

## WYKORZYSTANIE ANALIZY DYSKRYMINACYJNEJ ORAZ SIECI NEURONOWYCH DO PROGNOZOWANIA SYTUACJI FINANSOWEJ GOSPODARSTW ROLNICZYCH Z WERYFIKACJĄ ROLI CZASU

Joanna Kisielińska

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

**Streszczenie.** Celem badań było przede wszystkim zbudowanie funkcyjnych i sieciowych modeli dyskryminacyjnych, służących do klasyfikacji gospodarstw rolniczych. Efektem klasyfikacji było prognozowanie ich sytuacji finansowej na podstawie wielu cech, obejmujących między innymi zestawy wskaźników finansowych. Modele budowano oddzielnie dla każdego roku, a następnie przeprowadzono ich weryfikacje w latach pozostałych. Cele dodatkowe polegały na wyodrębnieniu cech o największych mocach wyróżniania klas oraz porównaniu obydwu narzędzi – liniowych funkcji dyskryminacyjnych i sieci neuronowych.

**Słowa kluczowe:** analiza dyskryminacyjna, sieci neuronowe

### WSTĘP

Celem prezentowanych badań było wyznaczenie liniowej funkcji dyskryminacyjnej oraz sieci neuronowej do tworzenia prognoz sytuacji finansowej gospodarstw rolniczych. Podstawę konstrukcji modeli dyskryminacyjnych stanowił zestaw wskaźników finansowych uzupełniony o dodatkowe informacje o gospodarstwie, natomiast kryterium klasyfikacji zostało oparte na dochodzie rolniczym. Analizy przeprowadzono na podstawie danych pochodzących z kilku lat, co pozwoliło na zbadanie wpływu czasu na jakość uzyskanych modeli dyskryminacyjnych. Chodziło o sprawdzenie możliwości zastosowania modeli dyskryminacyjnych wyznaczonych dla określonego roku w innych latach.

Cel dodatkowy polegał na określeniu cech o największych zdolnościach prognostycznych, czyli takich, których wpływ na wartość funkcji dyskryminacyjnej jest najistotniejszy.

Modele dyskryminacyjne utworzono na podstawie danych pochodzących z gospodarstw rolniczych prowadzących rachunkowość rolną pod kierunkiem Instytutu Ekonomiki Rolnictwa i Gospodarki Żywnościowej w latach 1992–2002. Obliczenia przeprowadzono korzystając z danych z lat 1992–2001. Rok 2002 służył jedynie do weryfikacji modeli.

Do tworzenia modeli wykorzystano pakiet STATISTICA, natomiast obróbkę danych i analizę wyników wykonano w arkuszu kalkulacyjnym EXCEL wykorzystując język VISUAL BASIC FOR APPLICATION.

## ZASADY BUDOWY MODELI DISKRYMINACYJNYCH

Budowanie modelu klasyfikacyjnego opartego na analizie dyskryminacyjnej wymaga określenia dwóch podstawowych elementów. Pierwszym z nich jest jednoznacznie sformułowane kryterium przypisania obiektów do klas wynikające z potrzeb, jakim klasyfikacja ma służyć. Drugim natomiast dobranie zestaw cech opisujących obiekty, na których opierać się będzie taka klasyfikacja.

Prezentowane dalej modele dyskryminacyjne będą służyły prognozowaniu sytuacji finansowej gospodarstw. Jako miarę tej sytuacji przyjęto dochód rolniczy. Wprawdzie wydaje się, że zysk lepiej spełniałby tę funkcję, jednak dla gospodarstw rolniczych wyznaczenie zysku jest kłopotliwe, ponieważ wymaga założenia pewnych umownych i hipotetycznych kosztów (jak wynagrodzenie za pracę własną rolnika, oprocentowanie własnego kapitału czy renta gruntowa).

Badana zbiorowość zostanie podzielona na dwie klasy. Klasę I będą stanowić gospodarstwa słabsze, natomiast II – lepsze. Wartość dochodu rolniczego rozgraniczającego klasy przyjęto na poziomie mediany. Gospodarstwa osiągające dochód mniejszy od mediany będą zaliczone do klasy I, pozostałe będą stanowiły klasę II.

Podstawę konstrukcji modeli stanowił szeroki zestaw wskaźników finansowych uzupełniony o pewne dodatkowe informacje o gospodarstwach. Zestaw wskaźników finansowych dobrano na podstawie prac: Kulawik [1995], Rachunek ekonomiczny i analiza finansowa [1994] i Wyszowska [1996]. Zestaw ten objął 29 wskaźników, w tym wskaźniki płynności finansowej, rentowności, obrotowości, sprawności gospodarowania, wspomagania finansowego oraz stanowiących charakterystykę majątku trwałego. Dodatkowo uwzględniono powierzchnię użytków rolnych (wyrażoną w ha przeliczeniowych), wiek osoby prowadzącej gospodarstwo, poziom intensywności produkcji oraz poziom intensywności organizacji produkcji (wg Kopcia). Poziom intensywności produkcji oznacza nakłady materiałowo-pieniężne na 1 ha użytków rolnych. Poziom intensywności organizacji produkcji natomiast pokazuje, na ile intensywnie zorganizowana jest działalność. Sposób obliczenia tego wskaźnika, uwzględniający poziom obsady zwierząt gospodarskich oraz strukturę zasiewów, można znaleźć w pracy Olko-Bagieńska, Ziętara [1995].

Należy zaznaczyć, że w celu zapewnienia funkcji prognostycznej modelu kryterium klasyfikacji oparto na dochodzie rolniczym z roku następnego po tym, z którego pochodziły wskaźniki finansowe.

Ponieważ analizy prowadzono na podstawie danych pochodzących z kilku lat, wszystkie wielkości wyrażone w złotych przeliczono na ceny stałe odniesione do roku 1992.

Ze względu na wymagania zastosowanego do obliczeń algorytmu dokonano wstępnej eliminacji cech. Jeśli współczynnik korelacji między dwiema cechami był większy od 0,8, to w dalszych badaniach uwzględniano tę, której poziom skorelowania z dochodem rolniczym był większy. Należy dodać, że w poszczególnych latach wstępnie wyeliminowano zwykle różne wskaźniki, co oznacza, że modele w kolejnych latach wyznaczano na podstawie różnych zbiorów cech.

Przedstawiony problem klasyfikacji gospodarstw rolniczych z punktu widzenia matematyczno-statystycznego jest analogiczny do budowy prognoz ostrzegających przed bankructwem firm oraz prowadzonych przez banki ocen wiarygodności kredytobiorców. Badana zbiorowość jest dzielona na dwie klasy. W przypadku prognoz ostrzegawczych jedną klasę stanowią firmy zagrożone bankructwem, drugą natomiast będące w dobrej kondycji finansowej, którym bankructwo nie zagraża. Banki chcą odróżnić kredytobiorców wiarygodnych od tych, w przypadku których istnieje groźba niespłacenia zaciągniętych zobowiązań.

Problemom przedstawionym powyżej poświęconych jest wiele pozycji literatury, wśród których fundamentalną jest artykuł Altmana [1968]. Jako przykłady konstrukcji prognoz ostrzegających przed bankructwem można podać prace: Altman, Giancarlo, Varetto [1994], Hadasik [1998], Harmol, Czajka, Piechocki [2004], Hołda [2001] i Mączyńska [2004]; ocenie wiarygodności kredytobiorców poświęcono między innymi publikacje Staniec [2004] i Witkowska, Staniec [2002]. Szeroki przegląd systemów wczesnego ostrzegania zawiera praca Zalewskiej [2002].

Budowa typowych systemów wczesnego ostrzegania dla gospodarstw rolniczych nie jest właściwa i celowa, ponieważ w ich przypadku problem bankructwa praktycznie nie występuje. Wynika to z niechęci rolników do zaciągania kredytów oraz z niskich kosztów utrzymania gospodarstwa (niskie podatki, składki emerytalno-rentowe i zdrowotne). W pracy Kisielińskiej [2004] przedstawiono propozycję systemu wczesnego ostrzegania dla gospodarstw. Kryterium klasyfikacji oparto również na dochodzie rolniczym, przyjmując jako wartość graniczną zero.

W wymienionych wyżej publikacjach, dotyczących zastosowania analizy dyskryminacyjnej do prognozowania bankructwa, oceny wiarygodności kredytobiorców czy prognozowania sytuacji gospodarstw rolniczych, pomijany był wpływ czasu. Obliczenia prowadzono w obrębie jednego roku. Naturalne jest wówczas pytanie, czy wyznaczone w ten sposób modele dyskryminacyjne mogą być wykorzystane w latach innych. Próba poszukiwania odpowiedzi na to pytanie są badania przedstawione w niniejszym artykule.

Do tworzenia modeli klasyfikacyjnych wykorzystano dwie metody: analizę dyskryminacyjną oraz sieci neuronowe, w celu porównania ich skuteczności. Przykłady zastosowania sieci neuronowych do zagadnień dyskryminacyjnych można znaleźć np. w pracach: Altman, Giancarlo, Varetto [1994], Kisielińska [2004], Yang, Platt, Platt [1999]. Ich autorzy nie stwierdzili wyraźnej przewagi modeli sieciowych, a niektóre wyniki wskazywały na przewagę metody klasycznej, czyli funkcji dyskryminacyjnej.

## WYNIKI KLASYFIKACJI PRZEPROWADZONEJ PRZY UŻYCIU LINIOWEJ FUNKCJI DYSKRYMINACYJNEJ ORAZ SIECI NEURONOWYCH

W tabeli 1 przedstawiono liczebności zbiorów danych w kolejnych latach oraz medianę dochodu rolniczego w cenach zmiennych i cenach stałych (odniesionych do roku 1992). Mediana dochodu rolniczego była najniższa w 1999 r., a w latach 1994, 2000 i 2001 na zbliżonym niskim poziomie. Wartość najwyższa cechuje rok 1996, a nieco niższe dochody wystąpiły w latach 1995, 1997 i 2002.

Tabela 1. Liczebności zbiorów danych z kolejnych lat oraz mediany dochodu rolniczego wyrażone w cenach stałych i zmiennych

Table 1. The size of data sets in the following years and the median values expressed in constant and changeable prices.

Rok, dla którego budowana jest prognoza	Liczba gospodarstw	Mediana dochodu rolniczego (ceny stałe) PLN	Mediana dochodu rolniczego (ceny zmienne) PLN
1993	663	2 822,44	38 187 570,00
1994	703	2 355,34	42 129 233,00
1995	858	3 541,91	8 096,50
1996	770	3 765,49	10 320,50
1997	997	3 060,78	9 639,00
1998	998	2 673,83	9 414,00
1999	998	1 588,62	6 001,50
2000	912	2 225,81	9 258,00
2001	851	2 468,93	10 834,00
2002	663	2 989,81	13 369,00

Źródło: badania własne.

Source: own calculations.

W celu prognozowania sytuacji finansowej gospodarstwa rolniczego dla każdego roku oddzielnie wyznaczono funkcyjne (w postaci liniowych funkcji dyskryminacyjnych) i sieciowe modele klasyfikacyjne. Modele funkcyjne oznaczone będą dalej jako LFDt, natomiast sieciowe jako SNt, gdzie  $t = 1993, \dots, 2002$  jest rokiem, dla którego jest konstruowana prognoza sytuacji finansowej.

W następnym kroku dane ze wszystkich lat połączono, uzyskując jeden zbiór danych obejmujący 7750 przypadków. Model funkcyjny wyznaczony dla tego zbioru oznaczono jako LFDW, natomiast model sieciowy jako SNW.

Liniowe funkcje dyskryminacyjne obliczono metodą krokowej analizy postępującej. W metodzie tej do modelu kolejno wprowadzane są cechy, które w największym stopniu przyczyniają się do dyskryminacji klas.

Do budowy sieciowych modeli wykorzystano narzędzie zwane automatycznym projektantem, który testuje wiele sieci, sam dobierając ich strukturę i stopień złożoności. Obliczenia dla każdego zbioru danych powtarzane były wielokrotnie i spośród wyznaczonych sieci wybierano najlepszą. W 8 przypadkach siecią najlepszą był perceptron z jedną warstwą ukrytą, w dwóch natomiast były to sieci o radialnych funkcjach bazo-

wych. Należy podkreślić, że obliczenia prowadzące do wyznaczenia sieci neuronowej były długotrwałe, zwłaszcza dla zbioru obejmującego dane ze wszystkich lat.

W tabelach 2 i 3 przedstawiono procent poprawnie rozpoznanych gospodarstw modelami odpowiednio funkcyjnymi i sieciowymi. W wierszach tych tabel podano lata, z których pochodziły dane (dokładnie cechy opisujące gospodarstwa). W kolumnach natomiast określono model wykorzystany do klasyfikacji gospodarstw. Indeks w nazwie modelu jest rokiem, dla którego budowana jest prognoza (aby zapewnić funkcje prognozy, jest on o jeden większy niż rok, z którego pochodzą cechy).

Największa różnica (ponad 15%) w wynikach klasyfikacji przy użyciu liniowych modeli dyskryminacyjnych występuje dla roku 1992. Najlepsze rezultaty uzyskano używając funkcji LFDP<sub>1993</sub> (dla tego roku zbudowano model), najgorsze dla LFDP<sub>1997</sub>. Najbardziej jednolita jest klasyfikacja dla roku 1998. Różnica między modelem najlepszym (LFDP<sub>1999</sub>) i najgorszym (LFDP<sub>1997</sub>) jest poniżej 4%. Średni rozrzut w wynikach klasyfikacji był powyżej 8%.

Na przekątnej tabeli 2 są podane wyniki klasyfikacji gospodarstw, uzyskane za pomocą funkcji wyznaczonej dla roku, z którego pochodziły dane. Należy podkreślić, że w pięciu przypadkach nie był to rezultat najlepszy, gdyż modele w kolejnych latach budowano na podstawie innych zestawów cech. Cechy eliminowano na podstawie wartości współczynników korelacji, co wynika z wymagań stosowanych do obliczeń algorytmów. Uzyskane rezultaty wskazują, że warto zwrócić uwagę na właściwy dobór cech, poprawiając tą drogą jakość klasyfikacji. Współczynniki korelacji nie gwarantują otrzymania najlepszego zestawu cech<sup>1</sup>.

Najlepsze wyniki klasyfikacji uzyskano stosując funkcję oznaczoną jako LFDP<sub>1993</sub>. Dość bezpieczne jest zastosowanie funkcji LFDP<sub>W</sub>, wyznaczonej na podstawie danych ze wszystkich lat. Wprawdzie nie daje ona rezultatów najlepszych, ale również nie są one najgorsze.

Sieci neuronowe dały w zdecydowanej większości przypadków lepsze wyniki klasyfikacji niż modele funkcyjne.

Porównując wyniki klasyfikacji tych samych danych różnymi modelami sieciowymi, należy stwierdzić, podobnie jak w przypadku modeli funkcyjnych, pewne zróżnicowanie. Największa różnica wystąpiła także dla roku 1992 (nieco ponad 11%). Najgorsza okazała się sieć zbudowana do prognozowania kondycji finansowej gospodarstwa w roku 1998 (SNP<sub>1998</sub>), najlepsza natomiast była oczywiście sieć SNP<sub>1993</sub>. Najmniejsze zróżnicowanie uzyskano w roku 1993 (różnica między siecią najlepszą – SNP<sub>1994</sub> i najgorszą – SNP<sub>1998</sub> była poniżej 5%). Średni rozrzut wyników wyniósł 7,5%.

Należy zauważyć, że najwięcej poprawnych klasyfikacji występuje na przekątnej. Oznacza to, że najlepiej spisują się modele sieciowe zbudowane na podstawie danych klasyfikowanych. W przypadku sieci neuronowych nie występuje bowiem konieczność wstępnej eliminacji cech. Model sam dobiera zestaw najlepiej spisujący się w danych warunkach.

---

<sup>1</sup>Właściwy dobór cech umożliwiając algorytmy genetyczne. Ich zastosowanie ma jednak również wady, polegające w przypadku pakietu STATISTICA na eksperymentalnym doborze parametrów regulujących złożoność modeli.

Tabela 2. Wyniki klasyfikacji gospodarstw wykonanych LFDP<sub>t</sub>, gdzie t = 1993, ..., 2001 oraz LFDP<sub>W</sub>Table 2. Results of farms classification done with LFDP<sub>t</sub>, where t = 1993, ..., 2001 and LFDP<sub>W</sub>

Rok, z którego pochodzą wskaźniki finansowe	Liczba gospodarstw	LFDP <sub>1993</sub>	LFDP <sub>1994</sub>	LFDP <sub>1995</sub>	LFDP <sub>1996</sub>	LFDP <sub>1997</sub>	LFDP <sub>1998</sub>	LFDP <sub>1999</sub>	LFDP <sub>2000</sub>	LFDP <sub>2001</sub>	LFDP <sub>W</sub>
		%									
1992	663	81,45	75,72	76,92	68,48	65,91	69,38	70,74	73,30	71,49	72,10
1993	703	76,96	76,96	76,10	69,42	73,12	72,55	72,97	68,71	70,27	73,54
1994	858	82,28	76,92	80,42	74,83	69,58	71,45	72,38	72,26	71,10	72,84
1995	770	78,05	77,01	77,92	78,31	72,73	75,32	71,56	69,48	70,52	74,16
1996	997	75,03	75,53	76,43	76,93	74,82	73,02	71,01	69,01	70,21	73,92
1997	998	79,36	78,46	81,26	79,56	75,05	78,16	76,85	74,15	74,75	77,66
1998	998	77,15	76,35	77,76	74,85	74,45	75,75	78,06	76,45	75,25	77,25
1999	912	80,92	75,44	80,26	73,57	70,83	74,34	74,78	78,84	73,68	75,66
2000	851	82,14	81,32	81,20	78,97	76,26	78,38	80,96	80,26	80,49	79,67
Razem	7750	79,16	77,08	78,80	75,33	72,79	74,48	74,57	73,77	73,23	75,38

Źródło: badania własne.

Source: Own calculations.

Tabela 3. Wyniki klasyfikacji gospodarstw wykonanych SNP<sub>t</sub>, gdzie t = 1993, ..., 2001 oraz SNP<sub>W</sub>Table 3. Results of farms classification done with SNP<sub>t</sub>, where t = 1993, ..., 2001 and SNP<sub>W</sub>

Rok, z którego pochodzą wskaźniki finansowe	Liczba gospodarstw	SNP <sub>1993</sub>	SNP <sub>1994</sub>	SNP <sub>1995</sub>	SNP <sub>1996</sub>	SNP <sub>1997</sub>	SNP <sub>1998</sub>	SNP <sub>1999</sub>	SNP <sub>2000</sub>	SNP <sub>2001</sub>	SNP <sub>W</sub>
		%									
1992	663	83,71	77,53	78,28	77,98	75,41	72,55	79,79	76,77	76,17	78,13
1993	703	78,24	79,09	75,53	74,82	74,11	74,11	78,24	75,96	77,52	77,67
1994	858	83,33	81,24	84,62	82,17	80,42	77,62	81,12	79,95	76,34	82,52
1995	770	78,57	77,27	80,65	85,84	83,38	82,34	82,34	76,75	77,66	82,34
1996	997	75,63	72,72	76,63	79,84	81,85	80,24	77,53	73,52	74,02	77,83
1997	998	78,76	76,55	80,66	83,27	83,37	83,67	81,56	79,66	78,96	82,16
1998	998	75,85	72,75	76,15	78,06	78,56	76,35	80,66	78,66	78,56	79,56
1999	912	80,37	76,10	79,61	78,40	80,37	78,95	79,39	82,24	79,17	82,24
2000	851	80,14	77,91	80,14	81,90	83,90	80,73	83,43	82,73	84,49	84,14
Razem	7750	79,18	76,56	79,14	80,34	80,41	78,78	80,27	78,54	78,11	80,80

Źródło: badania własne.

Source: own calculations.

Sieć neuronowa zbudowana na podstawie danych ze wszystkich lat, podobnie jak jej odpowiednik funkcyjny, również zwykle dawała średnie wyniki klasyfikacji – wprawdzie nie najlepsze, ale i nie najgorsze. We wszystkich latach natomiast była wyraźnie lepsza od modelu funkcyjnego (LFDP<sub>W</sub>). Najmniejsza poprawę klasyfikacji uzyskano w roku 1998 (nieco ponad 2%), największą natomiast dla roku 1994 (prawie 10%). Liczba poprawnych klasyfikacji dla wszystkich lat w przypadku SNP<sub>W</sub> była równa 80,80%, podczas gdy przy użyciu LFDP<sub>W</sub> uzyskano zaledwie 75,38%.

Podsumowując, należy stwierdzić, że zastosowanie modelu dyskryminacyjnego wyznaczanego dla wybranego roku do klasyfikacji danych z innych lat może dawać wyraźnie gorsze rezultaty.

Wielu autorów, widząc zalety analizy dyskryminacyjnej, zwraca uwagę na jej wady. Mączyńska [2004] podkreśla „ograniczenia mechanicznego przenoszenia modeli opracowanych dla warunków konkretnego kraju czy sektora na inny obszar”. Harmol, Czajka i Piechocki [2004] wymieniają dużą wrażliwość modeli na zmiany przedsiębiorstw zebranych w próbie oraz możliwą dezaktualizację modelu, wynikającą z upływu czasu, co potwierdzają przedstawione wyniki klasyfikacji gospodarstw rolniczych.

Wszystkie modele funkcyjne i sieciowe zostały następnie użyte do klasyfikacji danych pochodzących z roku 2001 w celu budowy prognozy sytuacji finansowej gospodarstwa na rok 2002. Wyniki klasyfikacji przedstawiono w tabeli 4.

Największy procent poprawnych klasyfikacji był równy 78,36 i został uzyskany za pomocą dwóch modeli – LFDP<sub>1996</sub> i SNP<sub>1999</sub>. Wynik najgorszy (72,13%) dały SNP<sub>1993</sub> i SNP<sub>1994</sub>. Rozrzut wyników klasyfikacji jest więc nieco większy od 6%. Dobre wyniki uzyskano stosując modele zbudowane na podstawie danych ze wszystkich lat. SNP<sub>W</sub> poprawnie sklasyfikowała 77,16%, podczas gdy LFDP<sub>W</sub> 76,68%.

Prognoza na rok 2002 jest dobrym sprawdzianem skuteczności modeli, ponieważ dane te nie były użyte do budowy żadnego z modeli.

Tabela 4. Wyniki klasyfikacji gospodarstw prognozującej ich sytuację finansową w roku 2002  
Table 4. Results of farms classification done for forecasting their financial situation in 2002

Zastosowana LFDP	Procent poprawnych klasyfikacji przy użyciu LFDP (%)	Procent poprawnych klasyfikacji przy użyciu SNP (%)	Zastosowana SNP
LFDP <sub>1993</sub>	72,70	72,13	SNP <sub>1993</sub>
LFDP <sub>1994</sub>	77,25	72,13	SNP <sub>1994</sub>
LFDP <sub>1995</sub>	76,49	72,99	SNP <sub>1995</sub>
LFDP <sub>1996</sub>	78,39	76,49	SNP <sub>1996</sub>
LFDP <sub>1997</sub>	75,26	77,25	SNP <sub>1997</sub>
LFDP <sub>1998</sub>	75,64	77,73	SNP <sub>1998</sub>
LFDP <sub>1999</sub>	75,83	78,39	SNP <sub>1999</sub>
LFDP <sub>2000</sub>	74,03	75,36	SNP <sub>2000</sub>
LFDP <sub>2001</sub>	74,79	76,11	SNP <sub>2001</sub>
LFDP <sub>W</sub>	76,68	77,16	SNP <sub>W</sub>

Źródło: badania własne.  
Source: own calculations.

## WSKAŹNIKI FINANSOWE O NAJWIĘKSZEJ MOCY PROGNOSTYCZNEJ

Moc dyskryminacyjną cechy w przypadku modeli funkcyjnych określają standaryzowane współczynniki funkcji dyskryminacyjnej. Im są większe co do wartości bezwzględnej, tym udział cechy w modelu jest znaczniejszy. W modelach sieciowych natomiast moc cechy określa jej ranga w tzw. analizie wrażliwości. W jednym i drugim przypadku można łatwo uszeregować cechy w porządku wynikającym z ich znaczenia w poszczególnych modelach.

Wyodrębnienie cech o największej mocy prognozyjnej wymagało jej oceny we wszystkich modelach. Cecha o największej randze w modelu sieciowym lub o największym standaryzowanym współczynniku funkcji dyskryminacyjnej otrzymała 10 punktów, cechy kolejne o jeden punkt mniej. Punkty uzyskane we wszystkich modelach następnie zsumowano i na tej podstawie określono 10 cech o największej mocy dyskryminacyjnej (pozostałe otrzymały znacznie niższe oceny). Zostały one przedstawione w tabeli 5 wraz z formułą obliczeniową i sumą uzyskanych punktów.

Na prognozę największy wpływ ma powierzchnia użytków rolnych, nieco mniejszy wpływ mają dwa wskaźniki rentowności (kapitału własnego i sprzedaży). Kolejne cechy mają ocenę punktową zdecydowanie niższą. Są to wskaźnik produktywności ziemi, poziom intensywności organizacji produkcji, wskaźnik wydajności pracy i wskaźnik cash flow. Znacznie mniejszą rolę w modelach pełnią poziom intensywności produkcji, wskaźnik wyposażenia przedsiębiorstwa w kapitał stały oraz wiek kierownika gospodarstwa.

Tabela 5. Cechy o największych mocach prognozyjnych wraz z oznaczeniami  
Table 5. Indicators with the best forecasting properties and their notations

Nazwa cechy	Formuła obliczeniowa	Suma rang
Powierzchnia użytków rolnych w ha przeliczeniowych (PUR)	–	172
Wskaźnik rentowności kapitału własnego (WRKW)	Dochód rolniczy/Kapitał własny	139
Wskaźnik rentowności sprzedaży (WRS)	Dochód rolniczy/Produkcja końcowa brutto	135
Wskaźnik produktywności ziemi (WPZ)	Produkcja końcowa netto/Powierzchnia użytków rolnych w ha przeliczeniowych	74
Poziom intensywności organizacji produkcji (PIOP)	Obliczony według B. Kopcia	66
Wskaźnik wydajności pracy (WWP)	Produkcja końcowa netto/Nakłady pracy (liczba godzin pracowanych w gospodarstwie)	60
Wskaźnik cash flow (WCF)	(Dochód rolniczy + Amortyzacja)/Produkcja końcowa brutto	56
Poziom intensywności produkcji (PIP)	Nakłady materiałowo-pieniężne na 1 ha użytków rolnych	42
Wskaźnik wyposażenia przedsiębiorstwa w kapitał stały (WWPKS)	Kapitał stały/Pasywa ogółem	35
Wiek kierownika gospodarstwa (WIEK)	–	32

Źródło: badania własne  
Source: own calculations



Spośród dziesięciu cech jedynie cztery to wskaźniki finansowe (WRKW, WRS, WCF i WWPKS), dwa są wskaźnikami sprawności gospodarowania (WPZ i WWP), pozostałe stanowią dodatkowe cechy opisujące gospodarstwo (PUR, PIOP, PIP i WIEK). Modele zbudowane jedynie na podstawie wskaźników finansowych były znacznie niższej jakości i dawały mało trafne prognozy.

## WNIOSKI I PODSUMOWANIE

Przedstawione w artykule wyniki badań pozwalają na wyciągnięcie następujących wniosków:

1. Modele dyskryminacyjne (zarówno funkcyjne, jak i sieciowe) wyznaczone do prognozowania sytuacji finansowej gospodarstw rolniczych na podstawie danych z jednego roku w przypadku zastosowania ich do klasyfikacji w latach innych mogą dawać zdecydowanie gorsze wyniki. Najbezpieczniejsze jest wykorzystanie modeli opracowanych na podstawie danych wieloletnich (wprawdzie nie pozwalają uzyskać wyników najlepszych z możliwych, ale również nie są najgorsze).

2. Sieciowe modele klasyfikacyjne w większości przypadków były lepsze od modeli funkcyjnych. Wyraźną poprawę jakości modeli uzyskano dzięki wielokrotnemu powtarzaniu obliczeń.

3. Porównując modele klasyfikacyjne funkcyjne i sieciowe, można stwierdzić, że pierwsze z nich są znacznie wygodniejsze w użyciu i bardziej uniwersalne. Modele funkcyjne mają bowiem formuły obliczeniowe, sieciowe natomiast nie. Zastosowanie opracowanego modelu sieciowego wymaga konkretnego oprogramowania oraz pliku zawierającego gotową i „nauczoną” sieć. Czas potrzebny do wyznaczenia modelu sieciowego jest znacznie dłuższy niż w przypadku modelu funkcyjnego. Zaletą sieci natomiast jest możliwość poprawiania modelu (można zbudować wiele sieci o różnej strukturze i wybrać z nich najlepszą).

4. Procent poprawnych klasyfikacji gospodarstw rolniczych w celu budowy prognoz ich sytuacji finansowej dla lat 1993–2001 mieścił się w przedziale od 66 do 86%. Uzyskane rezultaty wskazują na konieczność poprawy jakości modeli przez uwzględnienie dodatkowych cech (np. położenie gospodarstwa, wykształcenie kierownika), podział zbiorowości na podgrupy i opracowanie dla nich oddzielnych modeli (np. według typów produkcyjnych czy powierzchni).

## PIŚMIENNICTWO

- Altman E.I., Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* 23, September 1968.
- Altman E.I., Giancarlo M., Varetto F., Corporate Distress Diagnostic: Comparison Using Linear Discriminant Analysis and Neural Networks (the Italian Experience), *Journal of Banking and Finance* 18, 1994.
- Hadasik D., Upadłość przedsiębiorstw w Polsce i metody jej prognozowania, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 1998.

- Harmol M., Czajka B., Piechocki M., Analiza dyskryminacyjna. Przegląd najważniejszych modeli, Przegląd Organizacji 4/2004.
- Hołda A., Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej ZH, Rachunkowość 5, 2001.
- Kisielińska J., Wykorzystanie liniowej funkcji dyskryminacyjnej oraz sieci neuronowych do prognozowania kondycji finansowej gospodarstw rolniczych, Przegląd Statystyczny nr 2, Tom 51/2004.
- Kulawik J., Wskaźniki finansowe i ich systemy w zarządzaniu gospodarstwami rolniczymi, Studia i Monografie, z. 72, IERiGŻ, Warszawa 1995.
- Mączyńska E., Systemy wczesnego ostrzegania, Nowe Życie Gospodarcze, Nr 12, 2004.
- Olko-Bagieńska T., Ziętara W., Materiały do ćwiczeń z organizacji i ekonomiki gospodarstw rolniczych, Wydawnictwo SGGW, Warszawa 1995.
- Staniec I., Analiza dyskryminacyjna w ocenie wiarygodności kredytobiorców, Taksonomia 11, Wrocław 2004.
- Witkowska D., Staniec I., Dychotomiczna klasyfikacja kredytobiorców przy użyciu wielowymiarowej analizy dyskryminacyjnej, Acta Universitatis Lodzensis, Folia Oeconomica 156, 2002.
- Wyszowska Z., Wybrane elementy wskaźnikowej analizy finansowej na przykładzie przedsiębiorstw rolniczych, ATR, Bydgoszcz 1996.
- Yang Z.R., Platt M.B., Platt H.D., Probabilistic neural Networks in Bankruptcy Prediction, Journal of Business Research 44, 1999.
- Zalewska M., Ocena ekonomiczno-finansowa przedsiębiorstwa przez analityka bankowego, Szkoła Główna Handlowa, Warszawa 2002.
- Ziętara W., Kosiarek M., Tehorzewska E., Kondraszuk T., Rachunek ekonomiczny i analiza finansowa w przedsiębiorstwie rolniczym, Materiały metodyczne Centrum Doradztwa i Edukacji w Rolnictwie, Brwinów 2004.

## APPLICATION OF DISCRIMINANT ANALYSIS AND NEURAL NETWORKS TO FORECASTING THE FINANCIAL STANDING OF FARMS WITH CONSIDERATION THE INFLUENCE OF THE TIME

**Abstract.** The aim of the research was to determinate a linear discriminant function and neural network that could be applied for classification of farms. The result of classification was forecasting financial situation of them, based on set of many variables, which including financial indicators. Models were built separately for each year, but they were verified in the rest. The additional aims were to determine the set of indicators with large forecasting ability, and to compare two classification methods – linear discriminant function and neural network.

**Key words:** linear discriminant function, neural networks

Zaakceptowano do druku – Accepted for print: 10.12.2006