Piotr Jardanowski

OCENA KONDYCJI EKONOMICZNEJ PRZEDSIĘBIORSTW: ANALIZA DYSKRYMINACYJNA A DRZEWA KLASYFIKACJI I REGRESJI (CART)

Wstęp

Wyobraźmy sobie pewną zbiorowość, o której posiadamy informacje, że:

- Składa się z N – obiektów, które przynależą do jednej z grup: G_0 albo G_1;
- Każdy obiekt jest opisany za pomocą $K+1$ – zmiennych;
- Istnieje jedna zmienna x_{K+1}, która pozwala, w zależności od wartości jakie przyjmuje, podzielić całą zbiorowość na dwie rozłączne populacje (grupy) G_0 i G_1;
- Znane są prawdopodobieństwa a priori[1] – π_0 i π_1 przynależności obiektów do poszczególnych populacji;
- Ustalone są koszty tzw. błędnej klasyfikacji obiektu[2] c_{01} (koszt błędnego zaklasyfikowania podmiotu z grupy G_0 do G_1) i c_{10} (koszt błędnego zaklasyfikowania podmiotu z grupy G_1 do G_0).

Wyobraźmy sobie teraz, że w zbiorowości pojawił się nowy obiekt, opisany za pomocą K zmiennych. Brakująca zmienna to ta, na podstawie której dokonywany jest podział naszej zbiorowości na dwie populacje. Naszym zadaniem jest podjęcie decyzji, czy nowy obiekt powinien się znaleźć w grupie G_0 czy G_1.

Podjęcie decyzji o zaklasyfikowaniu obiektu może nastręczać wiele kłopotów. Pomocą w przeprowadzeniu klasyfikacji służyć mogą metody wielo-

---

[1] Prawdopodobieństwo to jest traktowane jako informacja o względnych częstościach, z którymi obiekty przynależą do populacji. Zakłada się również, że wielkości π_0 i π_1 są znane. Najczęściej przyjmuje się pewne szacunkowe wielkości, które wynikają np. z obserwacji gospodarki, albo zakłada się, że ich wielkości są zgodne z udziałem poszczególnych grup w całym zbiorze, który podlega badaniu.

[2] Koszty te są przyjmowane przez badacza i zależą od jego intuicji. W badaniach bardzo często poprzez ich zmianę poszukuje się modelu o mniejszym błędzie.

wymiarowej analizy danych (WAD). W niniejszym artykule postanowiono sprawdzić:

1. Jak z klasyfikacją obiektów poradzą sobie:
   - liniowa funkcja dyskryminacyjna,
   - drzewo klasyfikacji i regresji, zbudowane za pomocą algorytmu CART?
2. Jaki wpływ na klasyfikację ma zmiana kosztu błędnej klasyfikacji?

Liniowa funkcja dyskryminacyjna

Analiza dyskryminacyjna wykorzystuje do klasyfikacji obiektów m.in. tzw. bayesowską regułę klasyfikacyjną. Punktem wyjścia bayesowskiej reguły jest założenie, że obiekt powinien być zaklasyfikowany do tej z rozważanych grup, do której „z większym prawdopodobieństwem" należy. Zdefiniujmy prawdopodobieństwa a posteriori przynależności obiektu opisanego wektorem x do populacji G_0:

$$\gamma_0(x) = \frac{\pi_0 f_0(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)};$$

oraz obiektu opisanego wektorem x do populacji G_1:

$$\gamma_1(x) = \frac{\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}.$$

gdzie $f_0(x)$ i $f_1(x)$ to funkcje gęstości rozkładów wektorów zmiennych losowych opisujących obiekty odpowiednio w grupie G_0 i grupie G_1.

Zgodnie z bayesowską regułą klasyfikacyjną:

- jeżeli $\gamma_0(x) < \gamma_1(x)$, to obiekt zostaje zaklasyfikowany do populacji G_1;
- jeżeli $\gamma_0(x) > \gamma_1(x)$, to obiekt zostaje zaklasyfikowany do populacji G_0[3].

[3] Ani w pierwszym przypadku, ani w drugim nie została uwzględniona sytuacja $\gamma_0(x) = \gamma_1(x)$. W takim przypadku można zastosować dwa rozwiązania:

1. na początku badania wprowadzić założenie, do której grupy zostaje zaklasyfikowany obiekt w takim przypadku;
2. można przyjąć występowanie strefy błędu (wątpliwej, niepewności) i wtedy obiekty, które się w niej znajdą, wymagają wykorzystania jeszcze innych metod w celu udzielenia odpowiedzi, do której grupy należą.

Zdefiniujmy następujący iloraz:

$$Z(x)=\frac{\gamma_0(x)}{\gamma_1(x)}=\frac{\dfrac{\pi_0 f_0(x)}{\pi_0 f_0(x)+\pi_1 f_1(x)}}{\dfrac{\pi_1 f_1(x)}{\pi_0 f_0(x)+\pi_1 f_1(x)}}=\frac{\pi_0 f_0(x)}{\pi_1 f_1(x)}.$$

Zgodnie z bayesowską regułą klasyfikacyjną:

a) jeżeli $Z(x)<1$, to obiekt zostaje zaklasyfikowany do populacji G_1;

b) jeżeli $Z(x)>1$, to obiekt zostaje zaklasyfikowany do populacji G_0.

Konkretna postać $Z(x)$ zależy od postaci funkcji $f_i(x)$. Jeżeli funkcje gęstości $f_0(x)$ oraz $f_1(x)$ są opisane K-wymiarowym rozkładem normalnym o [8, s. 213]:

– różnych wektorach wartości oczekiwanych w każdej grupie – $\mu_0 \neq \mu_1$;

– jednakowej macierzy wariancji-kowariancji w obu grupach – $\Sigma_0=\Sigma_1=\Sigma$;

to można, korzystając z postaci analitycznej funkcji gęstości rozkładu normalnego, zapisać funkcję $Z(x)$ w następujący sposób:

$$Z(x)=\frac{\pi_0 f_0(x)}{\pi_1 f_1(x)}=\frac{\pi_0}{\pi_1}\frac{(2\pi)^{0,5K}|\Sigma|^{0,5}\exp\left[-0,5(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right]}{(2\pi)^{0,5K}|\Sigma|^{0,5}\exp\left[-0,5(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right]}=$$

$$=\frac{\pi_0}{\pi_1}\exp\left[-0,5(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)+0,5(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right]=$$

$$=\frac{\pi_0}{\pi_1}\exp\left[(\mu_0-\mu_1)^T\Sigma^{-1}x\ -0,5(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1)\right].$$

gdzie:

Σ^{-1} – to odwrócona macierz wariancji – kowariancji;

μ_0 i μ_1 – to wektory wartości oczekiwanych zmiennych x_i ($i = 1, ..., K$) w grupach G_0 i G_1.

Jeżeli dokonamy podstawienia:

$$D(x)=(\mu_0-\mu_1)^T\Sigma^{-1}x\ -0,5(\mu_0-\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1),$$

a następnie wykorzystamy własności funkcji wykładniczej, to reguła klasyfikacyjna przyjmie wówczas postać:

obiekt, opisany wektorem x, dla którego zachodzi $D(x) > \ln\frac{\pi_1}{\pi_0}$, zostaje zaklasyfikowany do grupy G_0, a w przeciwnym wypadku do grupy G_1.

W sytuacji gdy uwzględniamy koszty błędnej klasyfikacji, reguła klasyfikacyjna ulega modyfikacji i przyjmuje postać:

obiekt, opisany wektorem x, dla którego zachodzi $D(x) > \ln\frac{\pi_1 c_{01}}{\pi_0 c_{10}}$, zostaje zaklasyfikowany do grupy G_0, a w przeciwnym wypadku do grupy G_1.

Właśnie równanie $D(x) = (\mu_0 - \mu_1)^T \Sigma^{-1} x - 0{,}5(\mu_0 - \mu_1)^T \Sigma^{-1}(\mu_0 + \mu_1)$ nazywa się liniową funkcją dyskryminacyjną. Przy posługiwaniu się tą funkcją należy pamiętać, że:

- u podstaw jej konstrukcji leży założenie, że zmienne opisujące obiekty mają K-wymiarowy rozkład normalny o jednakowej dyspersji w obu populacjach ($\Sigma_0 = \Sigma_1 = \Sigma$) i różnych średnich;
- jeżeli wektory średnich μ_0 oraz μ_1 w poszczególnych grupach będą do siebie bardzo zbliżone (statystycznie nie będą się różnić), to wartość różnicy $\mu_0 - \mu_1$ będzie wynosiła „0"; w konsekwencji wartość funkcji dyskryminacyjnej będzie, dla dowolnego x, wynosiła „0" i nie będzie możliwa klasyfikacja obiektów;
- aby obliczyć wartość funkcji dyskryminacyjnej musimy dokonać operacji odwrócenia macierzy wariancji-kowariancji zmiennych, co jest możliwe pod warunkiem, że wyznacznik tej macierzy wariancji-kowariancji jest różny od „0"; aby tak było, zmienne opisujące obiekty nie mogą być ze sobą skorelowane.

Tworzenie drzewa klasyfikacji i regresji

Procedura CART (Classification And Regression Tree) jest jedną z kilku procedur służących do tworzenia drzew, za pomocą których przeprowadza się klasyfikację obiektów. Została wprowadzona na grunt finansów przez H. Frydman, E. I. Altmana, D. Kao w artykule *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*. Jest ona stosunkowo prosta i szybko zyskała sobie popularność. Wcześniej była z powodzeniem stosowana w innych dyscyplinach naukowych.

Przy tworzeniu drzewa za pomocą procedury CART (Classification And Regression Tree) korzysta się z algorytmu rekurencyjnego podziału. Na każdym

etapie tworzenia drzewa przeprowadza się pewien cykl obliczeń. Na ich podstawie wyciąga się wnioski, które powodują zatrzymanie algorytmu bądź ponowne (rekurencyjne) wykonanie obliczeń. Poniżej zostaną przedstawione kroki, które wykonuje się na każdym etapie budowy drzewa za pomocą procedury CART. Drzewo składa się tzw. węzłów i łączących je gałęzi. Jego tworzenie polega na dzieleniu kolejnych węzłów, na dwa inne, możliwie homogeniczne węzły. Podziału węzła dokonuje się wzdłuż tzw. osi podziału i to właśnie jej znalezienie jest celem procedury CART. Kolejne kroki poszukiwania osi podziału przebiegają następująco:

1. Wybieramy dowolną zmienną x_j i ustalamy jej wartość dla poszczególnych obiektów.

2. Porządkujemy wartości tej zmiennej od najmniejszej do największej: $x_{j1} < x_{j2} < \ldots < x_{jN}$.

3. Ustalamy N–1 potencjalnych osi podziału. Za potencjalną oś podziału przyjmujemy średnią arytmetyczną sąsiednich, uporządkowanych rosnąco, wartości zmiennej.

4. Dzielimy węzeł na N–1 sposobów, zgodnie z potencjalnymi osiami podziału.

5. Celem określenia przynależności nowo powstałych węzłów do grupy G_0 lub G_1 obliczamy tzw. ryzyko resubstytucji, które dla węzła t opisane jest wzorem:

$$R_i(t) = c_{ji}\, p(j,t) = c_{ji}\pi_j p(t/j) = c_{ji}\pi_j \frac{n_j(t)}{N_j}$$

gdzie:

$p(j,t)$ – oznacza prawdopodobieństwo, że obiekt pochodzący z grupy j wpadnie do węzła t;

$p(t/j) = \frac{n_j(t)}{N_j}$ – oznacza prawdopodobieństwo warunkowe, że obiekt z grupy j wpadnie do węzła t;

$n_j(t)$ – liczba podmiotów z grupy j, która wpadła do węzła t;

c_{ij} – koszt błędnej klasyfikacji obiektu;

π_i – prawdopodobieństwo *a priori*.

Widać wyraźnie, że ryzyko resubstytucji zależy zarówno od kosztów błędnej klasyfikacji obiektu c_{ij}, jak i prawdopodobieństwa *a priori*. Ryzyko to jest utożsamiane z kosztem błędnej klasyfikacji węzła.

6. Przyporządkowujemy nowe węzły do jednej z dwóch badanych grup G_0 lub G_1 wg reguły:

$$R(t) = \min\{R_0(t), R_1(t)\}.$$

Jest to tzw. bayesowska reguła klasyfikacyjna. Jej zrozumienie jest bardzo intuicyjne. Ponieważ $R_i(t)$ reprezentuje koszt błędnej klasyfikacji węzła, dlatego badaczowi zależy, aby dany węzeł przydzielić do tej grupy, dla której koszt ten jest minimalny.

7. Obliczamy zanieczyszczenie I(t) nowo powstałych węzłów. Jest ono zdefiniowane w następujący sposób:

$$I(t) = R_0(t)\,p(0\,|\,t) + R_1(t)\,p(1\,|\,t).$$

Wykorzystując wcześniej podane wzory na $R_0(t)$ i $R_1(t)$ i dokonując odpowiednich przekształceń, dochodzimy do następującej postaci analitycznej:

$$I(t) = (c_{01} + c_{10})\,p(0\,|\,t)\,p(1\,|\,t)\,p(t),$$

gdzie:

$p(i\,|\,t)$ – prawdopodobieństwo warunkowe, że obiekt w węźle t pochodzi z grupy i (dla $i = 0,1$) i definiuje się je wzorem:

$$p(i\,|\,t) = \frac{\dfrac{\pi_i n_i(t)}{N_i}}{\sum_{k=0}^{1} \dfrac{\pi_k n_k(t)}{N_k}};$$

$p(t)$ – prawdopodobieństwo, że obiekt z próby wpadnie do węzła t i jest ono zdefiniowane w następujący sposób:

$$p(t) = \sum_{k=0}^{1} p(k,t) = \sum_{k=0}^{1} \frac{\pi_k n_k(t)}{N_k}.$$

8. Obliczamy spadek zanieczyszczenia:

$$\Delta I(t) = I(t) - I(t_R) - I(t_L)$$

gdzie:

$I(t)$ – zanieczyszczenie węzła t, który będzie dzielony na dwa węzły t_R i t_L;

$I(t_R)$ i $I(t_L)$ – zanieczyszczenia nowo powstałych węzłów – odpowiednio prawego i lewego.

9. Wykonujemy czynności 5–9 dla kolejnych potencjalnych osi podziału danej zmiennej.

10. Powtarzamy czynności 1–10 dla kolejnych zmiennych.

11. Wybieramy taką oś podziału, dla której spadek zanieczyszczenia jest największy:

$$\max \Delta I(t).$$

Reguła wyboru osi podziału mówi, że najlepszy podział zbioru, znajdującego się w danym węźle, uzyskuje się dla maksymalnej wartości $\Delta I(t)$, czyli maksymalnego spadku zanieczyszczenia węzła.

W efekcie dokonywany jest podział węzła na dwa nowe węzły, które należy podzielić w kolejnej iteracji wykonując opisane wyżej kroki 1–11. Widać wyraźnie, że działania wykonywane są rekurencyjnie i stąd właśnie nazwa „algorytm rekurencyjnego podziału".

Podziału kolejnych węzłów możemy dokonywać do momentu, aż wszystkie węzły będą homogeniczne. Jednak na pewnym etapie budowane drzewo przestanie być czytelne, a korzystanie z niego stanie się bardzo trudne. Celem ustrzeżenia się przed nadmierną komplikacją stosuje się tzw. reguły stopu. Do podstawowych zaliczamy:

1. Minimum $n(t)$ – reguła ta polega na określeniu minimalnej liczby obiektów w węźle, który może być dalej dzielony. RPA zostaje zatrzymane, gdy liczba podmiotów w węźle jest mniejsza lub równa $n(t)$. Wielkość $n(t)$ określa badacz.

2. Brak znaczącego spadku zanieczyszczenia węzła – jeżeli w wyniku podziału węzła spadek zanieczyszczenia $-\Delta I(t)$ – jest bardzo mały, wtedy rezygnuje się z dzielenia tego węzła.

3. Maksymalny poziom – badacz może określić, z jak wielu poziomów ma się składać budowane drzewo. Algorytm zostaje zatrzymany, gdy dotrzemy do odpowiedniego poziomu.

Ostatnim etapem budowania drzewa jest wybór („wycięcie") odpowiedniego poddrzewa. Takie działanie ma na celu uniknięcie sytuacji, zwanej „przeuczeniem drzewa". Przeuczenie jest skutkiem dzielenia kolejnych węzłów, mimo że wynikający z tego spadek zanieczyszczenia jest bardzo mały. Drzewo „przeuczone" doskonale klasyfikuje obiekty, na podstawie których zostało utworzone, ale zupełnie nie radzi sobie z klasyfikacją nowych elementów.

W swoim artykule H. Frydman, E. I. Altman, D. Kao, jako metodę „wycięcia" poddrzewa zaproponowali rozwiązanie następującego zadania decyzyjnego:

$$R(t) + K \times NT \longrightarrow \min,$$

gdzie:

$R(t)$ – ryzyko resubstytucji dla całego drzewa,

NT – liczba węzłów końcowych,

$K \geq 0$ – kara za kompleksowość (nadmierne rozbudowanie drzewa).

Na szczególną uwagę zasługuje nieujemna stała K. W sytuacji gdy jej wartość będzie wynosiła „0", rozwiązaniem powyższego zadania będzie „przeuczone" drzewo, otrzymane w wyniku działania algorytmu. Jednak jakakolwiek wartość dodatnia będzie powodowała „wycięcie" poddrzewa. Konsekwencją takiego

działania będzie powstanie całej grupy poddrzew, z których każde kolejne będzie się charakteryzowało:

- zmniejszeniem kompleksowości,
- większą przejrzystością i łatwiejszym korzystaniem z niego,
- zmianą ryzyka resubstytucji.

Przy odpowiednio dużej wartości stałej K zbiór wejściowy nie ulegnie podziałowi.

Ostatnim etapem jest wybór optymalnego drzewa spośród powstałej grupy poddrzew. Wybór ten służy określeniu poziomu drzewa, do którego następuje wyraźny spadek kosztu błędnej klasyfikacji – $R(T)$, określanego jako suma kosztów błędnej klasyfikacji wszystkich węzłów ostatecznych (niepodlegających dalszemu podziałowi). Takie postępowanie wynika z faktu, że w miarę dzielenia $R(T)$ najpierw spada bardzo szybko, a następnie osiąga pewien minimalny poziom, wokół którego oscyluje. Pomimo wykonywania kolejnych podziałów, nie uzyskuje się spadku $R(T)$ i dlatego kolejne podziały są zbędne.

Oceny $R(T)$ dla kolejnych drzew dokonuje się za pomocą testu V-krotnej walidacji krzyżowej (ang. V-fold cross validation test). „Technika walidacji skrośnej[4] polega na losowym podziale[5] posiadanego zbioru danych na dwie części: uczącą i testującą. Pierwszego podzbioru używamy do stworzenia modelu, drugiego zaś do jego testowania". [12, s.14]. Dokładniej rzecz ujmując, zbiór – znajdujący się w rdzeniu – jest dzielony na V podzbiorów. V-1 grup jest używanych do wytworzenia drzewa, V-ta do jego testowania. Następnie procedurę powtarza się V razy, za każdym razem wybierając inny zbiór jako testujący, a z pozostałych buduje się drzewo. W ten sposób otrzymujemy V różnego rodzaju ryzyka resubstytucji dla drzewa. Obliczamy z nich średnią arytmetyczną i uzyskany wynik przyjmujemy za zanieczyszczenie dla wyciętego poddrzewa. „Jeżeli budowane drzewa bazowały na szumie, a nie na istotnej informacji, wynik walidacji skrośnej będzie bardzo słaby" [12, s. 14].

Liniowa funkcja dyskryminacyjna a drzewa klasyfikacji i regresji

Intuicyjne różnice

Chociaż liniowa funkcja dyskryminacyjna oraz drzewa klasyfikacji i regresji są oparte na bayesowskiej regule klasyfikacyjnej, której celem jest minimalizacja

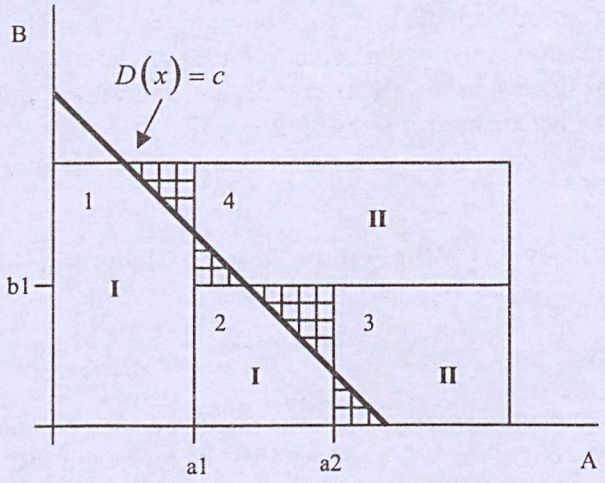
[4] W literaturze spotyka się różne nazwy testu m.in. walidacja krzyżowa i walidacja skrośna.

[5] Przez losowy podział należy rozumieć sytuację, w której żadna reguła nie decyduje o tym, w której grupie znajdzie się dany obiekt. Jest to podział czysto przypadkowy.

kosztu błędnej klasyfikacji, można wskazać pewne różnice w podejściu do dyskryminacji.

Pierwszą różnicą, którą daje się na wstępie zauważyć, są założenia. Zmienne, które wykorzystuje się do szacowania liniowej funkcji dyskryminacyjnej, powinny spełniać założenia wymienione w punkcie 2. niniejszego artykułu. Zmienne wykorzystywane do zbudowania drzewa za pomocą algorytmu RPA nie muszą spełniać żadnych założeń.

Drugą różnicą, którą można wskazać pomiędzy analizą dyskryminacyjną i RPA, jest sposób, w jaki obie techniki dzielą przestrzeń zmiennych na obszary klasyfikacyjne. Rozważmy hipotetyczną liniową funkcję dyskryminacyjną $D(x)$. Reguła klasyfikacyjna jest następująca: obiekt, dla którego zachodzi $D(x) > \ln \frac{\pi_1 c_{01}}{\pi_0 c_{10}}$ zostaje zaklasyfikowany do grupy G_0, a przeciwnym wypadku do grupy G_1. Widać wyraźnie, że obszar klasyfikacji jest dzielony na dwie półpłaszczyzny odpowiadające grupom G_0 i G_1, za pomocą prostej $D(x) = c$, gdzie $c = \ln \frac{\pi_1 c_{01}}{\pi_0 c_{10}}$. Wszystkie obiekty, które znajdą się powyżej tej prostej będą należały do grupy G_0, a wszystkie poniżej – do grupy G_1. RPA natomiast dzieli przestrzeń na pewną liczbę prostokątnych obszarów, a każdy z nich oznacza jedną z dwóch grup przeznaczenia – G_0 lub G_1. Na rysunku 1 widać różnicę w podejściu do klasyfikacji obiektów w obu technikach.



Rysunek 1. Różnica w klasyfikacji liniowej funkcji dyksryminacyjnej i RPA

Źródło: H. Frydman, E. I. Altman, D. Kao, *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*, „The Journal of Finance" 1985, marzec, vol. 40, s. 277

Obiekty, które znajdą się w obszarach zaznaczonych kratką, zostaną przez każdą z metod zaklasyfikowane w odmienny sposób.

Trzecią różnicą pomiędzy analizowanymi metodami jest sposób, w jaki zmiana kosztu błędnej klasyfikacji wpływa na rezultaty badania. W przypadku liniowej funkcji dyskryminacyjnej zmiana tej wielkości powoduje przesunięcie punktu podziału, czyli wartości stałej „c". Natomiast w przypadku RPA zmiana kosztu błędnej klasyfikacji wpływa na podział każdego węzła. Wynika to z faktu, że przy każdym dzielonym węźle oblicza się zanieczyszczenie oraz ryzyko resubstytucji, których wartość zależy od tych kosztów. Zmiana tych wielkości ma wpływ na wybór osi podziału, a w konsekwencji na kształt drzewa.

Porównanie wyników klasyfikacji na podstawie liniowej funkcji dyskryminacyjnej oraz drzew klasyfikacji i regresji

Dane wykorzystane w badaniu

Próba[6], którą wykorzystano w badaniu, składała się z obiektów pochodzących z dwóch populacji G_0 i G_1 odpowiednio: przedsiębiorstw w złej kondycji finansowej („złe") oraz dobrej kondycji finansowej („dobre"). Podmioty zostały tak dobrane, aby każdemu podmiotowi w złej kondycji finansowej odpowiadał jeden w dobrej kondycji finansowej. Obie populacje zawierały po 33 przedsiębiorstwa[7].

Obiekty zostały opisane za pomocą wskaźników finansowych, obliczonych na podstawie sprawozdań finansowych. W przypadku podmiotów „złych" za podstawę obliczeń przyjęto sprawozdania finansowe z roku poprzedzającego rok, w którym zgłoszono wniosek o upadłość. Na podstawie sprawozdań finansowych oszacowano 21 wskaźników finansowych, które opisywały każdy badany podmiot. Brakujące dane zastąpiono średnią wartością wskaźnika w danej grupie[8]. Wskaźniki prezentuje tabela 1.

---

[6] Dane wykorzystane w badaniu pochodziły z projektu numer 5 HO2B 014 21, realizowanego w Katedrze Ekonometrii Akademii Ekonomicznej w Poznaniu pod tytułem *Analiza i prognozowanie upadłości przedsiębiorstw w Polsce.*

[7] W dalszej części przedsiębiorstwa będą również nazywane obiektami, podmiotami.

[8] Średnie szacowano na podstawie wartości wskaźnika w danej grupie G_0 lub G_1. Nie szacowano średniej na podstawie całej zbiorowości, ponieważ z założenia analizy dyskryminacyjnej przyjmuje się, że średnie powinny być jak najbardziej zróżnicowane. Jest to warunek wysokiej skuteczności funkcji dyskryminacyjnej.

Tabela 1. Wykaz $K = 21$ wskaźników wykorzystanych w badaniu

L.p.	Nazwa wskaźnika	Przyporządkowany symbol
1	Marża zysku brutto ze sprzedaży	W1
2	Marża zysku operacyjnego	W2
3	Marża zysku brutto	W3
4	Marża zysku netto	W4
5	Stopa zwrotu z kapitału własnego	W5
6	Stopa zwrotu z aktywów	W6
7	Kapitał pracujący	W7
8	Wskaźnik płynności bieżącej	W8
9	Wskaźnik płynności szybkiej	W9
10	Wskaźnik podwyższonej płynności	W10
11	Rotacja należności	W11
12	Rotacja zapasów	W12
13	Cykl Operacyjny	W13
14	Rotacja zobowiązań	W14
15	Cykl konwersji gotówki	W15
16	Rotacja aktywów obrotowych	W16
17	Rotacja aktywów	W17
18	Wskaźnik pokrycia majątku	W18
19	Stopa zadłużenia	W19
20	Wskaźnik dźwigni finansowej	W20
21	Dług / EBITDA	W21

Źródło: Opracowanie własne.

Przedstawione wskaźniki obrazują 4 najważniejsze, podlegające ocenie, obszary działalności przedsiębiorstwa, odpowiednio:

- wskaźniki rentowności – W1, W2, W3, W4, W5, W6;
- wskaźniki płynności – W7, W8, W9, W10;
- wskaźniki sprawności działania – W11, W12, W13, W14, W15, W16, W17;
- wskaźniki zadłużenia – W18, W19, W20, W21.

Porównywane modele

Celem porównania wyników klasyfikacji przedsiębiorstw oszacowano następujące modele oparte na liniowej funkcji dyskryminacyjnej:

1. Pierwszy z modeli został oszacowany metodą krokową w przód[9]. Estymacji dokonano przy następujących założeniach (por.: D. Hadasik [9, s. 152]):

- minimalna tolerancja: 0,05;
- poziom istotności: 5%;
- wartość statystyki F służącej do wprowadzania zmiennych do modelu: F = 1;
- ze zbioru zmiennych usunięty został wskaźnik W3, jako najbardziej skorelowany z pozostałymi zmiennymi;
- model nie może zawierać więcej niż 13 zmiennych[10].

W rezultacie przeprowadzonej analizy uzyskano następujący model ($DA_1(W)$):

$$DA_1(W) = -2{,}28951 \times W1 + 3{,}87841 \times W2 - 1{,}837 \times W4 - 0{,}03207 \times W7 + \\ + 0{,}66176 \times W8 + 0{,}09781 \times W9 - 1{,}89738 \times W10 - 1{,}67831 \times W12 + \\ + 0{,}45204 \times W17 + 1{,}02303 \times W18 - 0{,}16074 \times W20 - 0{,}00937 \times W21;$$

Dla modelu tego wartość statystyki lambda Wilksa[11] wynosi $\lambda = 0{,}43334$, co świadczy o dobrej zdolności dyskryminacyjnej.

2. Drugi model, który zostanie zaprezentowany, został oszacowany zupełnie inaczej. W trakcie szacowania drzew decyzyjnych, wyselekcjonowano zmienne, które najlepiej różnicowały populacje. Drugi model liniowej funkcji dyskryminacyjnej oszacowano wykorzystując te właśnie zmienne. Minimalna dopuszczalna tolerancja wynosiła 0,05. W efekcie otrzymano model ($DA_2(W)$) postaci:

$$DA_2(W) = -0{,}40133 \times W1 + 1{,}996 \times W2 + 0{,}20428 \times W5 + 0{,}14978 \times W6 + \\ + 0{,}01006 \times W8 + 1{,}85554 \times W9 - 2{,}14632 \times W10 + 0{,}84722 \times W18 + \\ + 0{,}1533 \times W\,19 - 0{,}00916 \times W21 - 1{,}75103.$$

Otrzymany w ten sposób model charakteryzował się statystyką lambda Wilksa $\lambda = 0{,}53326$, świadczy to o tym, że model ma gorsze zdolności dyskryminacyjne niż $DA_1(W)$.

---

[9] Szacowanie modelu liniowej funkcji dyskryminacyjnej za pomocą analizy krokowej jest zbliżone do podejścia jakie stosuje się w RPA. Wpływ każdej zmiennej na model jest analizowany oddzielnie. Takie oszacowanie modelu umożliwi porównanie wyników z rezultatami uzyskanymi z RPA.

[10] Zgodnie z sugestiami w literaturze na jedną zmienną w modelu powinno przypadać co najmniej 5 obserwacji.

[11] Im bliższy „0" jest ten wskaźnik, tym lepsze własności dyskryminacyjne ma oszacowany model.

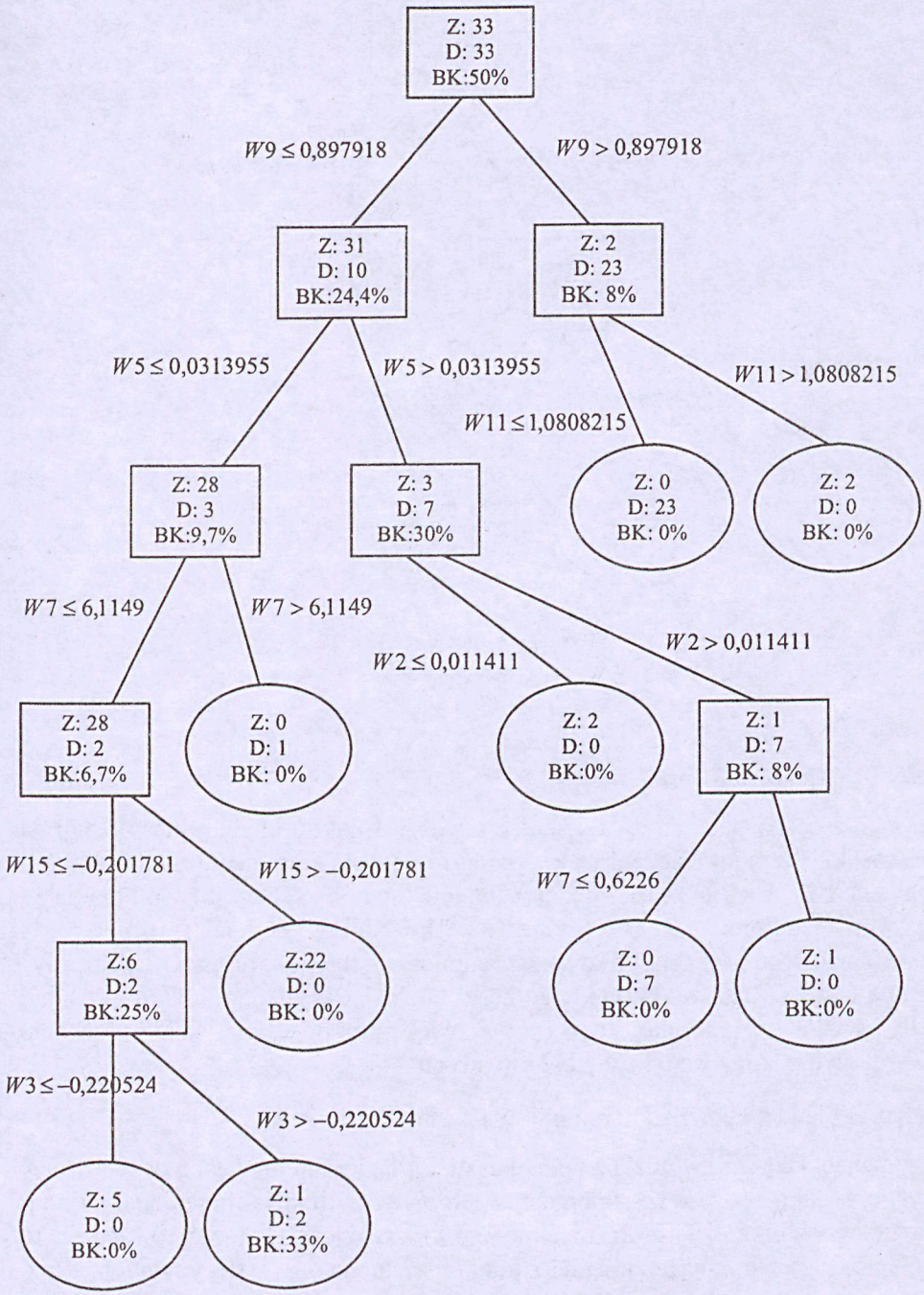
W niniejszym artykule wspomniano, że koszt błędnej klasyfikacji ma wpływ na podział węzła, a w konsekwencji na kształt drzewa. W związku z tym nie można, jak to miało miejsce w przypadku liniowej funkcji dyskryminacyjnej, przedstawić gotowego kształtu drzewa, gdyż będzie on ulegał zmianie wraz ze zmianą kosztów. Można natomiast określić zasady konstruowania drzewa. Pierwszy model (RPA1) był budowany przy zachowaniu następujących zasad:

- drzewo nie mogło mieć więcej niż 10 poziomów;
- węzeł był dzielony w sytuacji, gdy zawierał co najmniej 7 podmiotów;
- nie wycinano poddrzewa.

Uzyskane w ten sposób drzewo, dla równych kosztów błędnej klasyfikacji $c_{01} = c_{10} = 1$, prezentuje rysunek 2. Drzewo ma 6 poziomów i 9 węzłów ostatecznych. Poszczególne węzły opisane są za pomocą następujących informacji:

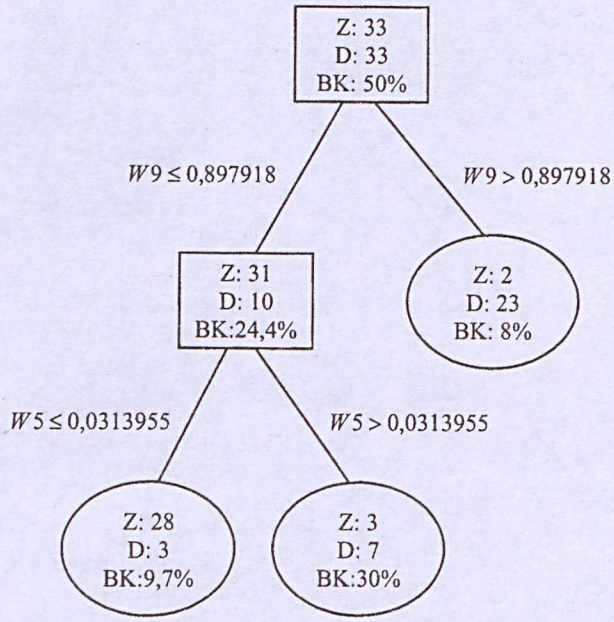
- Z – liczba podmiotów określonych jako „złe" (w złej kondycji finansowej);
- D – liczba podmiotów określonych jako „dobre" (w dobrej kondycji finansowej);
- BK – ogólny procent błędnie zaklasyfikowanych podmiotów w danym węźle.

Wykorzystany w badaniu, drugi model (RPA2) uzyskano za pomocą techniki V-fold Cross Validation, dokładniej: 10-krotnego testu walidacji krzyżowej. Wycięte poddrzewo przedstawia rysunek 3. Jest ono dużo mniej skomplikowane i czytelniejsze od drzewa RPA1. W sytuacji gdy próba badawcza składa się z 66 podmiotów, opisanych za pomocą 21 wskaźników, stosowanie drzewa RPA1 nie sprawia problemu i przewaga RPA2 nie jest zauważalna. Sytuacja uległaby jednak zdecydowanej zmianie, gdybyśmy mieli do czynienia z próbą 10 000 obiektów opisanych za pomocą 50 zmiennych.



Rysunek 2. Model RPA1

Źródło: Opracowanie własne



Rysunek 3. Model RPA2

Źródło: Opracowanie własne

Analiza wyników badania

Jednym z celów badania była symulacja wpływu kosztu błędnej klasyfikacji na jej wyniki. Przyjęto założenie, że z punktu widzenia klasyfikacji, większe konsekwencje negatywne może mieć popełnienie błędu klasyfikacji I typu (zaklasyfikowanie podmiotu „złego" (grupa G_0) jako „dobry" (grupa G_1)), niż błędu klasyfikacji typu II (zaklasyfikowanie podmiotu „dobrego" (grupa G_1) jako „zły" (grupa G_0)). W badaniu przyjęto, że koszt błędnej klasyfikacji podmiotu c_{10} był stały i wynosił 1, natomiast koszt c_{01} był zwiększany od 1 do 5. Miało to na celu „wymuszenie" mniejszej liczby błędów I typu.

a) Koszt błędnej klasyfikacji podmiotu a kształt drzewa

Analizując, jaki wpływ miała zmiana kosztu c_{01} na kształt drzewa, należy zwrócić uwagę na fakt, że we wszystkich drzewach pierwszy podział przebiegał na podstawie wskaźnika *W9* (wskaźnika płynności szybkiej). Oznacza to, że jest to najlepiej dyskryminująca zmienna z punktu widzenia drzew decyzyjnych. Również wartość wskaźnika, według której dokonywano pierwszego podziału, była stała, aż do kosztu $c_{01} = 1{,}6$. Dopiero w ostatnim drzewie wartość ta zwiększyła się. Wielkość drzewa również ulegała zmianie. W przypadku modelu RPA1

pierwsze drzewo było bardzo rozbudowane (rysunek 2). Każde zwiększenie kosztu powodowało, że drzewo traciło jeden podział, począwszy od samego dołu. Z kolei przy modelu RPA2 pierwsze drzewo składało się zaledwie z dwóch podziałów (rysunek 3). Zostało ono ograniczone do jednego podziału przy koszcie $c_{01} = 1{,}15$. Natomiast wzrosła wartość osi podziału, gdy koszt osiągnął poziom $c_{01} = 1{,}6$. Widać wyraźnie, że zgodnie z wcześniejszym opisem, procedura CART jest czuła na zmiany kosztu błędnej klasyfikacji podmiotu. Dotyczy to nie tylko wielkości i kształtu drzewa, ale również położenia osi podziału.

b) Koszt błędnej klasyfikacji podmiotu a trafność klasyfikacji

Wpływ zmiany kosztu na trafność klasyfikacji przedstawia tabela 2.

Tabela 2. Wpływ zmiany kosztu c_{01} na klasyfikację badanych firm

Model	C_{01}																				
	1			1,04			1,08			1,15			1,2			1,6			3		
	I[12]	II	T	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T	I	II	T
RPA 1	3%	0%	2%	0%	6%	3%	0%	9%	5%	0%	9%	5%	0%	9%	5%	0%	30%	15%	0%	45%	23%
RPA 2	15%	9%	12%	15%	9%	12%	6%	30%	18%	6%	30%	18%	6%	30%	18%	6%	30%	18%	0%	45%	23%
DA 1	12%	12%	12%	9%	15%	12%	6%	18%	12%	6%	21%	14%	3%	24%	14%	0%	30%	15%	0%	52%	26%
DA 2	21%	18%	20%	21%	21%	21%	21%	21%	21%	15%	21%	18%	15%	21%	18%	0%	42%	21%	0%	61%	30%

Źródło: Opracowanie własne.

W analizie posłużono się syntetyczną miarą charakteryzującą liczbę popełnionych błędów (I i II typu) przy klasyfikacji podmiotów na podstawie poszczególnych modeli. Miarę nazwano ryzykiem resubstytucji modelu i określono następująco:

$$RS = \frac{\pi_0 c_{01} n_{01}}{N_0} + \frac{\pi_1 c_{10} n_{10}}{N_1},$$

gdzie:

n_{01}, n_{10} – to liczba błędnych klasyfikacji odpowiednio typu I i typu II,

N_0, N_1 – to liczebność grupy odpowiednio „złe" i „dobre".

Na podstawie wartości RS stwierdzono (por.: tab. 2), że dla wszystkich wartości kosztu c_{01} zdecydowanie najlepiej klasyfikuje model RPA1. Do wielkości

[12] Zapis „I" oznacza błąd klasyfikacji typu I – zaklasyfikowanie podmiotu „złego" jako „dobry", natomiast „II" oznacza błąd klasyfikacji typu II – zaklasyfikowanie podmiotu „dobrego" jako „zły".

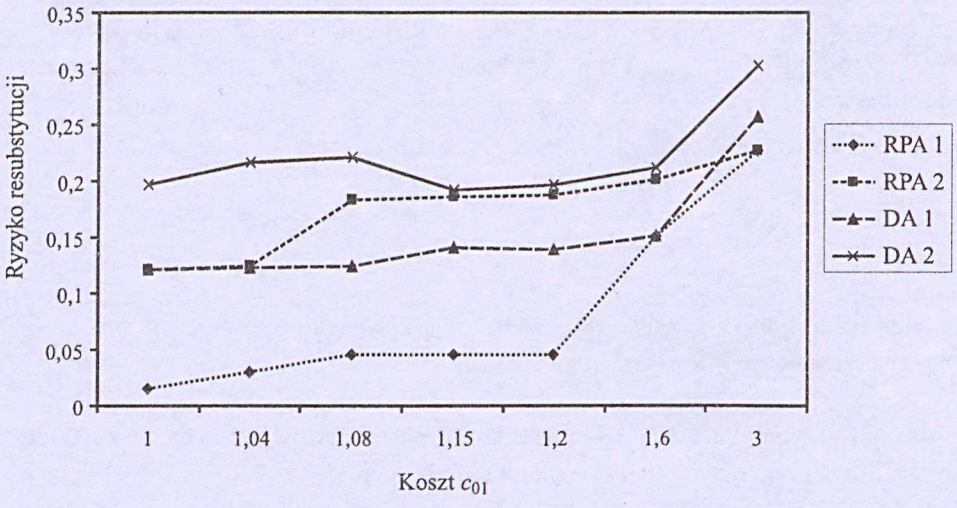
kosztu $c_{01} = 1,2$ jego przewaga nad pozostałymi metodami jest bardzo duża, co obrazuje wykres 1. Ryzyko resubstytucji związane z tym modelem jest najniższe i dopiero przy koszcie 1,6 zrównuje się z modelem DA1, aby dla kosztu 3 znowu osiągnąć najmniejsze ryzyko.

Model RPA2 dla kosztu $c_{01} = 1$ oraz $c_{01} = 1,04$ ma dokładnie takie samo ryzyko jak model DA1. Jeżeli jednak przyjrzymy się błędom, jakie popełniały oba te modele, to widać wyraźnie, że RPA2 popełnia zdecydowanie więcej błędów I typu. Z punktu widzenia naszego badania jest to gorsza sytuacja i dlatego można powiedzieć, że model DA1 jest lepszy. RPA2 klasyfikuje również gorzej od modelu DA1 dla wszystkich kosztów od 1,08 do 1,6. Dopiero przy koszcie $c_{01} = 3$ RPA2 zrównuje się ryzykiem z modelem RPA1 i tym samym jest lepszy od modeli DA1 i DA2.

Model DA1, z wyjątkiem klasyfikacji dla kosztu $c_{01} = 1,6$, klasyfikuje gorzej niż model RPA1. Jednocześnie jest on lepszy niż model RPA2, aż do kosztu $c_{01} = 3$.

Model DA2 klasyfikuje zdecydowanie najgorzej. Dla wszystkich wartości kosztu jego ryzyko jest największe. Różnica w klasyfikacji pomiędzy DA2 i RPA1 jest ogromna. Najmniej model ten odbiega od klasyfikacji modelu RPA2 i to tylko dla c_{01} przyjmującego wartości od 1,08 do 1,6.

Powyższe wnioski potwierdza poniższy wykres. Widać wyraźnie, że dla wszystkich poziomów kosztu błędnej klasyfikacji podmiotów najniżej położona jest linia RPA1 – świadczy to o najniższym ryzyku resubstytucji modelu.



Wykres 1. Zależność ryzyka resubstytucji od zmiany kosztu błędnej klasyfikacji podmiotu

Źródło: Opracowanie własne

Warto zwrócić uwagę, że wraz ze wzrostem kosztu błędnej klasyfikacji jakość klasyfikacji modeli RPA i DA zbliża się do siebie. Wynika to z faktu, że ze względu na rosnącą „karę" za złe zaklasyfikowanie podmiotu „dobrego" jako „złego" modele wolą popełnić błąd w drugą stronę. Dlatego rośnie ogólny koszt błędnej klasyfikacji i modele uzyskują podobne wyniki.

Na koniec postanowiono sprawdzić, ile ryzyka pozwalają nam zaoszczędzić modele w stosunku do „naiwnej" reguły klasyfikacyjnej. Przyjęto następującą „naiwną" regułę:

1) dla kosztu $c_{01} \leq 1{,}15$ wszystkie podmioty zostają zaklasyfikowane jako „złe",

2) dla kosztu $c_{01} > 1{,}15$ wszystkie podmioty zostają zaklasyfikowane jako „dobre".

Tabela 3 prezentuje wskaźniki obliczone, jako stosunek ryzyka resubstytucji danego modelu przy określonym koszcie do poziomu ryzyka dla „naiwnej" reguły klasyfikacyjnej przy tym samym koszcie. Wyraźnie daje się zauważyć, że im wyższy jest koszt, tym mniej zyskujemy ryzyka w stosunku do naiwnej reguły klasyfikacyjnej. Przy koszcie $c_{01} = 1{,}04$ ryzyko resubstytucji RPA1 stanowi zaledwie 3% ryzyka resubstytucji naiwnej reguły klasyfikacyjnej. Oznacza to, że na samym początku zyskiwaliśmy około 97% ryzyka w stosunku do reguły naiwnej, a przy koszcie $c_{01} = 3$ jest to już tylko około 54% zysk.

Obliczony wskaźnik pokazuje, że zysk wynikający z zastosowania CART jest duży. Uzasadnia to potrzebę korzystania z różnych technik, pomimo nakładów poniesionych na oszacowanie modeli.

Tabela 3. Stosunek ryzyka resubstytucji danego modelu do ryzyka „naiwnej" reguły klasyfikacyjnej

c_{01}	Modele			
	RPA1	RPA2	DA1	DA2
1,04	0,0288	0,2327	0,2327	0,3788
1,08	0,0556	0,2296	0,2278	0,4000
1,15	0,0783	0,3200	0,2157	0,3843
1,20	0,0900	0,3760	0,2780	0,3940
1,60	0,3040	0,4040	0,3040	0,4240
3,00	0,4540	0,4540	0,5160	0,6060

Źródło: Opracowanie własne.

Wnioski

W przeprowadzonych badaniach widać przewagę modeli opartych na drzewach decyzyjnych, skonstruowanych za pomocą algorytmu rekurencyjnego podziału, nad liniowymi funkcjami dyskryminacyjnymi (model RPA2 był gorszy od DA1 tylko w niewielkim stopniu). Przyczyn takiej sytuacji można upatrywać przede wszystkim w nieparametryczności techniki RPA. Brak wymagań formalnych sprawia, że praktycznie w każdej sytuacji można dopasować model do zmiennych. W przypadku analizy dyskryminacyjnej, niespełnienie któregokolwiek z założeń może powodować pogorszenie modelu. Jest to tym bardziej istotne, że w praktyce gospodarczej bardzo trudno jest znaleźć zmienne o rozkładzie normalnym, czy równych macierzach wariancji-kowariancji w różnych grupach obiektów.

Kolejną przyczyną przewagi metody RPA jest jej jednowymiarowe podejście na każdym poziomie procedury klasyfikacyjnej. Przy każdym podziale węzła wyszukuje się zmiennej najlepiej dyskryminującej obiekty. Dzięki temu można jedną zmienną wykorzystać więcej niż raz. Taka sytuacja nie jest możliwa w przypadku analizy dyskryminacyjnej. Metoda RPA okazała się lepsza od analizy dyskryminacyjnej. Nie oznacza to, że zawsze będzie ona lepsza od innych technik klasyfikacyjnych. Prostota szacowania i uzyskane wyniki zachęcają jednak do korzystania z tej metody. Ma ona również jeszcze jedną zaletę, której nie posiada liniowa funkcja dyskryminacyjna – możliwość uwzględniania zmiennych jakościowych.

Literatura

[1] Aczel, A. D., *Statystyka w zarządzaniu: pełny wykład*, Wydawnictwa Naukowe PWN, Warszawa 2000

[2] Altman, E. I., *Financial Ratios, Dyscryminant analysis and the Prediction of Corporate Bankruptcy*, „The Journal of Finance" 1968, nr 4, vol. 23

[3] Bao, H. T. *Data Miting and Knowledge Discovery Techniques and Practice*, www.netnam.vn/unescocourse/knowlegde/knowlegd.htm.

[4] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. *Classification and Regression Trees*, CHAPMAN& HALL/CRC, 1984

[5] Elias, A., *Ekonometria przestrzenna*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1991

[6] Fisher, R. A. *The Use of Multiple Measurements In Taxonomic Problems*, „Annals of Eugenics" 1936, nr 7

[7] Frydman, H., Altman, E. I., Kao, D., *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*, „The Journal of Finance" 1985, marzec, vol. 40

[8] Guzik, B., Appenzeller, D., Jurek, W., *Prognozowanie i symulacje: wybrane zagadnienia*, Wyd. AE w Poznaniu, Poznań 2004
[9] Hadasik, D. *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Zeszyty Naukowe – Wyd. AE w Poznaniu, Seria 2 – Prace Habilitacyjne i Doktorskie, Poznań 1998
[10] Laitinen, T., Kankaapaa, M., *Comparative Analysis of Failure Prediction Methods*, The University of Vaasa, Accounting and Finance
[11] Leszczyński, Z., *Analiza dyskryminacyjna w przewidywaniu bankructwa firmy*, „Expert" 1993, nr 3
[12] Moczko, J., *Wybrane metody eksploracji danych i wspomagania procesów decyzyjnych* [w:] *Metody statystyczne i techniki data miting w badaniach naukowych: seminarium Warszawa 15 października 2002*, StatSoft Polska, Kraków 2002
[13] Szapiro, T., *Co decyduje o decyzji?*, Wydawnictwo Naukowe PWN, Warszawa 1993
[14] Zhang, H. P., Singer, B., *Recursive Partitioning in the Health Science*, Springer, New York 1999

Piotr Jardanowski

EVALUATION OF THE FINANCIAL STANDING OF A COMPANY: LINEAR DISCRIMINANT ANALYSIS AND CLASSIFICATION AND REGRESSION TREES

Summary

The purpose of this article is to present two classification procedures: linear discriminant analysis (LDA) and classification and regression trees based on recursive partitioning algorithm (RPA) and to compare them in the context of firm bankruptcy. The results from both procedures study were compared and RPA was found to outperform LDA in original sample.