

Tomasz Strąk

Uniwersytet Szczeciński

WYKORZYSTANIE METODY REKURENCYJNEGO PODZIAŁU DO OCENY ZAGROŻENIA BANKRUCTWEM POLSKICH PRZEDSIĘBIORSTW

1. Wstęp

Ocena zagrożenia bankructwem przedsiębiorstwa jest procesem mającym na celu określenie na podstawie danych historycznych lub planowych prawdopodobieństwa wystąpienia tego zjawiska w przyszłości. Prawdopodobieństwo bankructwa przedsiębiorstwa w czasie t niekoniecznie musi być przy tym oszacowane w sposób ilościowy, może być także wyrażone w sposób opisowy przy użyciu następujących określeń: wysokie, średnie lub niskie. Ocena taka ma zatem charakter *ex ante* i stanowi ważny element monitoringu sytuacji finansowej firmy.

Istota metod oceny zagrożenia bankructwem sprowadza się do porównania – na podstawie zmiennych, które uznano za symptomy niewypłacalności – badanego podmiotu z przedsiębiorstwami wypłacalnymi i tymi, które stały się bankrutami w czasie t . Im firma jest bardziej podobna do firm, które zbankrutowały, tym prawdopodobieństwo bankructwa jest większe. Porównania tego dokonuje się za pomocą modeli, których parametry szacuje się na podstawie badań empirycznych *ex post* dużej grupy przedsiębiorstw obejmujących firmy wypłacalne i niewypłacalne. Metody te na ogół nie wykorzystują teorii, co skłoniło niektórych krytyków do określenia ich mianem „brutalnego empiryzmu” [Ball, Foster 1982, s. 161-234].

Modele oceny zagrożenia bankructwem przedsiębiorstwa konstruuje się z wykorzystaniem metod zarówno parametrycznych (szczególnie: analizy dyskryminacyjnej, analizy logitowej i probitowej) jak i eksploracyjnych, a więc mających charakter nieparametryczny. W technikach eksploracyjnych – obecnie powszechnie wykorzystywanych w Data Mining [Sung, Chang, Lee 1999, s. 63-85] – nie zakłada się znajomości rozkładów cech ani postaci analizowanego związku między nimi. Dobór zmiennych następuje automatycznie na podstawie przyjętych wcześniej kryteriów [Gatnar 2001, s. 7-8]. Można tu wymienić w szczególności takie techniki, jak: anali-

za skupień, metody programowania matematycznego, sieci neuronowe oraz metodę rekurencyjnego podziału. Metodę rekurencyjnego podziału do oceny zagrożenia bankructwem wykorzystali m.in.: E.I. Altman, H.N. Frydman i D.L. Kao [Altman, Frydman, Kao 1985, s. 269-291] oraz T.E. McKee [McKee 1995, s. 26-36].

Celem badań zaprezentowanych w artykule była analiza możliwości wykorzystania metody rekurencyjnego podziału do oceny zagrożenia bankructwem polskich przedsiębiorstw. Przedmiotowe badania przeprowadzono z wykorzystaniem sprawozdań finansowych 2092 polskich przedsiębiorstw.

2. Metoda rekurencyjnego podziału

Metoda rekurencyjnego podziału¹ polega na stopniowym podziale wielowymiarowej przestrzeni cech, w której znajduje się zbiór obserwacji, na rozłączne segmenty (podzbiory), aż do uzyskania ich homogeniczności ze względu na wyróżnioną cechę (y) [Gatnar 2001, s. 8].

Metoda ta pozwala przezwyciężyć ograniczenia klasycznych metod statystycznych i zbudować model bez znajomości postaci rozkładów cech oraz związków między nimi. Do realizacji tego celu konieczne jest jednak posiadanie stosunkowo dużej próby. Drugą cechą tej metody jest jej adaptacyjny charakter. Metody adaptacyjne polegają na doborze zmiennych do modelu przez optymalizację pewnego kryterium, podczas gdy metody klasyczne wykorzystują z góry zadany zbiór predyktorów. Trzecią cechą omawianej metody jest to, iż zamiast budować jeden model globalny, tworzy się K modeli lokalnych $g_k(x, \beta)$ zmiennej y , będących elementem składowym modelu:

$$y = \alpha_0 + \sum \alpha_k g_k(x, \beta_k)$$

gdzie: $g_k(x, \beta)$ – funkcja składowa o wewnętrznych parametrach β [Gatnar 2001, s. 14-17].

Jeżeli przedmiotem analizy jest zbiór obserwacji S , którego każdy element jest charakteryzowany przez wektor zmiennych niezależnych (predyktorów) $x = [x_1, x_2, \dots, x_M]$, to wielowymiarowe obserwacje można zapisać w postaci macierzy:

$$\begin{bmatrix} x_{11} & \dots & x_{1M} & y_1 \\ x_{21} & \dots & x_{2M} & y_2 \\ \dots & \dots & \dots & \dots \\ x_{N1} & \dots & x_{NM} & y_N \end{bmatrix}$$

gdzie: y – zmienna zależna [Gatnar 2001, s. 13].

¹ Pierwszym całościowym opracowaniem dotyczącym metody rekurencyjnego podziału była praca [Breiman i in. 1984].

W rezultacie rekurencyjnego podziału zbiór S zostanie rozbity na K rozłącznych podzbiorów $S_1, S_2, \dots, S_K$, z których każdy liczy $N(k)$ obiektów ($k = 1, 2, \dots, K$).

Spełniony jest przy tym warunek: $\sum_{k=1}^K N(k) = N$. Podział ten dokonywany jest tak długo,

aż osiągnie się maksymalną jednorodność podzbiorów ze względu na wartość zmiennej zależnej y , która w wypadku oceny zagrożenia bankructwem jest cechą nominalną, przyjmującą wartości: 0 lub 1. Jego kryterium jest wartość wybranych zmiennych niezależnych x_n lub ich kombinacja liniowa. Ten hierarchiczny proces można przedstawić graficznie w postaci drzewa klasyfikacyjnego. Jeżeli z każdego jego wierzchołka wychodzą dwie krawędzie (podział na dwie grupy), to drzewo takie nosi nazwę drzewa binarnego [Gatnar 2001, s. 19-27].

W procesie budowy drzew klasyfikacyjnych można wyróżnić cztery podstawowe etapy: 1) określenie kryteriów oceny jakości modelu, 2) wybór metody podziałów, 3) wyznaczenie końca podziałów i 4) wybór drzewa właściwej wielkości [Moduł drzewa...].

Najważniejszym kryterium oceny jakości modelu drzewa klasyfikacyjnego jest precyzja określenia wartości zmiennej y dla obiektów ze zbioru rozpoznawanego. Ocena precyzji może być dokonana m.in. przy użyciu kosztów błędnej klasyfikacji. Koszty te reprezentują „karę” za przydzielenie pewnemu obiektowi wartości y_j zmiennej zależnej y , podczas gdy faktycznie wartość ta wynosi y_j . Minimalizacja kosztów może być zatem podstawowym celem analizy drzew klasyfikacyjnych [Gatnar 2001, s. 101-102].

Drugi etap w analizie drzewa klasyfikacyjnego polega na wybraniu podziałów na podstawie zmiennych objaśniających, które są wykorzystane do przewidywania przynależności przypadków lub obiektów do klas wyznaczonych przez zmienną zależną. Wyróżniamy trzy podstawowe metody wyboru podziału. Pierwsza z tych metod to dyskryminacyjny podział jednowymiarowy dla nominalnych lub porządkowych zmiennych predykcyjnych. Drugą metodą wyboru podziału jest dyskryminacyjny podział z wykorzystaniem kombinacji liniowych dla zmiennych objaśniających. Trzecią metodą wyboru podziału jest metoda CART, wyczerpującego poszukiwania podziałów jednowymiarowych dla nominalnych lub porządkowych zmiennych predykcyjnych [Moduł drzewa...].

Trzeci etap analizy drzewa klasyfikacyjnego polega na rozstrzygnięciu, kiedy należy zakończyć dzielenie. Sposoby zatrzymania podziału w różnych algorytmach klasyfikacyjnych są różne. W pakiecie Statistica są dostępne dwie opcje sterowania decyzją zakończenia procesu podziału: minimalna liczebność podklasy, reguła błędu standardowego i frakcja obiektów [Moduł drzewa...].

Ostatni etap algorytmu rekurencyjnego podziału to określenie właściwej wielkości drzewa. Istnieją dwie podstawowe strategie wyboru drzewa właściwej wielkości spośród wszystkich możliwych drzew. Pierwsza strategia polega na budowaniu

drzewa do określonej, właściwej wielkości, gdzie właściwa wielkość jest wyznaczona przez użytkownika na podstawie wiedzy z poprzednich badań, informacji diagnostycznych uzyskanych w poprzednich analizach lub choćby intuicji. Druga strategia polega na wykorzystaniu do wyboru drzewa właściwej wielkości automatycznego wyboru drzewa na podstawie techniki, którą opracowali Breiman i in. (w 1984 r.), nazywanej przycinaniem na podstawie minimalizacji kosztów i złożoności drzewa w sprawdzanie krzyżowym [Breiman i in. 1984].

W literaturze przedmiotu przyjmuje się dodatkowo, iż model powinien być tak zbudowany, aby kryterium podziału było zgodne z teorią i logiką, np. im mniejsza wartość wskaźnika bieżącej płynności, tym większe ryzyko bankructwa, a nie na odwrót [Breiman i in. 1984, s. 60].

3. Charakterystyka próby badawczej

Próbkę badawczą, stanowiącą podstawę budowy modeli oceny zagrożenia bankructwem tworzyło 1147 podmiotów, z tego 544 zbankrutowało w latach 1996-2002, a pozostałe były przedsiębiorstwami wypłacalnymi². Dane o sytuacji finansowej analizowanych przedsiębiorstw pochodziły z: Krajowego Rejestru Sądowego, sądów upadłościowych, Monitora Polskiego B oraz z kwartalnika „Notoria Serwis”. W grupie badawczej 252 przedsiębiorstwa były spółkami publicznymi.

Struktura branżowa analizowanej grup przedsiębiorstw zaprezentowana jest w tab. 1, a w tab. 2 statystyki dotyczące ich aktywów, zobowiązań i przychodów ze sprzedaży.

Tabela 1. Struktura branżowa analizowanej grup przedsiębiorstw (w %)

Rodzaj działalności	Firmy niewypłacalne	Firmy wypłacalne
Przemysł	35	38
Handel	20	12
Produkcja i przetwórstwo rolne	8	7
Transport	2	3
Informatyka i telekomunikacja	3	7
Usługi budowlano-montażowe	17	17
Pozostałe usługi	15	16

Źródło: opracowanie własne.

² Podmioty te prowadziły swoją działalność w roku 2003 i spełniały przedstawione wyżej kryterium.

Tabela 2. Aktywa, zobowiązania i przychody ze sprzedaży analizowanej grup firm (w zł)

Zmienna	Przedsiębiorstwa niewypłacalne		
	aktywa	zobowiązania ogółem	przychody ze sprzedaży
Średnia	43 800 182,49	35 592 620,00	19 914 733,19
Mediana	1 198 979,92	1 558 316,99	2 270 000,00
Dolny Kwartyl	264 300,00	387 800,00	534 927,11
Górny Kwartyl	6 973 267,23	7 818 905,05	10 424 177,43
Zmienna	Przedsiębiorstwa wypłacalne		
	aktywa	zobowiązania ogółem	przychody ze sprzedaży
Średnia	128 608 940,84	66 080 055,93	151 665 199,27
Mediana	14 445 100,00	5 030 615,72	23 194 079,87
Dolny Kwartyl	5 684 228,16	1 428 267,57	8 523 991,41
Górny Kwartyl	58 732 334,82	26 758 000,00	75 672 000,00

Źródło: opracowanie własne.

Posiadany materiał empiryczny pozwolił na zbudowanie modeli oceny zagrożenia bankructwem w okresie do 36 miesięcy. Próba analityczna dla poszczególnych modeli obejmowała:

1) dla modeli oceny zagrożenia bankructwem w okresie do 12 miesięcy: zbiór uczący – 300 niewypłacalnych firm i tyle samo wypłacalnych; zbiór testowy – 214 niewypłacalnych przedsiębiorstw i 214 wypłacalnych;

2) dla modeli oceny zagrożenia bankructwem w okresie od 12 do 24 miesięcy: zbiór uczący – 300 niewypłacalnych firm oraz 300 wypłacalnych; zbiór testowy – 189 niewypłacalnych firm i 213 wypłacalnych;

3) dla modeli oceny zagrożenia bankructwem w okresie od 24 do 36 miesięcy: zbiór uczący – 200 niewypłacalnych firm i 200 wypłacalnych; zbiór testowy – 142 niewypłacalnych firm i 299 podmiotów wypłacalnych.

4. Modele oceny zagrożenia bankructwem polskich przedsiębiorstw

Modele oceny zagrożenia bankructwem zostały zbudowane przy pomocy modułu „drzewa klasyfikacyjne” pakietu Statistica 6.0. Użyto metod dyskryminacyjnych podziału na podstawie kombinacji linowej dla predyktorów porządkowych. Przyjęto, iż relacja kosztu błędnego zakwalifikowania firmy zagrożonej bankructwem do grupy podmiotów niezagrażonych bankructwem w stosunku do kosztu błędnego zakwalifikowania firmy niezagrażonej bankructwem do grupy podmiotów zagrożo-

nych bankructwem jest równa 1,5³. Jako opcji zatrzymania użyto reguły „przytnij przy błędzie złej klasyfikacji”, przy parametrach zatrzymania: minimalna liczność 5 i regule błędu standardowego. W celu sprawdzenia krzyżowego przyjęto $V = 10$.

Zmienne objaśniające do modeli oceny zagrożenia bankructwem dobierano spośród wyjściowego zbioru zmiennych (obejmującego 123 zmienne), który wyznaczono na podstawie analizy: wyników dotychczasowych badań, przyczyn trudności finansowych i ich symptomów, czynników wpływających na podatność przedsiębiorstwa na kryzys, literatury dotyczącej analizy finansowej i oceny ryzyka kredytowego oraz dostępnego materiału statystycznego. Zbiór ten następnie poddano selekcji. W wyniku zastosowania procedury [Stępień, Strak 2004, s. 448-449] wybrano trzy zmienne: X_1 – zysk netto/aktywa ogółem, X_{34} – kapitał pracujący/aktywa ogółem i X_{72} – zobowiązania ogółem/aktywa ogółem.

Zbudowano przy tym trzy modele. Model M1 pozwala na ocenę zagrożenia bankructwem w okresie do 12 miesięcy, model M2 – w okresie od 12 do 24 miesięcy, a model M3 w okresie od 24 do 36 miesięcy. Dla modelu M1 ogólna trafność klasyfikacji wynosi dla próby uczącej 92% (dla firm niewypłacalnych i wypłacalnych – 92%), dla próby testowej – 92% (firmy niewypłacalne – 89% i firmy wypłacalne – 94%). Dla modelu M2 ogólna trafność klasyfikacji wynosi dla próby uczącej 86% (dla firm niewypłacalnych – 86% i firm wypłacalnych – 85%), dla próby testowej – 82% (firmy niewypłacalne – 83% i firmy wypłacalne – 81%). Dla modelu M3 ogólna trafność klasyfikacji wynosi dla próby uczącej 78% (dla firm niewypłacalnych – 76% i firm wypłacalnych – 80%), dla próby testowej – 76% (firmy niewypłacalne – 75% i firmy wypłacalne – 77%).

Współczynniki dla podziałów kombinacji liniowych poszczególnych modeli zaprezentowano w tab. 3.

Tabela 3. Współczynniki dla podziałów kombinacji liniowych

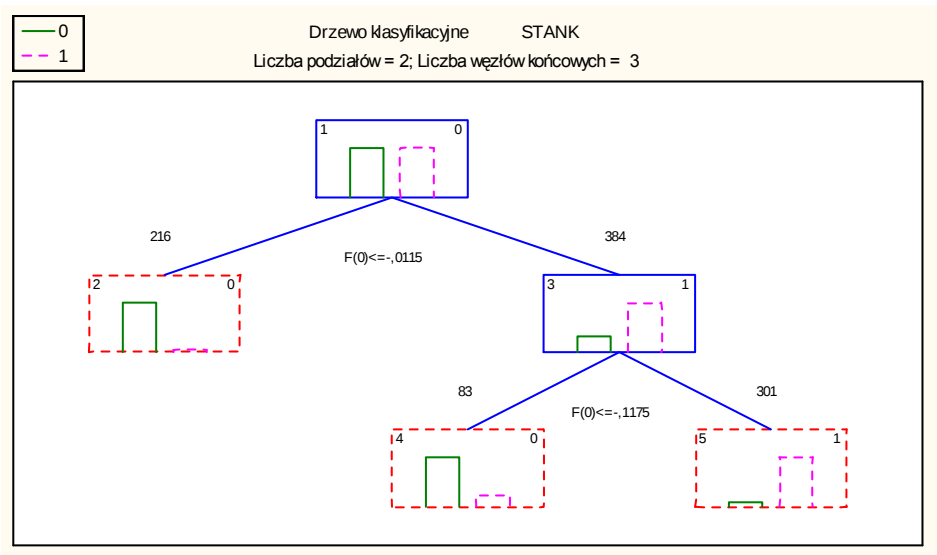
M1	X_1	X_{34}	X_{72}	M3	X_1	X_{72}
F(1)	0,013	0,025	-0,007	F(1)	0,043	-0,024
F(2)	0,134	0,000	-0,154	F(2)	0,204	-0,210
M2	X_1	X_{34}	X_{72}			
F(1)	0,007	0,002	-0,004			
F(2)	0,109	0,047	-0,139			

Źródło: opracowanie własne.

Model M1 zaprezentowano na rys. 1, model M2 na rys. 2, a model M3 na rys. 3. Modele te pozwalają na zakwalifikowanie ocenianego przedsiębiorstwa do jednej z

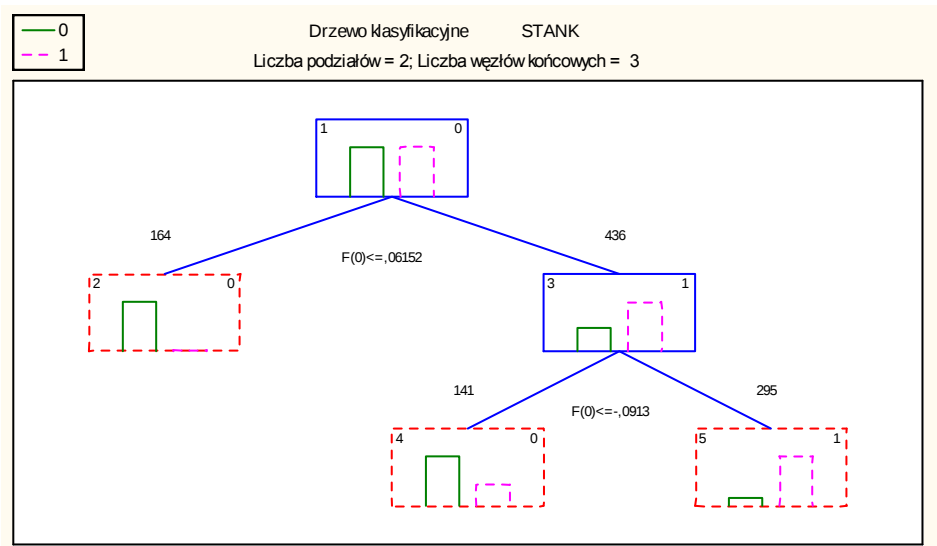
³ Dla takiego współczynnika otrzymano zadowalającą jakość klasyfikacji firm niewypłacalnych.

trzech grup: wysokie ryzyko bankructwa, średnie ryzyko bankructwa i niskie ryzyko bankructwa.



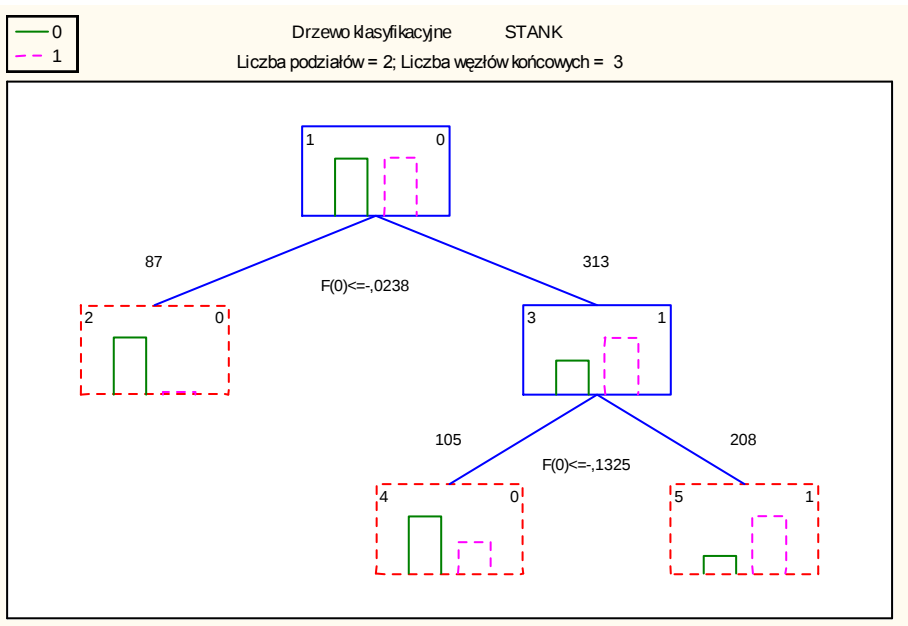
Rys. 1. Model M1

Źródło: opracowanie własne.



Rys. 2. Model M2

Źródło: opracowanie własne.



Rys. 3. Model M3

Źródło: opracowanie własne.

Tabela 4. Wyniki grupowania dla próby uczącej

Ryzyko bankructwa	Rok 1, próba ucząca = 600			Rok 2, próba ucząca = 600			Rok 3, próba ucząca = 400		
	klasa 0	klasa 1	trafność klasyfikacji	klasa 0	klasa 1	trafność klasyfikacji	klasa 0	klasa 1	trafność klasyfikacji
Wysokie	207	9	96%	173	5	97%	83	4	95%
Średnie ⁴	68	15	82%	83	41	67%	68	37	65%
Niskie	25	276	92%	44	254	85%	49	159	76%

Źródło: opracowanie własne.

Jeżeli wartość $F(1) \leq$ wartość graniczna (1)⁵, to oceniany podmiot zaliczany jest do grupy o wysokim ryzyku bankructwa. Jeżeli $F(1) >$ wartość graniczna (1) i $F(2) \leq$ wartość graniczna (2), to oceniana firma zaliczana jest do grupy o średnim ryzyku bankructwa. Jeżeli $F(1) >$ wartość graniczna (1) i $F(2) >$ wartość graniczna (2), to

⁴ Na potrzeby oceny jakości klasyfikacji za poprawnie rozpoznane uznano te firmy, które w analizowanym okresie rzeczywiście zbankrutowały.

⁵ Wartości graniczne umieszczone są poszczególnych schematach. Na przykład dla modelu M1 wartość graniczna funkcji $F(1)$ wynosi $-0,0115$, a funkcji $F(2)$ jest równa $-0,1175$.

oceniane przedsiębiorstwo zaliczane jest do grupy o niskim ryzyku bankructwa. W tab. 4 i 5 zaprezentowano wyniki grupowania dla próby uczącej i testowej.

Tabela 5. Wyniki grupowania dla próby testowej

Ryzyko bankructwa	Rok 1, próba ucząca = 428			Rok 2, próba ucząca = 402			Rok 3, próba ucząca = 441		
	klasa 0	klasa 1	trafność klasyfikacji	klasa 0	klasa 1	trafność klasyfikacji	klasa 0	klasa 1	trafność klasyfikacji
Wysokie	123	4	97%	102	7	94%	46	8	85%
Średnie	70	16	81%	54	38	59%	61	61	50%
Niskie	21	194	90%	33	168	84%	35	230	87%

Źródło: opracowanie własne.

Z analizy zaprezentowanych danych wynika, iż w przypadku zaliczenia przedsiębiorstwa do grup wysokiego ryzyka bankructwa istnieje bardzo duże prawdopodobieństwo, że analizowany podmiot zbankrutuje we wskazanym okresie. Dokładność prognozy w grupach o niskim ryzyku bankructwa jest jednak słaba i waha się od 76 do 92%.

Dodatkowo zdolność prognostyczną modelu oceniono dla próby 945 przedsiębiorstw o przeciętnej wartości (mierzonej medianą) aktywów równej 33 695 000 zł, zobowiązań ogółem – 15 688 000 zł i przychodach ze sprzedaży – 54 183 000 zł⁶. Analizowaną grupę stanowiły jedynie podmioty niepubliczne, zarówno o dobrej, jak i o słabej kondycji finansowej, o których wiadomo, iż kontynuowały działalność przez cztery lata po roku (1998), z którego pochodziły dane do wyznaczenia zmiennych modelu. Podział na podgrupy 945 firm zaprezentowano w tab. 6.

Tabela 6. Podział na podgrupy 945 firm

Węzły	Rok 1		Rok 2		Rok 3	
	Liczba	%	Liczba	%	Liczba	%
Węzeł 2	67	7	77	8	69	7
Węzeł 4	130	14	208	22	242	26
Węzeł 5	748	79	660	70	634	67
Suma	945	100	945	100	945	100

Źródło: opracowanie własne.

Przedstawione dane potwierdzają wysoką zdolność prognostyczną analizowanych modeli.

⁶ Źródłem informacji o analizowanej grupie przedsiębiorstw był Krajowy Rejestr Sądowy i Monitor Polski B.

5. Zakończenie

Wykorzystanie drzew klasyfikacyjnych (metody rekurencyjnego podziału) pozwala na identyfikację ryzyka bankructwa polskich przedsiębiorstw. Oceniane przedsiębiorstwo, w zależności od aktualnych wyników finansowych, może być zakwalifikowane do grupy o wysokim, średnim lub niskim ryzyku bankructwa. Z przeprowadzonych badań wynika, iż jedynie ok. 7% firm zakwalifikowanych do grupy o wysokim ryzyku bankructwa w przyjętym okresie prognozy w rzeczywistości nie zbankrutuje. Dla grupy podmiotów o wysokim ryzyku bankructwa prezentowane modele wykazują wysoką trafność klasyfikacji w okresie do 3 lat.

Konsekwencją wyodrębnienia obszaru o średnim ryzyku bankructwa jest pogorszenie jakości klasyfikacji przedsiębiorstw nie zagrożonych bankructwem. Spadek poprawności klasyfikacji dla tej grupy może jednak wynikać także z trudność jednoznacznego zidentyfikowania przedsiębiorstw o niskim ryzyku bankructwa.

Literatura

- Altman E.I., Frydman H.N., Kao D.L., *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*, „Journal of Finance” 1985, vol. XL, nr 1.
- Ball R., Foster G., *Corporate Financial Reporting: A Methodological Review of Empirical Research*, „Journal of Accounting Research”, Supplement 1982.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., *Classification and Regression Trees*, Wadsworth Interbatiobal Grup, Belmont, 1984.
- Gatnar E., *Nieparametryczna metoda dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- McKee T.E., *Predicting Bankruptcy via Induction*, „Journal of Information Technology” 1995, nr 10.
- Moduł drzewa klasyfikacyjnego*, dokumentacja programu Statistica 6.0.
- Stepień P., Strąk T., *Wielowymiarowe modele logitowe oceny zagrożenia bankructwem polskich przedsiębiorstw*, [w:] *Zarządzanie finansami: finansowanie przedsiębiorstw w UE*, red. D. Zarzecki, t. 1, Szczecin 2004.
- Sung T.K., Chang N., Lee G., *Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction*, „Journal of Management Information Systems” 1999, vol. 16, nr 1.

RECURSIVE PARTITIONING IN PREDICTING CORPORATE BANKRUPTCY IN POLAND

Summary

The paper presents the results of the research which explored the applicability of recursive partitioning to predicting corporate bankruptcy in Poland. Three models were thus constructed. M1 model allows predicting bankruptcy within up to 12 months, M2 model – 12-24 months in advance,

and M3 model – 24-36 months in advance. The analyzed company, depending on its present financial performance, can be classified as a high-, medium- or low bankruptcy risk entity. The research shows that only ca. 7% of companies classified as high bankruptcy risk entities will not actually go bankrupt within the period of prediction.

The models presented in the paper were constructed on the basis of financial reports of 1147 companies, and then verified on the sample of 945 solvent companies.