke, koji mogu izazvati velike probleme u funkcioniranju banaka. S jedne strane odobravanje kredita klijentu koji neće biti u mogućnosti ispuniti svoje obveze direktni je trošak banke, no i neodobravanje kredita dobrom klijentu umanjuje njezin financijski učinak. Stoga je pitanju procjene kreditne sposobnosti zajmotražitelja potrebno posvetiti posebnu pozornost. Kako je u radu i istaknuto, kao podrška donositeljima odluka iz te domene razvijeni su i koriste se brojni modeli kreditnog scoringa. U ovom su radu prezentirane mogućnosti primjene modela stabla odlučivanja temeljenog na iscrpnom CHAID algoritmu. Uzorak korišten u analizi sastojao se od 200 klijenata.

Rezultati analize pokazali su da se važnim prediktorom može smatrati varijabla definirana kao broj nedopuštenih prekoračenja i starost klijenta. Naime, velika većina loših klijenata barem se jedanput nalazila u nedopuštenom prekoračenju, kao i klijenti

stari 55 ili manje godina koji ranije nisu niti jednom prekoračili dopušteno prekoračenje. Budući da je 92,5% klijenata bilo korektno klasificirano, model se može smatrati zadovoljavajućim. Iako je tek nakon implementacije u praksi moguće izvršiti adekvatnu validaciju modela, na temelju navedenoga ipak se može zaključiti da primjena prezentiranoga modela doprinosi poboljšanju upravljanja kreditnim rizikom.

U cilju prevencije i minimiziranja vjerojatnosti pojavljivanja statusa neispunjavanja obveza, banke su prisiljene nastaviti razvijati kreditne scoring modele. Zbog nedostatka iskustva i vrlo malo istraživačkoga rada unutar područja kreditnog scoringa, banke u Hrvatskoj nisu dosegle odgovarajuću razinu primjene vlastitih modela procjene. Ukazujući na mogućnosti korištenja modela kreditnog scoringa, u ovom je radu aktualizirana važnost kvantitativnog pristupa pri procjeni kreditne sposobnosti zajmotražitelja. Na taj su način postavljeni temelji za daljnja istraživanja u području kreditne analize klijenata.

LITERATURA

1. Berry, M.J.A., Linoff, G.S.: *Data Mining Techniques - For Marketing, Sales, and Customer Relationship Management*, Second Edition, Wiley Publishing, Inc., Indianapolis, 2004.
2. Bramer, M.: *Principles of Data Mining (Undergraduate Topics in Computer Science)*, Springer-Verlag, London, 2007.
3. Buć, S.: *Model upravljanja rizicima kod javnih projekata*, magistarski rad, Sveučilište u Zagrebu, Fakultet organizacije i informatike Varaždin, Varaždin, 2007.
4. Goodwin, P., Wright, G.: *Decision Analysis for Management Judgment*, Third Edition, John Wiley & Sons Ltd, Chichester, 2004.
5. Han, J., Kamber, M.: *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann Publishers, San Francisco, 2006.
6. Joos, P., Vanhoof, K., Ooghe, H., Sierens N.: *Credit Classification: A Comparison of Logit Models and Decision Trees*, In: Nakhaezadeh, G., Steurer, E. (eds.): *Proceedings of the Workshop on Application of Machine Learning and Data Mining in Finance*, 10th European Conference on Machine Learning, Chemnitz, Germany, 1998., str. 59. - 72.
7. Kvesić, Lj.: *Modeli kreditnih rizika za predviđanje statusa neispunjavanja obveza*, doktorska disertacija, Sveučilište Josipa Jurja Strossmayera u Osijeku, Ekonomski fakultet, Osijek, 2010.
8. Larose, D.T.: *Discovering Knowledge in Data - An Introduction to Data Mining*, John Wiley & Sons, Inc., Hoboken, 2005.
9. Pang, S.-L., Gong, J.-Z.: *C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks*, *Systems Engineering - Theory & Practice*, Vol. 29, No. 12, 2009., str. 94. - 104.
10. Satchidananda, S.S., Simha, J.B.: *Comparing Decision Trees with Logistic Regression for Credit Risk Analysis*, SAS APaugC, Mumbai, 2006., http://abibasystems.com/white_paper/credit_risk.pdf
11. Xiao, W., Zhao, Q., Fei, Q.: *A Comparative Study of Data Mining Methods in Consumer Loans Credit Scoring Management*, *Journal of Systems Science and Systems Engineering*, Vol. 15, No. 4, 2006., str. 419. - 435.
12. Yap, B.H., Ong, S.H., Husain, N.H.M.: *Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models*, *Expert Systems with Applications*, Vol. 38, No. 10, 2011., str. 13274. - 13283.
13. Zenzerović, R., Peruško, T.: *“Kratki osvrt na modele za predviđanje stečaja”*, *Economic Research*, Vol. 19, No. 2, 2006., str. 132. - 151.
14. Zhang, D., Zhou, X., Leung, S.C.H., Zheng, J.: *Vertical Bagging Decision Trees Model for Credit Scoring*, *Expert Systems with Applications*, Vol. 37, No. 12, 2010., str. 7838. - 7843.
15. Zurada, J.: *Rule Induction Methods for Credit Scoring*, *Review of Business Information Systems*, Vol. 11, No. 2, 2007., str. 11. - 22.
16. Zurada, J., Zurada, M.: *How Secure Are “Good Loans”: Validating Loan-Granting Decisions And Predicting Default Rates On Consumer Loans*, *The Review of Business Information Systems*, Vol. 6, No.3, 2002., str. 65. - 83.

Ljiljanka Kvesić

Original scientific paper

APPLICATION OF DECISION TREES IN CREDIT SCORING

ABSTRACT

Banks are particularly exposed to credit risk due to the nature of their operations. Inadequate assessment of the borrower directly causes losses. The financial crisis the global economy is still going through has clearly shown what kind of problems can arise from an inadequate credit policy. Thus, the primary task of bank managers is to minimise credit risk. Credit scoring models were developed to support managers in assessing the creditworthiness of borrowers. This paper presents the decision tree based on exhaustive CHAID algorithm as one such model. Since the application of credit scoring models has not been adequately explored in the Croatian banking theory and practice, this paper aims not only to determine the characteristics that are crucial for predicting default, but also to highlight the importance of a quantitative approach in assessing the creditworthiness of borrowers.

Keywords: credit scoring, model, decision tree, exhaustive CHAID algorithm