

**Mariola Chrzanowska**

Wyższa Szkoła Ekonomii i Administracji w Kielcach

**Dorota Witkowska**

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

**ANALIZA EFEKTYWNOŚCI KLASYFIKACJI  
KREDYTOBIORCÓW UZYSKANEJ ZAGREGOWANYMI  
DRZEWAMI KLASYFIKACYJNYMI  
DLA PRÓB O RÓŻNEJ STRUKTURZE\***

**1. Wstęp**

Dynamiczny wzrost liczby kredytów udzielanych klientom indywidualnym wymaga odpowiednich narzędzi wspomagających podejmowanie decyzji kredytowych. Z jednej strony kredytodawcy są zainteresowani rozpoznawaniem klientów niewiarygodnych w celu minimalizacji ryzyka kredytowego, z drugiej strony zaś poszukują efektywnych metod umożliwiających rozpatrywanie wniosków w krótkim czasie. Stąd rosnące zainteresowanie metodami klasyfikacji.

Przy ocenie wiarygodności kredytowej zauważa się asymetrię danych, która – w przypadku dychotomicznej klasyfikacji – polega na tym, że liczebność jednego wariantu cechy znacznie przewyższa liczebność drugiego. Innymi słowy w bazie danych liczba klientów wiarygodnych jest na ogół znacznie wyższa od liczby klientów, którzy zostali zaliczeni przez ekspertów do grupy klientów niewiarygodnych<sup>1</sup>.

Przyczyny występowania takiego zróżnicowania liczebności obu typów klientów mogą być różne. Po pierwsze, można przyjąć, że w przypadku ubiegania się o kredyt z wnioskami występują przede wszystkim klienci przypuszczający, że ocena ich wiarygodności kredytowej wypadnie pomyślnie. Natomiast osoby przekonane,

---

\* Artykuł powstał w ramach projektu badawczego nr N113-012-32/1158 finansowanego ze środków Ministerstwa Nauki i Szkolnictwa Wyższego; tytuł projektu „Ocena ryzyka spłaty kredytów indywidualnych za pomocą wybranych metod statystycznych”.

<sup>1</sup> Problem ten jest znany i omawiany w literaturze przedmiotu. Maddala [2006] opisuje analizę wniosków kredytowych w Columbii, gdzie liczba zaakceptowanych wniosków była równa 4000, a liczba wniosków odrzuconych tylko 150. Z kolei w monografii [Gruszczyński 2001] opisano badanie 1000 wniosków kredytowych (700 dobrych i 300 złych) złożonych w Niemczech. Porównaj także pracę [Tian-Shyug i in. 2006].

że zostaną przez inspektora kredytowego ocenione negatywnie, w ogóle nie przystępują do procedury ubiegania się o kredyt. Po drugie, banki i inne instytucje finansowe chętnie zbierają i przechowują informacje o wiarygodnych klientach, są to bowiem potencjalni przyszli kredytobiorcy. Informacje zaś o klientach, którzy w procedurze oceny zdolności kredytowej wypadli negatywnie, nie są przechowywane. Po trzecie, banki niechętnie udzielają informacji na temat tzw. trudnych i straconych kredytów.

Dla banków podstawową kwestią jest poprawne rozpoznawanie klientów niewiarygodnych, pojawia się więc pytanie, czy asymetria informacji może powodować niższą jakość klasyfikacji dla grupy o mniejszej liczebności. Jeśli odpowiedź jest twierdząca, to kolejnym problemem do rozwiązania jest określenie sposobu postępowania w takim przypadku. Rozwiązaniem tego problemu może być wykorzystanie tzw. próby dobieranej. Jest to próba, która zawiera wszystkie obserwacje z grupy o mniejszej liczebności, a z klasy o większej liczebności dobiera się (losuje) elementy tak, aby liczebność obu grup była jednakowa (zob. [Gruszczyński 2001]). Oczywiście taka próba nie jest reprezentatywna, spełnia natomiast główny cel, tj. poprawia jakość klasyfikacji do grupy o mniejszej liczebności<sup>2</sup>. Jednakże, co jest wadą tego podejścia, następuje utrata informacji na temat obiektów należących do grupy o większej liczebności, ponieważ część obiektów jest eliminowana z próby<sup>3</sup>. Oprócz tego przy znacznej różnicy w liczebnościach obu grup, co ma często miejsce w analizach zdolności kredytowych, może się okazać, że mimo zasobnej w obiekty bazy danych, do analiz wykorzystuje się znacznie zredukowaną liczebnie próbę. A to z kolei może mieć istotny wpływ na jakość klasyfikacji, która zazwyczaj poprawia się wraz ze wzrostem liczby elementów w próbie.

Celem badań jest ocena wpływu struktury próby na jakość dychotomicznej klasyfikacji kredytobiorców. Szczególnie interesująca wydaje się odpowiedź na pytanie, czy identyfikacja niewiarygodnych klientów jest lepsza dla próby o jednakowej liczebności obiektów pochodzących z obu klas.

Klasyfikację przeprowadzono za pomocą drzew klasyfikacyjnych. Ze względu na stwierdzoną niestabilność rozwiązań uzyskanych dla pojedynczych drzew<sup>4</sup> w badaniach wykorzystano modele zagregowane<sup>5</sup>, stosując agregację bootstrapową, znaną jako *bagging* [Breiman 1996] i losowanie adaptacyjne – *boosting* [Freund, Schapire 1997].

---

<sup>2</sup> Maddala [1983] wykazał, że w przypadku modelu logitowego próba dobierana zmienia jedynie oszacowania wyrazu wolnego.

<sup>3</sup> Do budowy modeli klasyfikacyjnych wykorzystywane są tylko te obiekty, które zostaną wylosowane do próby (dobieranej). Obiekty, które nie zostaną wylosowane, są pomijane, a informacje o nich nie są uwzględniane w analizach.

<sup>4</sup> Por. [Breiman 1996; Breiman i in. 1998; Dietterich 1999; Gatnar, 2001 s. 120; Gatnar 2002].

<sup>5</sup> Wybór metod klasyfikacji nie jest przypadkowy. Za pomocą modeli zagregowanych często uzyskuje się lepsze wyniki niż za pomocą pojedynczych modeli. Breiman [1996] udowodnił, że *bagging* jest efektywnym algorytmem w przypadku, gdy małe zmiany w zbiorze uczącym powodują duże zmiany wyników predykcji. Por. [Cutler, Zhao 2001].

## 2. Opis badania

W badaniach wykorzystano dane dotyczące spłaty 2576 kredytów hipotecznych zaciągniętych w jednym z banków działających w Polsce. Po oczyszczeniu bazy z danych nietypowych skonstruowano trzy próby o różnej strukturze, ale o identycznej liczbie obiektów reprezentujących klientów zalegających ze spłatami. Próba A jest symetryczna (zbilansowana), tzn. liczebności kredytobiorców spłacających pożyczkę terminowo i zalegających ze spłatą kredytu są identyczne. Struktura próby B odpowiada sytuacji zaobserwowanej w Polsce w czerwcu 2006 r. (por. [Chmielewski i in. 2006, s. 17]), tzn. 9,4% próby to kredytobiorcy, którzy nie spłacają kredytu w terminie. Struktura próby C jest odzwierciedleniem udostępnionej przez bank bazy danych, w której 16% klientów nie spłaca kredytu w terminie.

Dostarczone przez bank dane charakteryzujące kredytobiorców oraz opisujące uzyskane kredyty zawierają informacje o 12 cechach ilościowych i jakościowych:

- $y$  – status kredytobiorcy (1 – klient spłaca kredyt w terminie; 0 – klient zalega ze spłatą kredytu),
- $x_1$  – grupa klienta (1 – klient standard; 0 – klient VIP),
- $x_2$  – płeć (0 – mężczyzna; 1 – kobieta),
- $x_3$  – waluta kredytu (0 – inne; 1 – zł),
- $x_4$  – wiek (w latach),
- $x_5$  – miejsce udzielenia kredytu (M1 – miasta powyżej 500 tys. mieszkańców, M2 – dawne miasta wojewódzkie i miasta liczące od 10 tys. do 50 tys. mieszkańców, M3 – miasta poniżej 10 tys. mieszkańców, o niskim poziomie rozwoju gospodarczego),
- $x_6$  – aktualnie spłacona kwota (w zł),
- $x_7$  – czas spłaty kredytu (w latach),
- $x_8$  – wartość kredytu (w zł),
- $x_9$  – posiadane lokaty (0 – do 1 roku, 1 – powyżej 1 roku),
- $x_{10}$  – aktualna stopa procentowa,
- $x_{11}$  – poprzednia stopa procentowa.

Po przeprowadzeniu badań pilotażowych dodatkowo uwzględniono transformacje zmiennych:

$$x_{12} = \frac{x_{10}}{x_{11}} \cdot 100,$$

$$x_{13} = \frac{x_6}{x_8} \cdot 100,$$

opisujące zmiany stopy procentowej oraz stan spłaty kredytu.

W eksperymentach wykorzystano zmienną grupującą  $y$  oraz 6 zmiennych diagnostycznych:  $x_2$ ,  $x_4$ ,  $x_5$ ,  $x_7$ ,  $x_{12}$ ,  $x_{13}$ , ponieważ wcześniejsze eksperymenty wykaza-

ły, że tylko wymienione zmienne posiadają zdolności dyskryminacyjne (por. [Chrzanowska, Witkowska 2007]). Badania przeprowadzono w trzech etapach.

W pierwszym wylosowano obiekty do trzech wcześniej zdefiniowanych prób: A, B i C. Utworzone próby podzielono na zbiory: uczący i testowy (tab. 1) w taki sposób, aby 68% obserwacji pochodzących z każdej podpróby (tj. klientów wiarygodnych i niewiarygodnych) wykorzystać do estymacji modeli (zbiór uczący), a pozostałe 32% obiektów potraktowano jako elementy zbioru testowego.

Tabela 1. Struktura poszczególnych prób

Oznaczenie prób	Próba A		Próba B		Próba C	
	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$
Wariant cechy $y$	136	136	1446	136	850	136
Zbiór uczący	64	64	681	64	400	64
Zbiór testowy	200	200	2127	200	1250	200

Źródło: opracowanie własne.

W drugim etapie dla każdej z trzech rozpatrywanych prób o różnej strukturze obiektów pochodzących z obu klas przeprowadzono szereg eksperymentów za pomocą drzew klasyfikacyjnych. Wykorzystano w tym celu modele zagregowane z wykorzystaniem algorytmów *bagging* i *boosting*. Wyznaczono rozwiązania dla zbiorów uczącego i testowego.

W ostatnim etapie przeprowadzono ocenę otrzymanych wyników oraz zbadano efektywność klasyfikacji za pomocą dwóch mierników:

1) błędu klasyfikacji obiektów do każdej (dla  $j = 1, 2$ ) grupy:

$$E_j = \frac{N_j}{N(j)} \cdot 100\%,$$

2) ogólnego błędu klasyfikacji:

$$E = \frac{\sum_{j=1}^2 N_j}{\sum_{j=1}^2 N(j)} \cdot 100\% = \frac{N_2 + N_1}{N} \cdot 100\%,$$

gdzie: dla  $j = 1, 2$ :  $N_j$  – liczba niepoprawnie rozpoznanych obiektów należących do klasy  $j$ -tej, tzn.  $N_1$  – liczba klientów banku, którzy nie spłacają kredytu, przez model zaś zostali rozpoznani jako spłacający kredyt,  $N_2$  – liczba klientów banku, którzy spłacają kredyt, a przez model zostali rozpoznani jako niespłacający kredyt,  $N(j)$  – liczebność klasy  $j$ -tej,  $N$  – liczba klientów banku  $N = N(1) + N(2)$  znajdujących się w zbiorze testowym lub uczącym.

Należy przy tym zauważyć, że o jakości klasyfikacji wnioskuje się głównie na podstawie błędów uzyskanych dla zbioru testowego (walidacyjnego) oraz poprzez porównanie błędów klasyfikacji w obu zbiorach. O poprawności modelu świadczy bowiem małe zróżnicowanie błędów wyznaczonych dla obu zbiorów. Warto wspomnieć, że z punktu widzenia banku podstawowe znaczenie ma błąd  $E_1$ , który pozwala ocenić skuteczność modelu przy rozpoznawaniu niewiarygodnych klientów.

### 3. Wyniki klasyfikacji

Wyniki klasyfikacji przedstawiono w tabelach zawierających liczebności obiektów zaliczanych do klas kredytów spłacanych i straconych, które zostały zaobserwowane i wygenerowane przez modele oraz wartości wykorzystanych mierników jakości klasyfikacji. Pogrubioną czcionką oznaczono liczebności poprawnie rozpoznanych w każdej klasie obiektów.

Wykorzystując algorytm *bagging*, dla każdej z prób zbudowano po 40 pojedynczych drzew klasyfikacyjnych. Następnie, stosując zasadę „głosowania większościowego”, każdy obiekt przypisano do jednej z grup: kredytów spłacanych ( $y = 1$ ) lub kredytów straconych ( $y = 0$ ). Wyniki badań dla zbioru uczącego i testowego przedstawiono w tab. 2.

Tabela 2. Wyniki klasyfikacji uzyskane dla drzew zagregowanych za pomocą algorytmu *bagging*

Rozwiązania modeli	Próba A		Próba B		Próba C	
	obserwacje		obserwacje		obserwacje	
Próba ucząca	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$y = 1$	106	52	1437	95	843	92
$y = 0$	30	84	9	41	7	44
<b>Razem</b>	136	136	1446	136	850	136
Błąd klasyfikacji	$E$	30,15	6,57	10,04		
	$E_1$	38,24	69,85	67,65		
	$E_2$	22,06	0,62	0,82		
Próba testowa	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$y = 1$	45	28	671	48	395	49
$y = 0$	19	36	10	16	5	15
<b>Razem</b>	64	64	681	64	400	64
Błąd klasyfikacji	$E$	36,72	7,79	11,64		
	$E_1$	43,75	75,00	76,56		
	$E_2$	29,69	1,47	1,25		

Źródło: obliczenia własne.

Analizując wyniki klasyfikacji (tab. 2), można zauważyć wyraźny wpływ liczebności próby na jakość klasyfikacji. Najlepsze bowiem rezultaty dla obu zbiorów uzyskano dla najbardziej licznej próby B (liczącej 2327 obserwacji, w tym w

zbiorze uczącym 1582 obiekty), a najgorsze dla próby B (zawierającej zaledwie 400 obserwacji, przy czym w zbiorze uczącym było ich 272).

Wyniki uzyskane w poszczególnych eksperymentach nie są w pełni satysfakcjonujące. Wprawdzie dla prób B i C ogólny błąd klasyfikacji dla zbioru uczącego wynosi odpowiednio 6 i 10%, jednak rozpoznawalność wiarygodnych klientów ( $100 - E_2$ ) przekracza 99%, poprawna klasyfikacja do grupy niewiarygodnych klientów (dla tych prób) jest już jednak znacznie niższa i wynosi niewiele ponad 30%. W przypadku próby symetrycznej rozpoznawalność obu grup klientów jest do siebie zbliżona i wynosi odpowiednio 78% dla  $y = 1$  oraz 62% dla  $y = 0$ , ale ogólny błąd klasyfikacji wynosi 30%.

Efektywność klasyfikacji można poprawnie ocenić na podstawie wyników uzyskanych dla zbioru testowego. Rozpoznawanie kredytów niespłaconych za pomocą modeli B i C w tym zbiorze nie przekracza 25%, a najlepsze wyniki (choć jakość klasyfikacji jest zbliżona do przypadkowej) uzyskano dla próby A – 56% poprawnie zaklasyfikowanych klientów. Ogólne błędy klasyfikacji w zbiorze testowym wahały się od 8% dla zbioru B do 37% dla zbioru A. Biorąc pod uwagę zróżnicowanie wartości błędów wyznaczonych dla obu zbiorów, należy zauważyć, że jest ono podobne dla wszystkich trzech prób, co świadczy o zdolnościach prognostycznych modelu.

W dalszym postępowaniu agregację modeli przeprowadzono metodą *boosting* na podstawie 200 pojedynczych drzew skonstruowanych dla każdej z prób. Wyniki klasyfikacji przedstawiono w tab. 3.

Tabela 3. Wyniki klasyfikacji uzyskane za pomocą algorytmu *boosting*

Rozwiązania modeli	Próba A		Próba B		Próba C	
	obserwacje		obserwacje		obserwacje	
Próba ucząca	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$y = 1$	124	23	1443	65	817	92
$y = 0$	12	113	3	71	33	44
Razem	136	136	1446	136	850	136
Błąd klasyfikacji	$E$	12,87	4,3	12,68		
	$E_1$	16,91	47,79	67,65		
	$E_2$	8,82	0,21	3,88		
Próba testowa	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$y = 1$	57	12	676	39	398	35
$y = 0$	7	52	5	25	2	29
Razem	64	64	681	64	400	64
Błąd klasyfikacji	$E$	14,84	5,91	7,97		
	$E_1$	18,75	60,94	54,69		
	$E_2$	10,94	0,70	0,50		

Źródło: obliczenia własne.

Grupowanie przeprowadzone z wykorzystaniem modeli zagregowanych za pomocą algorytmu *boosting* obarczone jest mniejszym błędem. Za pomocą najlep-

szego z modeli (próbą A) poprawnie rozpoznano w zbiorze uczącym 83% kredytów straconych i 91% kredytów spłacanych terminowo, a ogólny błąd klasyfikacji wyniósł 13%. W zbiorze testowym błąd ten wzrósł do 15%, ale model rozpoznał 81% kredytów zaległych oraz 89% kredytów spłacanych w terminie. W przypadku modelu uzyskanego dla próby B poprawnie zaklasyfikowano 39% kredytów straconych w zbiorze testowym, przy ogólnym błędzie klasyfikacji w wysokości 6% (dla porównania w zbiorze uczącym było to odpowiednio 52 i 4%). Natomiast klasyfikacja kredytów straconych na podstawie modelu oszacowanego dla próby C była poprawna w 45% w zbiorze testowym oraz w 32% w zbiorze uczącym.

#### 4. Wnioski

Przeprowadzone eksperymenty potwierdziły, że struktura próby istotnie wpływa na jakość klasyfikacji. Im bardziej próba jest zbilansowana (symetryczna), tym lepszą jakość klasyfikacji uzyskiwano dla obu grup. Wraz z rosnącą dysproporcją lepiej klasyfikowane były także kredyty spłacane w terminie, które były liczniej reprezentowane w próbie, i coraz gorzej rozpoznawane kredyty stracone. Widoczne też było zmniejszanie się ogólnego błędu klasyfikacji w miarę wzrostu liczebności próby uczącej, najmniejsze bowiem ogólne błędy generowały wszystkie modele na podstawie najliczniejszej próby B, a najwyższe błędy obserwowano dla najmniej licznej próby A.

W przeprowadzonych eksperymentach najwyższą jakość klasyfikacji uzyskano dla drzew klasyfikacyjnych zagregowanych za pomocą algorytmu *boosting*. Dla prób A, B i C rozpoznano odpowiednio 81, 39 i 46% kredytów straconych w zbiorze testowym.

#### Literatura

- Breiman L., Friedman J.H., Ohlsen R.A., Stone Ch.J. (1998), *Classification and Regression Trees*, CRC, Washington D.C.
- Breiman L. (1996), *Bagging Predictors*, „Machine Learning” nr 26, s. 123-140.
- Chmielewski, T., Głogowski A., Gołajewska M., Grodzicki M., Hałaj G., Imielska M., Kozak S., Zajączkowski S., Zochowski D. (2006), *Przegląd stabilności systemu finansowego, I półrocze 2006*, NBP, Warszawa.
- Chrzanowska M., Witkowska D. (2007), *Drzewa klasyfikacyjne w rozpoznawaniu kredytobiorców*, „Metody ilościowe w badaniach ekonomicznych”, Wydawnictwo SGGW, Warszawa, s. 291-300.
- Cutler A., Zhao G. (2001), *PERT – Perfect Random Tree Ensembles*, „Computing Science and Statistics”, 33, /12001Proceedings/ACutler/ACutler.pdf.
- Dietterich G.T. (1999), *An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting and Randomization*, „Machine Learning”, s. 1-22.
- Freund Y., Schapire R.E. (1997), *A Decision-theoretic Generalization of On-line Learning and an Application to Boosting*, „Journal of Computer and System Sciences”, 55, s. 119-139.

- Gatnar E. (2001), *Nieparametryczna metoda dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Gatnar E. (2002), *Agregacja modeli dyskryminacyjnych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 942, AE, Wrocław, s. 217-226.
- Gruszczyński M. (2001), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Szkoła Główna Handlowa, Warszawa.
- Maddala G.S. (2006), *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa.
- Maddala, G.S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, „Econometric Society Monographs in Quantitative Economics” Cambridge University Press, Cambridge.
- Tian-Shyug L., Chih-Chou Chi., Yu-Chao Ch., Chi-Jie L. (2006), *Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines*, „Computational Statistics & Data Analysis” nr 50, s. 1113-1130.

## **EFFICIENCY ANALYSIS OF THE BORROWERS CLASSIFICATION OBTAINED BY AGGREGATED CLASSIFICATION TREES FOR DIFFERENT STRUCTURE OF SAMPLE**

### **Summary**

In the paper we present the results of investigation that are conducted for the dichotomous classification of borrowers, employing actual data from the bank. For our research we construct several classification models applying ensemble classification. These models are built on the basis of 3 samples with different structure of credible and incredible clients. The first sample is symmetric (i.e. count of representatives of both groups is the same), the second one contains 9.4% incredible clients (as it was observed in Poland in the first two quarters of 2006), and the third sample contains 16% of noncreditworthy clients (as it was observed in the investigated bank). The aim of the investigation is to evaluate if different structure of the sample influences the quality of the classification.