

Wielokryterialna ocena modeli prognozowania stanu ekonomiczno-finansowego małych i średnich przedsiębiorstw

Andrzej Burda

Wyższa Szkoła Zarządzania i Administracji w Zamościu

Streszczenie: *W artykule zaproponowano sposób oceny modeli prognozowania stanu ekonomiczno-finansowego małych i średnich przedsiębiorstw. Na bazie przyjętych kryteriów przeanalizowano mocne i słabe strony modeli budowanych z wykorzystaniem metod statystycznych oraz uczenia maszynowego. Zaproponowano koncepcję optymalizacji heterogenicznego modelu prognostycznego.*

Cel i zakres oceny

Małe i średnie przedsiębiorstwa (MŚP) stanowią ważny element gospodarki wszystkich krajów Unii Europejskiej, zaś szczególne znaczenie mają w Polsce, w obszarach słabo zurbanizowanych, m.in. na terenie woj. lubelskiego oraz podkarpackiego. Upadłość MŚP w każdym przypadku stanowi bardzo poważne zagrożenie bezrobociem, dlatego opracowanie wiarygodnej metodologii oceny ich stanu (np. **stan dobry** lub **stan zagrożenia bankructwem**) ma duże znaczenie dla nich samych, a także spełnia ważną funkcję społeczną.

Prezentowane opracowanie dotyczy oceny modeli prognozowania stanu ekonomiczno-finansowego małych i średnich przedsiębiorstw w woj. lubelskim i podkarpackim. Modele utworzono w oparciu o różne metody klasyfikacji, takie jak analiza dyskryminacyjna, regresja logitowa (metody statystyczne), sieci przekonań Bayesa oraz sztuczne sieci neuronowe (metody uczenia maszynowego).

W odniesieniu do dwustanowych modeli klasyfikujących przedsiębiorstwa na grupy firm o **dobrym** i **złym stanie** (zagrożonych upadłością) ekonomiczno-finansowym [Nowak 2006] bardzo ważnym elementem ich oceny jest prawdopodobieństwo poprawnej predykcji. Nie należy jednak bagatelizować faktu, że koszt popełnienia błędu I rodzaju (fałszywie pozytywna predykcja) jest znacznie większy od błędu II rodzaju (fałszywie negatywna). W przypadku budowania modeli w oparciu o zbiory niezbilansowane (duża różnica w liczbie obserwacji różnych stanów), a z taką sytuacją mamy do czynienia w analizowanym przypadku¹, małą wartość ogólnego błędu predykcji uzyskują również takie modele, które doskonale dopasują się do klasy większościowej (tu: firm o dobrej kondycji). Takie „przeskalowanie” modelu skutkuje bardzo dużym błędem I rodzaju. Dobierając kryteria oceny modeli, należy mieć to na uwadze.

¹ Średni udział procentowy firm o złej kondycji ekonomiczno-finansowej w zbiorze wszystkich dostępnych przypadków w woj. lubelskim wynosił 7,8%, a woj. podkarpackim – 9%.

Opis metodologii oceny

Modele użyte do porównań, osobne dla różnych działów gospodarki [GUS 2007], budowano na zbiorach danych pochodzących z lat 1999-2005 w woj. lubelskim i podkarpackim. Obejmowały one kolejno sekcje:

- (D) przetwórstwo przemysłowe,
- (F) budownictwo,
- (G) handel hurtowy i detaliczny; naprawa pojazdów samochodowych, motocykli oraz artykułów użytku osobistego i domowego,
- (J) transport, gospodarka magazynowa i łączność,
- (K) obsługa nieruchomości, wynajem i usługi związane z prowadzeniem działalności gospodarczej.

Tab. 1. Liczebność zbiorów testujących.

Sekcja	Liczba przypadków w zbiorze testującym (razem/firmy zagrożone)	
	lubelskie	podkarpackie
D	167/20	197/35
F	62/5	68/4
G	440/38	423/33
I	42/4	29/1
K	83/0	64/4

Źródło: Opracowanie własne.

Ocena modeli dokonana została na niezależnych zbiorach testujących, złożonych z obserwacji MŚP pochodzących z lat 2005-2006 (tab. 1). W sumie ocenie poddano po 8 modeli dla każdej z metod. Modele budowanych dla sekcji **K** z woj. lubelskiego oraz sekcji **I** dla podkarpackiego nie uwzględniono w wynikach ze względu na liczbę przypadków firm zagrożonych bankructwem w zbiorze testującym.

Do oceny przyjęto 4 kryteria:

- sprawność ogólną,
- sprawność I rodzaju,
- iloraz szans,
- wskaźnik optymalnej granicy decyzji.

Jeśli przyjmujemy następujące oznaczenia parametrów:

P_1 – liczba przedsiębiorstw zagrożonych upadłością poprawnie zaklasyfikowanych,

NP_1 – liczba przedsiębiorstw zagrożonych upadłością błędnie zaklasyfikowanych,

P_2 – liczba przedsiębiorstw w dobrym stanie poprawnie zaklasyfikowanych,

NP_2 – liczba przedsiębiorstw w dobrym stanie błędnie zaklasyfikowanych,

to sprawność ogólną można wyrazić wzorem [1]:

$$[1] \quad SP = \frac{P_1 + P_2}{P_1 + NP_1 + P_2 + NP_2} * 100\%.$$

Analogicznie, sprawność I rodzaju [2]:

$$[2] \quad SP_1 = \frac{P_1}{P_1 + NP_1} * 100\%.$$

Iloraz szans obliczany jest zgodnie ze wzorem [3]. Im wyższa jest jego wartość, tym model jest dokładniejszy. Iloraz szans przypadkowej klasyfikacji wynosi 1 [Gruszczyński 2002]:

$$[3] \quad IS = \frac{P_1 * P_2}{NP_1 * NP_2} .$$

Miarę optymalnej granicy decyzji zgodnie ze wzorem [4] stanowi moduł różnicy sprawności I i II rodzaju. Poprawna dyskryminacja obu klas powinna zapewniać zbliżone prawdopodobieństwo ich poprawnej predykcji. Model uznaje się za optymalny, gdy wartość $GD = 0$:

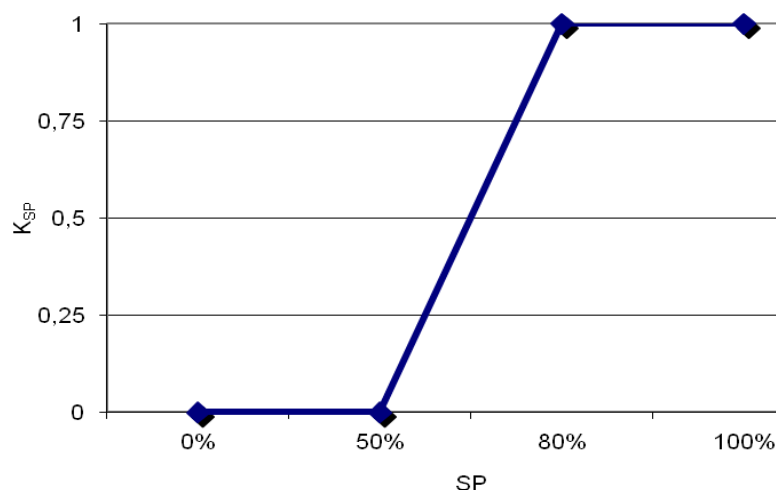
$$[4] \quad GD = |SP_1 - SP_2| .$$

Dobór funkcji użyteczności

W celu doprowadzenia do uniwersalizacji oceny dokonywanej wg różnych kryteriów wykorzystano aparat matematyczny zapożyczony z teorii zbiorów rozmytych, tzw. funkcję użyteczności [Dymowa, Sewastianow i Piłat 2004]. Jej wartość określono w przedziale $[0,1]$, a przebieg uzależniono od rodzaju kryterium.

W przypadku sprawności ogólnej uznano, że jej wartość poniżej 50% świadczy o tym, że model jest gorszy od prognozy losowej, i przyjęto, że wartość funkcji użyteczności dla tego kryterium wynosi $K_{SP} = 0$. Jednocześnie przyjęto, że ze względu na charakter danych, zarówno tych, które wykorzystywano do budowy modelu, jak i testujących, model o sprawności powyżej 80% jest dobry, a występujące różnice powyżej tej wartości mogą być nieistotne statystycznie. W związku z powyższym funkcja użyteczności dla kryterium pierwszego wyrażona jest wzorem [5]. Na rys. 1 przedstawiono wykres tej funkcji.

$$[5] \quad K_{SP} = \begin{cases} 0 & \text{dla } SP \leq 50\% \\ (SP - 50\%) / 30\% & \text{dla } 50\% < SP < 80\% . \\ 1 & \text{dla } SP \geq 80\% \end{cases}$$



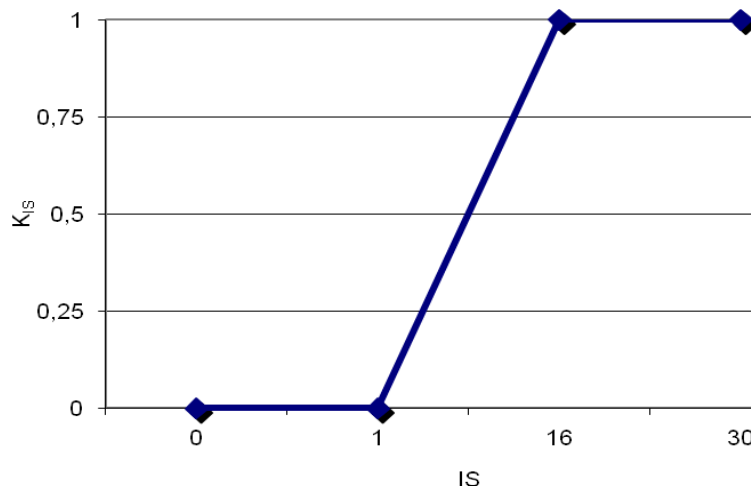
Rys. 1. Funkcja użyteczności $K_{SP} = f(SP)$ dla kryterium sprawności ogólnej.

Źródło: Opracowanie własne.

Analogicznie definiowana jest funkcja użyteczności $K_{SP1} = f(SP_1)$ dla sprawności I rodzaju.

Przy doborze przebiegu funkcji użyteczności dla ilorazu szans [6] kierowano się zasadą, że modele, dla których sprawność I rodzaju i sprawność II rodzaju są mniejsze od 50% ($IS \leq 1$), nie przedstawiają żadnej wartości predykcyjnej. Z kolei te, dla których sprawności wynoszą powyżej 80% ($IS \geq 16$), należy uznać za dobre i wartość funkcji użyteczności powinna wynosić $K_{IS} = 1$ (rys. 2).

$$[6] \quad K_{IS} = \begin{cases} 0 & \text{dla } IS \leq 1 \\ (IS - 1)/15 & \text{dla } 1 < IS < 16 \\ 1 & \text{dla } IS \geq 16 \end{cases}$$

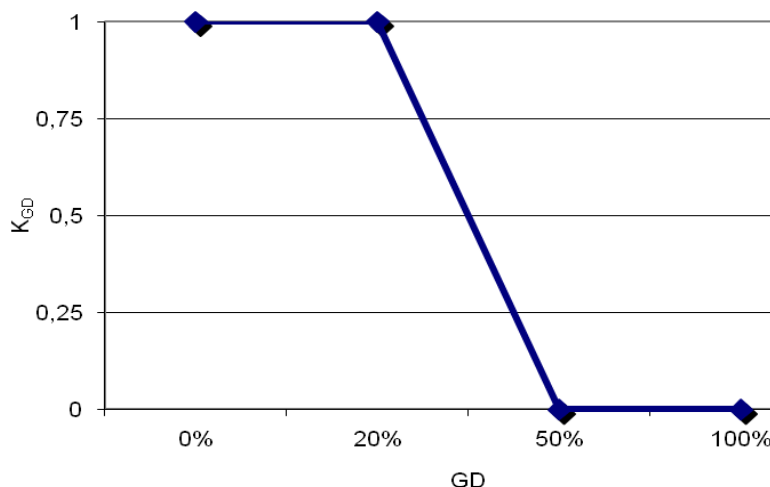


Rys. 2. Funkcja użyteczności $K_{IS} = f(IS)$ dla kryterium ilorazu szans.

Źródło: Opracowanie własne.

Funkcja użyteczności dla optymalnej granicy decyzji [7] premiuje te modele, które z równym prawdopodobieństwem rozpoznają obie klasy. Przy wartości $GD \leq 20\%$ wartość $K_{GD} = 1$. Jeśli jest ona większa od 50%, to $K_{GD} = 0$ (rys. 3).

$$[7] \quad K_{GD} = \begin{cases} 1 & \text{dla } GD \leq 20\% \\ (50\% - GD)/30\% & \text{dla } 20\% < GD < 50\% \\ 0 & \text{dla } GD \geq 50\% \end{cases}$$



Rys. 3. Funkcja użyteczności $K_{GD} = f(GD)$ dla kryterium optymalnej granicy decyzji.

Źródło: Opracowanie własne.

Sformułowanie kryterium globalnego

Dysponując szczegółowymi kryteriami oceny, sformułowano kryterium globalne K_G w celu zregrowania wszystkich uwzględnianych w ocenie informacji do jednej ilościowej oceny modeli. Jego wartość obliczono zgodnie ze wzorem [8]:

$$[8] \quad K_G = \sum_{i=1}^N r_i * K_i .$$

gdzie:

r_i – ranga i -tego kryterium,

K_i – wartość i -tego kryterium,

N – liczba przyjętych kryteriów.

Zgodnie z przyjętą zasadą, że suma rang równa jest 1, wartość kryterium globalnego K_G dla każdego modelu przyjmuje wartości w przedziale od 0 (model najgorszy z możliwych) do 1 (model idealny).

Rangowanie kryteriów

Trzy układy rang kryteriów pokazano w tab. 2. W pierwszym układzie nie wyróżniono żadnego z kryteriów. W układzie drugim uznano, że sprawność I rodzaju powinna być wiodącym kryterium oceny. W układzie trzecim najwyższą rangę nadano kryterium optymalnej granicy decyzji.

Tab. 2. Przyjęte układy rang kryteriów.

Kryterium	Układ 1	Układ 2	Układ 3
K_{IS}	0,25	0,1	0,1
K_{SP}	0,25	0,1	0,1
K_{SP1}	0,25	0,7	0,1
K_{GD}	0,25	0,1	0,7

Źródło: Opracowanie własne.

Wyniki oceny

Ocenie poddano modele: sztucznych sieci neuronowych SSN [Burda 2006], ENN_M i ENN² [Burda, Kuczmowska i Hippe 2007], analizy dyskryminacyjnej AD [Kasjaniuk 2006], regresji logitowej ML [Kowewski, Bielak i Długosz 2007] oraz sieci przekonań Bayesa SPB [Kuczmowska, Burda i Hippe 2007].

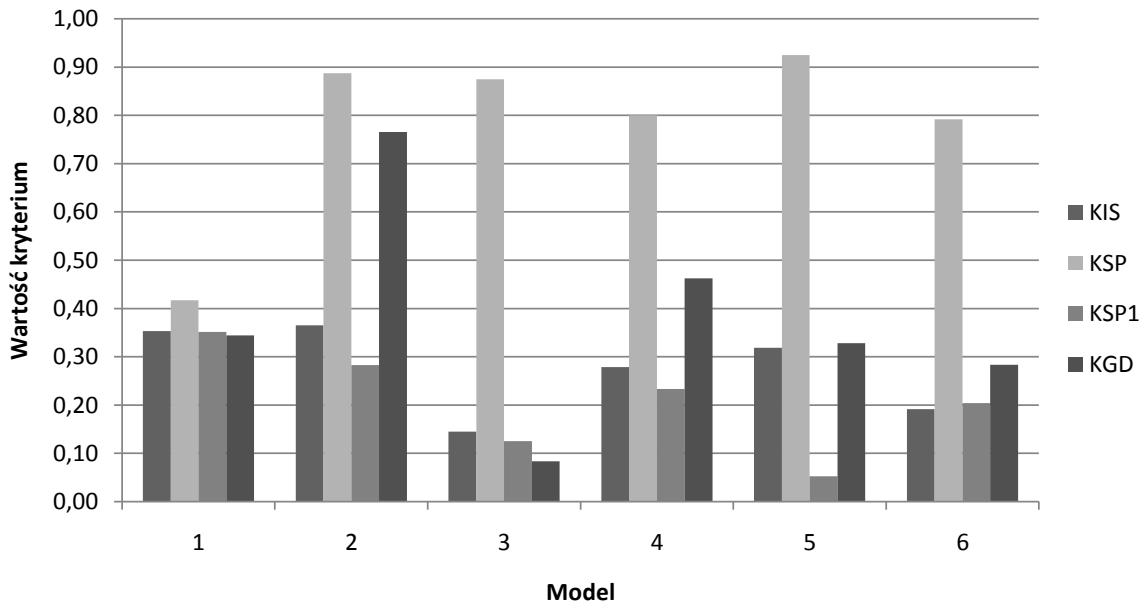
Tab. 3 i rys. 4 przedstawiają zbiór wartości wszystkich kryteriów lokalnych modeli poddanych ocenie.

Tab. 3. Wartości kryteriów lokalnych modeli poddanych ocenie.

Kryterium	ENN_M	ENN	AD	ML	SPB	SSN
K_{IS}	0,35	0,36	0,14	0,28	0,32	0,19
K_{SP}	0,42	0,89	0,88	0,80	0,93	0,79
K_{SP1}	0,23	0,28	0,13	0,23	0,05	0,20
K_{GD}	0,34	0,77	0,08	0,46	0,33	0,28

Źródło: Opracowanie własne.

² Ensemble Neural Networks (ENN), którego wektor wejściowy składał się wyłącznie z danych mikroekonomicznych, i ENN_M – wektor wejściowy składał się z danych mikro-mezzo-makro.



Rys. 4. Wartości kryteriów lokalnych modeli poddanych ocenie.

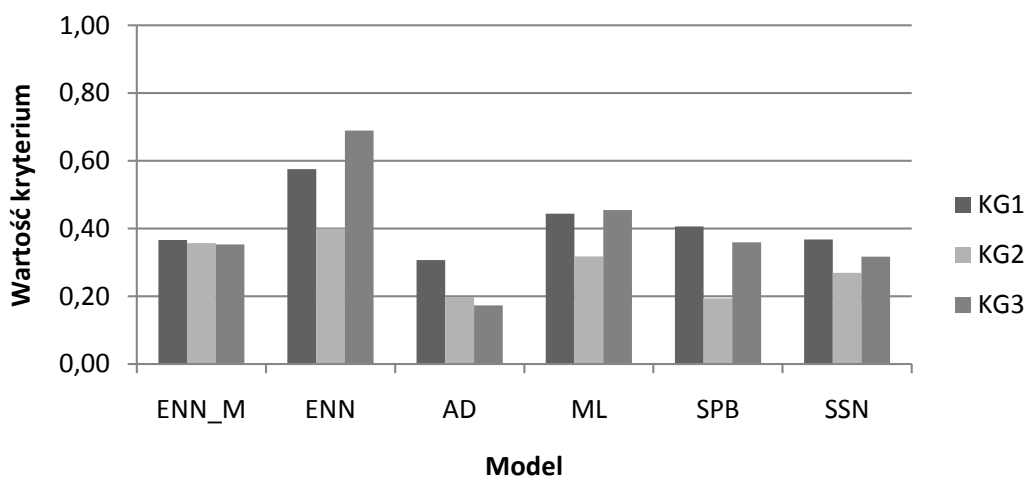
Źródło: Opracowanie własne.

Wartości kryteriów globalnych K_{Gi} (i – numer układu) dla wymienionych wyżej układów rang przedstawia tab. 4 i rys. 5.

Tab. 4. Wartości kryteriów globalnych modeli dla trzech układów rang.

	ENN_M	ENN	AD	ML	SPB	SSN
K_{G1}	0,37	0,58	0,31	0,44	0,41	0,37
K_{G2}	0,36	0,40	0,20	0,32	0,19	0,27
K_{G3}	0,35	0,69	0,17	0,45	0,36	0,32

Źródło: Opracowanie własne.



Rys. 5. Wartości kryteriów globalnych modeli dla trzech układów rang.

Źródło: Opracowanie własne.

Dyskusja wyników

Ocenie poddano 6 modeli. Cztery spośród nich to modele autonomiczne (AD, ML, SPB, SSN). Dwa pozostałe (ENN_M i ENN) stanowią zespoły 10 sztucznych sieci neuronowych, w których zastosowano metodę głosowania do wygenerowania wyjściowego sygnału predykcyjnego.

Przyjęte lokalne kryteria oceny opisują różne aspekty jakości predykcji modeli. Świadczą o tym przedstawione w tab. 3 i na rys. 4, zróżnicowane wartości kryteriów lokalnych. Spośród czterech z nich szczególnie istotne wydają się kryterium optymalnej granicy decyzji i kryterium sprawności I rodzaju. Pierwsze z nich podaje ocenę odporności metody na silnie niezbilansowany zbiór danych użyty do budowy modelu, a drugie ze względu na to, że błąd prognozy I rodzaju jest znacznie bardziej kosztowny dla firmy niż błąd II rodzaju.

Większość poddanych ocenie modeli, z wyjątkiem modelu ENN_M, uzyskała bardzo wysoką i porównywalną wartość kryterium sprawności ogólnej (K_{SP}). Znacznie gorsze oceny uzyskały one w odniesieniu do kryterium sprawności I rodzaju (K_{SP1}). Najlepiej w tym aspekcie prezentują się modele ENN_M i ENN, najgorzej – SPB i AD. W odniesieniu do kryterium optymalnej granicy decyzyjnej (K_{GD}) zdecydowanie najlepszy okazał się model ENN, bardzo dobry wynik osiągnął również model ML, a najslabszy – model AD. Należy zauważyć, że model ENN_M przy średnich wynikach ocen wykazał jako jedyny bardzo dużą stabilność dla wszystkich kryteriów oceny.

Analiza globalnego kryterium oceny (K_G), przeprowadzona na podstawie danych zawartych w tab. 4 i na rys. 5, skłania do wniosku, że niezależnie od przyjętego układu rang najefektywniejszym modelem prognostycznym spośród modeli autonomicznych jest model logitowy ML, a najgorszym – model utworzony w oparciu o metodę analizy dyskryminacyjnej AD. Różnica jakości pomiędzy nimi jest znaczna, bo wynosi ok. 25% wartości kryterium globalnego dla zrównoważonego układu rang. Model sieci przekonań Bayesa SPB tylko nieznacznie ustępuje modelowi ML (ok. 8%). W drugim układzie rang przy zwiększonej wadze sprawności I rodzaju, znacznie wzrasta globalna ocena modelu SSN i dorównuje on modelowi ML.

Niezależnie od przyjętego układu rang zdecydowanie najlepszą ocenę globalną uzyskał model ENN.

Wnioski

Przeprowadzona analiza pokazuje, że jakość ocenianych modeli nie jest wysoka, aczkolwiek nie odbiega od prezentowanych w literaturze, a odnoszących się do podobnych zbiorów badawczych.

Stosunkowo wysoka ocena predykcji stanu ekonomiczno-finansowego małych i średnich przedsiębiorstw zespołu sztucznych sieci neuronowych (ENN) oraz zróżnicowane oceny większości modeli w odniesieniu do różnych kryteriów, aspektów oceny skłania do przekonania, że należy zintensyfikować badania nad heterogenicznymi zespołami złożonymi z rozpatrywanych modeli. Proste głosowanie na ich wyjściu można usprawnić poprzez zastosowanie ważenia głosów w zależności od przyjętych ocen modeli składowych.

Wydaje się, że jednolity, globalny wskaźnik (kryterium globalne K_G) jakości modelu można wykorzystać jako współczynnik w_i , który we wzorze [9] występuje jako waga predykcji każdego z N modeli wchodzących w skład heterogenicznego zespołu predykcyjnego:

$$[9] \quad P = \sum_{i=1}^N w_i * P_i ,$$

gdzie:

P – predykcja przynależności przypadku do określonej klasy firm przez zespół,

P_i – predykcja przynależności przypadku do określonej klasy przez i -ty model zespołu,

w_i – współczynnik wagowy i -tego modelu,

N – liczba modeli wchodzących w skład zespołu.

Aby miara predykcji na wyjściu modelu utrzymana została w stałym przedziale $[0,1]$, należy wielkości wag uzależnione od K_G , unormować zgodnie ze wzorem [10]:

$$[10] \quad w_i = \frac{K_{Gi}}{\sum_{i=1}^N K_{Gi}},$$

gdzie:

w_i – współczynnik wagowy i-tego modelu,

K_{Gi} – wartość kryterium globalnego i-tego modelu,

N – liczba modeli wchodzących w skład zespołu.

Wyraźnie słabsze oceny modeli w odniesieniu do kryteriów K_{JS} , K_{SPI} i K_{GD} przy dobrych ocenach K_{SP} skłaniają do podjęcia zintensyfikowanych badań nad minimalizacją ograniczeń wynikających z dużego niezbilansowania zbiorów badawczych.

Literatura

- BURDA A. (2006): *Prognozowanie kondycji ekonomiczno-finansowej przedsiębiorstw z wykorzystaniem sztucznych sieci neuronowych*, „Barometr Regionalny” nr 6, s. 67-75.
- BURDA A., KUCZMOWSKA B., HIPPE Z.S. (2007): *Ensembles of Artificial Neural Networks for Predicting Economic Situation of Small and Medium Enterprises*, Computer Recognition Systems 2, Springer, s. 808-815.
- DYMOVA L., SEWASTINOW P., FIGAT P. (2004): *Problemy metodologiczne wspomagania decyzji przy ocenie projektów inwestycyjnych oraz odpowiedni system komputerowy*, <http://zsiie.icis.pcz.pl/artykuly/nowe/11a.pdf>.
- GRUSZCZYŃSKI M. (2002): *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza SGH, Warszawa, s. 83-84.
- Główny Urząd Statystyczny (2007): *Polska klasyfikacja działalności (PKD)*, <http://www.stat.gov.pl/klasyfikacje/PKD/pkd.htm>.
- KASJANIUK M. (2006): *Zastosowanie analizy dyskryminacyjnej do modelowania i prognozowania kondycji przedsiębiorstw*, „Barometr Regionalny” nr 6, s. 95-100.
- KOWERSKI M., BIELAK J., DŁUGOSZ D. (2007): *Ocena kondycji małych i średnich przedsiębiorstw w województwie lubelskim i podkarpackim za pomocą logitowych modeli mikro-makro*, „Zamojskie Studia i Materiały” nr 23, s. 83-98.
- KUCZMOWSKA B., BURDA A., HIPPE Z.S. (2007): *Prediction of Economic Situation of Small and Medium Enterprises Using Bayesian Network*, Computer Recognition Systems 2, Springer, s. 802-807.
- NOWAK E. (2006): *Propozycje zmiennych oceniających kondycję ekonomiczno-finansową przedsiębiorstw*, „Barometr Regionalny” nr 6, s. 35-41.