

**Urszula Gieraltowska**

Uniwersytet Szczeciński

## **WYKORZYSTANIE NIEKLASYCZNYCH METOD PODZIAŁU DO KLASYFIKACJI SPÓŁEK GIELDOWYCH NOTOWANYCH NA GPW**

### **1. Wstęp**

W ostatnich kilkunastu latach w badaniach ekonomiczno-społecznych coraz większe znaczenie mają metody statystycznej analizy wielowymiarowej. Człowiek, realizując czynności poznawcze w procesie uczenia oraz rozpoznawania, dokonuje podziału i klasyfikacji zdarzeń, ludzi, przedmiotów. Klasyfikacja jest zatem niejednokrotnie traktowana jako podstawowy cel nauki, ponieważ jest jednocześnie narzędziem i celem poznania. Zagadnienia klasyfikacji bardzo często występują w badaniach medycznych, socjologicznych, ekonomicznych czy społecznych, co wynika przede wszystkim z tego, że znaczna część analizowanych zjawisk empirycznych opisywana jest za pomocą wielu zmiennych, czyli ma charakter wielowymiarowy.

Zastosowanie metod analizy wielowymiarowej wymaga spełnienia wielu założeń. Metody klasyczne zakładają na ogół, że analizowane zmienne są mierzone na skalach przedziałowych lub mocniejszych. Ponadto część metod klasycznych w odniesieniu do zastosowania procedur wnioskowania statystycznego na podstawie losowej próby wymaga, aby rozkłady analizowanych zmiennych w populacji miały rozkład normalny. Ważnym ograniczeniem jest konieczność posiadania pełnej i precyzyjnej informacji o grupowanych obiektach. Trzeba dysponować danymi statystycznymi będącymi wartościami tych samych cech dla wszystkich rozważanych obiektów. Brak możliwości pomiaru cechy w przypadku pewnego obiektu powoduje, że nie można obliczyć odległości między nim a pozostałymi obiektami. Nie jest możliwa klasyfikacja obiektów na podstawie napływających danych, ponieważ uzyskany zbiór skupień odzwierciedla jedynie strukturę aktualnego zbioru obiektów. Pojawienie się nowego obiektu wymaga powtórzenia całego procesu klasyfikacji, co pociąga za sobą często wysoki koszt obliczeniowy. W praktyce spełnienie rygorystycznych założeń staje się dużym ograniczeniem metod klasycznych, ponieważ w znacznym stopniu może rzutować na jakość uzyskanych klasyfikacji. Alternatywą staje się zatem wykorzystanie metod,

których stosowanie nie jest obwarowane tak surowymi założeniami. Do takich metod bez wątpienia można zaliczyć metody nieparametryczne oparte m.in. na drzewach klasyfikacyjnych czy sieciach neuronowych, które pozwalają na budowę modelu bez znajomości postaci rozkładów cech oraz związków między nimi.

Drzewa klasyfikacyjne przedstawiają proces podziału zbioru obiektów na jednorodne klasy na podstawie wartości cech obiektów – liście odpowiadają klasom, do których należą obiekty, a krawędzie drzewa reprezentują wartości cech, na podstawie których dokonano podziału. Węzły reprezentują podzbiory, które mogą ulegać dalszym podziałom (w poszczególnych węzłach określone są reguły klasyfikacyjne) [Gatnar 1999]. Drzewa klasyfikacyjne są wykorzystywane w naukach stosowanych, ponieważ za pomocą czytelnego i prostego obrazu graficznego ułatwiają interpretację skomplikowanych zjawisk wielowymiarowych.

Proces budowy drzewa klasyfikacyjnego odbywa się przez rekurencyjny podział zbioru obiektów na podzbiory aż do uzyskania ich homogeniczności ze względu na przynależność do wyodrębnionych klas. Celem jest utworzenie drzewa o jak najmniejszej liczbie węzłów, aby otrzymać proste reguły klasyfikacyjne. Cechami charakterystycznymi drzew klasyfikacyjnych są [Misztal 2000]:

- hierarchiczna natura – polegająca na tym, że zależność liścia od drzewa, na którym rośnie, przez hierarchię podziałów gałęzi prowadzących do ostatniej gałęzi, z której wyrasta liść, można opisać w postaci pytań zadawanych w porządku hierarchicznym na poszczególnych wierzchołkach,
- elastyczność – pozwalająca na badanie pojedynczego wpływu zmiennych na klasyfikację, a także na badanie oddziaływania zmiennych niezależnych różnych typów i osłabienie założeń dotyczących ich pomiaru.

Wśród algorytmów obliczeniowych najczęściej stosowanymi są: CART (*Classification and Regression Trees*), QUEST (*Quick, Unbiased, Efficient Statistical Trees*) oraz CRUISE (*Classification Rule with Unbiased Interaction Selection and Estimation*). Do tworzenia drzew klasyfikacyjnych wykorzystać można m.in. moduł *Drzewa klasyfikacyjne* z pakietu STATISTICA czy program Answer Tree (produkt SPSS). W pracy badania empiryczne zostały przeprowadzone na podstawie pakietu statystycznego STATISTICA 8.0 z wykorzystaniem metody CART, która gwarantuje znalezienie podziału stanowiącego najlepszą klasyfikację próby uczącej.

W procesie klasyfikacji (z wzorcem lub bez wzorca) w ostatnich latach coraz gwałtowniej wzrasta zainteresowanie opartymi na wzorze mózgu człowieka, sztucznymi sieciami neuronowymi, zbudowanymi z neuronów, które przetwarzają informacje w sposób równoległy. Sieci neuronowe stosowane są do rozwiązywania spektrum wielu rozmaitych problemów badawczych związanych z klasyfikacją oraz predykcją w różnorodnych dziedzinach, takich jak: medycyna, fizyka, chemia, inżynieria oraz ekonomia i zarządzanie<sup>1</sup>. Zaletami sieci neuronowych w porówna-

---

<sup>1</sup> Problematyka sztucznych sieci neuronowych jest szeroko omawiana w literaturze, ale nie stanowi to istoty pracy. Wystarczy wspomnieć o pracach: W.S. McCullocha, W. Pittsa, E.M. Azoffa, C.M. Bishopa, R. Tadeusiewicza, J. Korbicza i innych.

niu z metodami klasycznymi są przede wszystkim nieliniowy charakter oraz brak założeń dotyczący danych wejściowych. Istotną cechą jest również umiejętność abstrakcji, która sprawia, że sieć neuronowa może uogólniać wiedzę zdobytą w procesie trenowania. Sztuczne sieci neuronowe w sposób szybki i efektywny przetwarzają dane, często chaotyczne, niekompletne, a nawet sprzeczne, funkcjonując nawet w momencie uszkodzenia lub zerwania połączenia.

Każda sieć, która ma mieć wartość użytkową, musi posiadać wejścia (odpowiadające nerwom sensorycznym w mózgu) oraz wyjścia (odpowiadające nerwom motorycznym), ale jednocześnie może zawierać neurony spełniające wewnętrzne funkcje w sieci, które pośredniczą w analizie informacji (tzw. neurony ukryte). W pracy wykorzystano stabilnie zachowujące się, najpopularniejsze sieci jednokierunkowe, w których przepływ sygnałów odbywa się tylko w jednym kierunku (nie ma sprzężeń zwrotnych). W trakcie procesu uczenia (trenowania) sieć ze wszystkich dostępnych połączeń wybiera tylko te, które są istotne, wykorzystując odpowiedni system wag. Współczynniki wagowe ustalane są dla każdego połączenia i mogą mieć charakter dodatni (połączenie pobudzające), ujemny (połączenie hamujące) lub zerowy (brak połączenia) [Tadeusiewicz 1993]. Do najbardziej znanych algorytmów w sposób iteracyjny pozwalających na wyznaczenie takich wag, które docelowo minimalizować będą odchylenia wartości wygenerowanych od zadanych, należą algorytmy wstecznej propagacji błędów oraz Levenberga-Marguardta.

Badania przeprowadzono na podstawie programu *Statistica Neural Networks*, który ma bardzo użyteczną funkcję automatycznego projektanta sieci, przez co nie wymaga teoretycznego przygotowania od potencjalnego użytkownika. Program ten pozwala w sposób automatyczny wyodrębnić grupy obiektów, które tworzą zbiory: uczący i testowy. Spośród wielu typów i rodzajów sieci neuronowych w pracy wykorzystano najpopularniejszą architekturę sieciową opartą na koncepcji perceptronu wielowarstwowego (MLP – *Multilayer Perceptron*). Założono, że sieci będą mieć tylko jedną warstwę neuronów ukrytych (perceptron trójwarstwowy), natomiast liczba neuronów wejściowych i wyjściowych zdeterminowana będzie kryterium badawczym. Zasadniczym celem sztucznej sieci neuronowej jest znalezienie takiej funkcji, która przydzieli każdy rozpatrywany obiekt do jednej ze zdefiniowanych wcześniej klas. Praktycznie proces dopasowania danych polega na jak najlepszym opisanu punktów rzeczywistych linią. Czasami jest to skomplikowana funkcja o dużych krzywiznach, co grozi nadmiernym dopasowaniem do błędów losowych i nie prowadzi do faktycznej prawidłowości statystycznej.

## 2. Opis grupy badawczej

W procesie badawczym wykorzystano spółki giełdowe notowane na warszawskiej Giełdzie Papierów Wartościowych (GPW), które stanowią zbiór obiektów wielowymiarowych opisanych przez zmienne diagnostyczne o charakterze ogólno-ekonomicznym, finansowym czy też organizacyjnym. Wielowymiarowa analiza

przeprowadzana została w latach 2006-2007 w ujęciu rocznym i kwartalnym w roku 2007. Badanie ograniczono do spółek notowanych w tym okresie w systemie ciągłym i podwójnego fixingu, które znajdowały się w obrocie giełdowym przynajmniej przez cały 2005 rok. Wyodrębniono zbiór 149 spółek giełdowych, dla których uzyskano pełen obraz wskaźników niezbędnych do przeprowadzenia dalszych analiz. Z analizy wyłączono banki, ponieważ zestaw wskaźników ekonomiczno-finansowych dla spółek sektorów niefinansowych nie jest tożsamy z zestawem wskaźników wykorzystywanych w ocenie standingu finansowego banków.

Jednym z najważniejszych etapów wstępnej analizy jest wybór kryterium badawczego pozwalającego na prawidłowe rozróżnienie spółek. Jest to bardzo skomplikowane, ponieważ wybór ten zależy przede wszystkim od indywidualnych preferencji inwestora, zatem powinien odbywać się na analizie merytorycznej uwzględniającej specyfikę przeprowadzanego badania. Zbyt mała rozróżnialność obiektów w wyniku przeprowadzonej klasyfikacji sprawia, że w badaniu należy uwzględnić inne kryterium bądź kryteria badawcze. W artykule zaproponowano kilka kryteriów klasyfikacji spółek giełdowych, ponieważ istotną kwestią miała być ocena stopnia dokładności przeprowadzanej klasyfikacji od przyjętego warunku podziału<sup>2</sup>. Zaproponowano klasyfikację spółek z punktu widzenia ich:

- kondycji ekonomiczno-finansowej mierzonej poziomem miary syntetycznej TMAI<sup>3</sup> (kryterium K I),
- dochodowości mierzonej poziomem stopy zwrotu (kryterium K II),
- zyskowności (kryterium K III),
- zadłużenia (kryterium K IV).

### 3. Wyniki badań

#### 3.1. Klasyfikacja z wykorzystaniem drzew klasyfikacyjnych

Badania zostały przeprowadzone dwuetapowo. W pierwszej kolejności dokonano analizy spółek giełdowych na podstawie danych rocznych. Wyniki uzyskanej klasyfikacji przedstawia tab. 1, natomiast graficzną prezentację podziału przy wykorzystaniu wybranych kryteriów – rys. 1.

---

<sup>2</sup> Każdorazowo dokonywano wstępnego podziału spółek według zamieszczonych kryteriów, przy czym kryterium podziału była mediana badanych wartości. Jeśli rozkład zmiennych ekonomicznych nie jest zbliżony do rozkładu normalnego (co w praktyce jest sytuacją częściej spotykaną), nie można stosować statystycznych miar klasycznych, a wskazane jest wykorzystanie jako miary przeciętnej mediany (wartość środkowa), która pozwala na wyodrębnienie dwóch równych liczebnie grup.

<sup>3</sup> Po raz pierwszy do klasyfikacji spółek ze względu na ich stopień atrakcyjności dla inwestora opracowaną przez siebie miarę syntetyczną wykorzystał W. Tarczyński [1994, s. 275 i dalsze]. W przykładzie jako zmienne diagnostyczne, które posłużyły do oceny siły fundamentalnej spółek, zaproponowano wykorzystanie takich wskaźników, jak: płynność bieżąca, stopa zadłużenia, ROE, rotacja zobowiązań, zysk na 1 akcję.

Tabela 1. Ocena dokładności klasyfikacji spółek giełdowych przy wykorzystaniu drzew klasyfikacyjnych – ujęcie roczne\*

Kryterium I (TMAI)	2006			dokładność	2007			
	wyjściowa	przewidywana			wyjściowa	przewidywana		
		gr. 1	gr. 2			gr. 1	gr. 2	
gr. 1		70	4	88,6	gr. 1	69	5	91,3
gr. 2		13	62		gr. 2	8	67	
Kryterium II (stopa zwrotu)	2006			dokładność	2007			
	wyjściowa	przewidywana			wyjściowa	przewidywana		
		gr. 1	gr. 2			gr. 1	gr. 2	
gr. 1		61	14	84,6	gr. 1	59	15	78,5
gr. 2		9	65		gr. 2	17	58	
Kryterium III (zyskowność)	2006			dokładność	2007			
	wyjściowa	przewidywana			wyjściowa	przewidywana		
		gr. 1	gr. 2			gr. 1	gr. 2	
gr. 1		60	4	83,9	gr. 1	62	12	83,9
gr. 2		10	65		gr. 2	12	63	
Kryterium IV (zadłużenie)	2006			dokładność	2007			
	wyjściowa	przewidywana			wyjściowa	przewidywana		
		gr. 1	gr. 2			gr. 1	gr. 2	
gr. 1		55	19	77,2	gr. 1	63	11	82,6
gr. 2		15	60		gr. 2	15	60	

\* W przypadku badanych kryteriów w grupie 1 znalazły się spółki, które przyjmowały wartości wskaźników powyżej mediany (w przypadku kryteriów I-III są to spółki lepsze z punktu widzenia badanego kryterium, natomiast w przypadku kryterium IV – spółki gorsze o wyższym zadłużeniu niż wynikałoby to z wartości mediany).

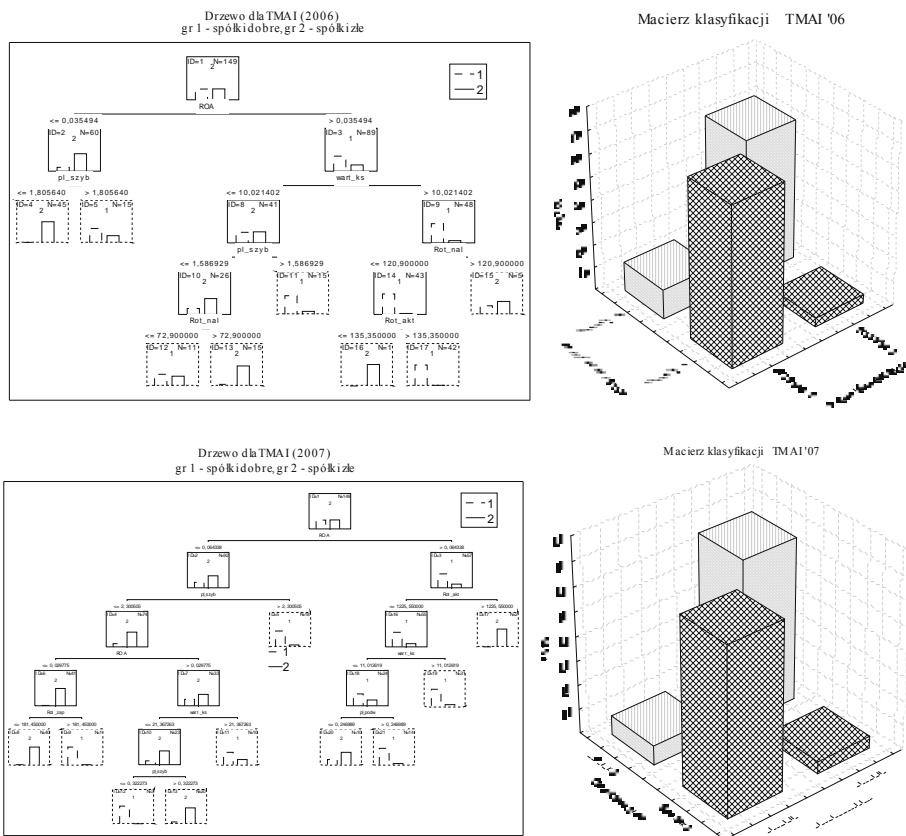
Źródło: opracowanie własne.

Należy zauważyć, że dokładność klasyfikacji spółek przy wykorzystaniu drzew klasyfikacyjnych jest dość wysoka (ponad 80%), a wyznaczone drzewa klasyfikacyjne nie ograniczają się do jednej liniowej kombinacji zmiennych, stąd umożliwiają znalezienie takich punktów, które pozwolą na najlepszą możliwą klasyfikację obiektów.

W drugim etapie dokonano klasyfikacji spółek w ujęciu kwartalnym (w artykule prezentowane są jedynie wyniki dotyczące poszczególnych kwartałów 2007 r.). Wyniki przedstawiono w tab. 2 i na rys. 2.

Podobnie jak w przypadku badania na podstawie wskaźników rocznych, tak i tu należy zauważyć, że dokładność klasyfikacji jest dość wysoka. Najniższą dokładność zanotowano w przypadku badania klasyfikacji spółek, opierając się na ich dochodowości, natomiast najwyższą dokładność otrzymano w przypadku klasyfikacji spółek ze względu na ich kondycję ekonomiczno-finansową. Zwraca uwagę fakt, iż przedstawione drzewa w równym stopniu niepoprawnie klasyfikowały zarówno spółki określane według kryterium jako „dobre”, jak i te „złe”<sup>4</sup>.

<sup>4</sup> Wyraźnie zauważyć można jedynie, iż niezależnie od przyjętego kryterium w II kwartale gorszą jakość klasyfikacji uzyskano dla spółek wyjściowo uznanych jako „dobre”, natomiast w IV kwartale gorszą dokładność otrzymano dla spółek „złych”.



Rys. 1. Drzewa klasyfikacyjne oraz macierz klasyfikacji dla kryterium I (TMAI)

Źródło: opracowanie własne na podstawie *Statistica 8*.

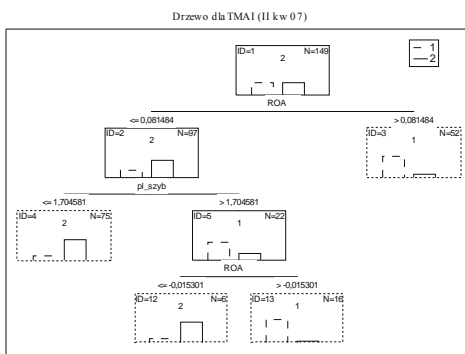
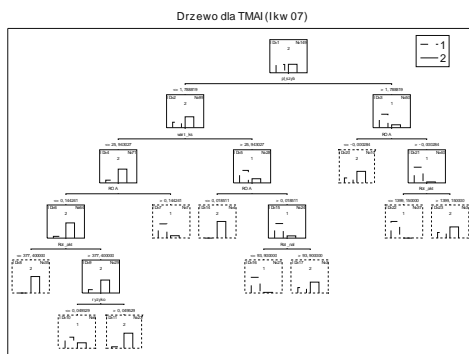
Tabela 2. Ocena dokładności klasyfikacji spółek giełdowych przy wykorzystaniu drzew klasyfikacyjnych – ujęcie kwartalne

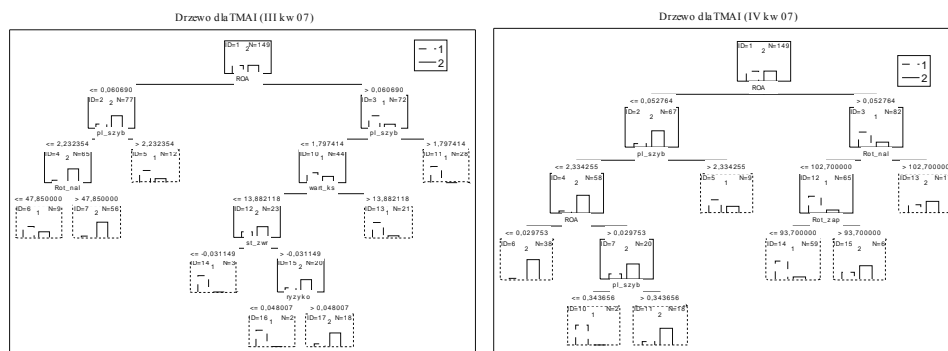
	I kwartał 2007			II kwartał 2007				
	wyjściowa	przewidywana		dokładność	wyjściowa	przewidywana		dokładność
		gr. 1	gr. 2			gr. 1	gr. 2	
Kryterium I (TMAI)	gr. 1	67	7	92,6	gr. 1	60	14	85,2
	gr. 2	4	71		gr. 2	8	67	
	III kwartał 2007			IV kwartał 2007				
	wyjściowa	przewidywana		dokładność	wyjściowa	przewidywana		dokładność
gr. 1	67	7	89,9		gr. 1	61	13	
gr. 2	8	67		gr. 2	9	66		
Kryterium II (stopa zwrotu)	I kwartał 2007			II kwartał 2007				
	wyjściowa	przewidywana		dokładność	wyjściowa	przewidywana		dokładność
	gr. 1	63	11		78,5	gr. 1	55	
	gr. 2	21	54	gr. 2		6	69	

Kryterium II (stopa zwrotu)	III kwartał 2007			IV kwartał 2007					
	wyjściowa	przewidywana		wyjściowa	przewidywana		dokładność		
		gr. 1	gr. 2		gr. 1	gr. 2			
	gr. 1	65	9	83,9	gr. 1	68	6	85,2	
	gr. 2	15	60		gr. 2	16	59		
Kryterium III (zysk na akcje)	I kwartał 2007			II kwartał 2007					
	wyjściowa	przewidywana		wyjściowa	przewidywana		dokładność		
		gr. 1	gr. 2		gr. 1	gr. 2			
		gr. 1	68	6	89,9	gr. 1	66	12	92,6
		gr. 2	9	66		gr. 2	7	72	
	III kwartał 2007			IV kwartał 2007					
wyjściowa	przewidywana		wyjściowa	przewidywana		dokładność			
	gr. 1	gr. 2		gr. 1	gr. 2				
	gr. 1	67	7	93,3	gr. 1	68	6	86,6	
	gr. 2	3	72		gr. 2	14	61		
Kryterium IV (stopa zadłużenia)	I kwartał 2007			II kwartał 2007					
	wyjściowa	przewidywana		wyjściowa	przewidywana		dokładność		
		gr. 1	gr. 2		gr. 1	gr. 2			
		gr. 1	73	1	93,3	gr. 1	63	11	87,2
		gr. 2	9	66		gr. 2	8	67	
	III kwartał 2007			IV kwartał 2007					
wyjściowa	przewidywana		wyjściowa	przewidywana		dokładność			
	gr. 1	gr. 2		gr. 1	gr. 2				
	gr. 1	57	17	84,6	gr. 1	67	7	86,6	
	gr. 2	6	69		gr. 2	13	62		

Źródło: opracowanie własne.

Wiele drzew ma prostą i łatwą do interpretacji strukturę, co z punktu widzenia inwestora przy poszukiwaniu dobrych modeli predykcyjnych pozwala na dokładne zrozumienie natury badanych zależności, które nie do końca można opisać za pomocą jednej zmiennej diagnostycznej. Oczywiście duży wpływ na jakość przeprowadzanych klasyfikacji ma dobór zmiennych diagnostycznych oraz ich dokładność. Zaprezentowany w badaniu dobór zmiennych ma charakter subiektywny,





Rys. 2. Drzewa klasyfikacyjne oraz macierz klasyfikacji dla kryterium I (TMAI) – ujęcie kwartalne  
 Źródło: opracowanie własne na podstawie *Statistica 8*.

jednak opierał się na badaniu wpływu zmiennych diagnostycznych na kryterium podziału (przy wykorzystaniu współczynników korelacji).

### 3.2. Klasyfikacja przy wykorzystaniu sieci neuronowych

Sieci neuronowe budowane były na podstawie zmiennych ekonomiczno-finansowych poszczególnych spółek. Na wyjściu otrzymywana była klasyfikacja według jednego z wcześniej zaprezentowanych kryteriów klasyfikacji (kryteria K I-K IV). Wyniki klasyfikacji oraz architekturę sieci (liczbę neuronów wejściowych–ukrytych–wyjściowych) przedstawiają: tab. 3 (dane roczne) i 4 (dane kwartalne)<sup>5</sup>.

W przypadku badania na podstawie danych rocznych (tab. 3) można zauważyć, że najwyższą dokładność otrzymano dla kryterium wykorzystującego ocenę kondycji badanych spółek (wysoki odsetek poprawnie zaklasyfikowanych obiektów otrzymano dla grupy zarówno uczącej, jak i testującej). W przypadku pozostałych 3 kryteriów dokładność jest już niższa i oscyluje na poziomie 60-70%, przy czym w większość przypadków wyższą dokładność klasyfikacji otrzymano w odniesieniu do grupy uczącej.

Tabela 3. Ocena dokładności klasyfikacji spółek giełdowych przy wykorzystaniu sztucznych sieci neuronowych (MLP) – ujęcie roczne

Kryterium	2006			2007		
	Typ	Dokładność gr. uczącej	Dokładność gr. testowej	Typ	Dokładność gr. uczącej	Dokładność gr. testowej
K I	8-3-2	91,7	96,4	8-6-2	99,2	92,9
K II	9-4-2	65,3	75	9-13-2	75,2	71,4
K III	10-12-2	63,6	60,7	10-8-2	79,3	60,7
K IV	10-4-2	70,3	67,9	10-5-2	66,1	35,7

Źródło: obliczenia własne.

<sup>5</sup> Grupa ucząca stanowiła 80% badanych spółek, natomiast 20% pozostałych spółek był to zbiór testujący.



Z danych zawartych w tab. 4 wynika natomiast, że jeśli wykorzystano się do analiz kwartalne dane ekonomiczno-finansowe, to jakość uzyskanej klasyfikacji jest dużo wyższa (na poziomie ok. 80-90%). Najniższą dokładność (podobnie jak w przypadku drzew klasyfikacyjnych) otrzymano w odniesieniu do kryterium K II (dochodowość).

Tabela 4. Ocena dokładności klasyfikacji spółek giełdowych przy wykorzystaniu sztucznych sieci neuronowych (MLP) – ujęcie kwartalne

Kryterium	Typ	Dokładność grupy uczącej	Dokładność grupy testowej	Typ	Dokładność grupy uczącej	Dokładność grupy testowej
K I	I 2007			II 2007		
	8-10-2	99,2	89,3	8-6-2	93,4	89,3
	III 2007			IV 2007		
	8-6-2	91,7	85,7	8-10-2	94,2	89,3
K II	I 2007			II 2007		
	9-12-2	81,8	64,3	9-7-2	71,9	71,4
	III 2007			IV 2007		
	9-11-2	81,8	64,3	9-7-2	71,9	71,4
K III	I 2007			II 2007		
	10-7-2	98,3	78,6	10-7-2	80,2	96,4
	III 2007			IV 2007		
	10-5-2	95	92,9	10-8-2	90,1	85,7
K IV	I 2007			II 2007		
	10-7-2	89,3	89,3	10-12-2	82,6	82,1
	III 2007			IV 2007		
	10-12-2	94,2	82,1	10-9-2	84,3	85,7

Źródło: obliczenia własne.

Podkreślić również należy, że niższe błędy klasyfikacji uzyskano, wykorzystując drzewa klasyfikacyjne w procesie rozpoznawania. W większości przypadków wygenerowane sztuczne sieci neuronowe gorzej klasyfikują spółki w porównaniu z wcześniej zaprezentowaną metodą.

#### 4. Porównanie wybranych klasycznych i nieklasycznych metod klasyfikacji

W celu zobrazowania wcześniej postawionej tezy, że wykorzystanie klasycznych metod klasyfikacji przy niespełnionych założeniach może rzutować na jakość klasyfikacji, postanowiono porównać wyniki klasyfikacji na podstawie sieci neuronowych i drzew klasyfikacyjnych z klasyfikacją przy wykorzystaniu funkcji dyskryminacyjnej. Istota funkcji dyskryminacyjnej jest szeroko omawiana w literaturze [Jajuga 1993], należy jednak podkreślić, że funkcja dyskryminacyjna ma tę przewagę nad prezentowanymi metodami nieklasycznymi, że nie tylko klasyfikuje obiekty do wyodrębnionych klas, ale pozwala również na ich uporządkowanie, co z punktu widzenia procesu inwestowania jest bardzo istotną kwestią. Dokładność klasyfikacji przedstawiono w tab. 5.

Tabela 5. Ocena dokładności klasyfikacji spółek giełdowych przy wykorzystaniu klasycznych i nieklasycznych metod podziału\*

Kryterium	FD	DK	SN	FD	DK	SN
K I	2006			2007		
	59,4	88,6	91,7	85,2	91,3	99,2
	I 2007			II 2007		
	83,9	92,6	99,2	83,2	85,2	93,4
	III 2007			IV 2007		
K II	2006			2007		
	64,4	84,6	65,3	63,1	78,5	75,2
	I 2007			II 2007		
	65,8	78,5	81,8	59,1	83,2	71,9
	III 2007			IV 2007		
K III	2006			2007		
	57,7	83,9	63,6	53,7	83,9	79,3
	I 2007			II 2007		
	82,6	89,9	98,3	84,6	92,6	80,2
	III 2007			IV 2007		
K I	2006			2007		
	55,7	77,2	70,3	55,7	82,6	66,1
	I 2007			II 2007		
	83,2	93,3	89,3	74,5	87,2	82,6
	III 2007			IV 2007		
	80,5	84,6	94,2	78,5	86,6	84,3

\* FD – funkcja dyskryminacyjna, DK – drzewa klasyfikacyjne, SN – sztuczne sieci neuronowe.

Źródło: obliczenia własne.

Z danych zawartych w tab. 5 wynika jednoznacznie, że każdorazowo, niezależnie od przyjętego kryterium podziału i okresu badawczego, dokładność klasyfikacji uzyskana przy wykorzystaniu funkcji dyskryminacyjnej przyjmuje najniższe wartości. Najczęściej najwyższą dokładnością klasyfikacji charakteryzują się drzewa klasyfikacyjne. W literaturze [Krause 1993] zwraca się uwagę, że stosowanie analizy dyskryminacyjnej oraz sieci neuronowych pozwala na uzyskiwanie zbliżonych wyników mierzonych poziomem błędów pierwszego i drugiego rodzaju, choć spotkać się można ze stwierdzeniem, że analiza dyskryminacyjna daje niższy błąd pierwszego rodzaju, natomiast sieci neuronowe – błąd drugiego rodzaju. Właściwość ta nie zawsze jest zauważalna w odniesieniu do badań na rynku kapitałowym.

Bardzo ciekawym zagadnieniem jest również sprawdzenie, jakie zdolności predykcyjne mają prezentowane metody. Potraktowanie prób badawczych jako prób uczących pozwala wykorzystać otrzymane reguły klasyfikacyjne do prognozowania przynależności obiektów w kolejnych okresach i ocenić jakość otrzymanych prognoz (mierzoną poziomem błędnej lub poprawnej klasyfikacji). Ze względu

du na ograniczenia objętościowe artykułu chciałabym tylko przytoczyć ogólne wyniki. Dokładność prognoz w przypadku badanych obiektów była bardzo zróżnicowana – od ok. 20% do ok. 70%. Podobne wyniki otrzymano we wcześniejszych badaniach [Gierałtowska 2005].

Omówione algorytmy nie wyczerpują problematyki konstrukcji reguł klasyfikacyjnych. W literaturze proponuje się również wykorzystanie zbliżonych do sieci neuronowych – algorytmów genetycznych, które cechuje uniwersalność, elastyczność i skuteczność oraz duża odporność na zakłócenia występujące w danych wejściowych. Możliwe jest również wykorzystanie modeli logitowych oraz łańcuchów Markowa. Założeń co do charakteru danych wejściowych nie wymaga także m.in. rozmyta metoda analizy skupień  $k$ -średnich.

## Literatura

- Gatnar E., *Drzewa klasyfikacyjne: nieparametryczna metoda dyskryminacji obiektów*, Badania Operacyjne i Decyzyjne nr 1, Wrocław 1999.
- Gierałtowska U., *Wykorzystanie funkcji dyskryminacyjnej do podejmowania optymalnych decyzji na rynku kapitałowym*, praca doktorska, Szczecin 2005.
- Jajuga K., *Statystyczna analiza wielowymiarowa*, Wydawnictwo Naukowe PWN, Warszawa 1993.
- Krause C., *Kreditwürdigkeitsprüfung mit Neuronalen Netzen*, [w:] *Schriften des Instituts für Revisionswesen der Westfälischen Wilhelms*, Universität Münster, Düsseldorf 1993.
- Miształ M., *Wykorzystanie drzew klasyfikacyjnych do wspomagania procesów podejmowania decyzji*, materiały StatSoft Polska, Kraków 2000.
- Tadeusiewicz R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza, Warszawa 1993.
- Tarczyński W., *Taksonomiczna miara atrakcyjności inwestycji w papiery wartościowe*, „Przegląd Statystyczny” 1994, R. XLI, z. 3.

## THE UTILIZATION OF NONCLASSICAL METHODS FOR CLASSIFICATION OF THE COMPANIES NOTED ON THE WARSAW STOCK EXCHANGE

### Summary

The subject of the elaboration is the utilization of Neural Networks (MPL – Multilayer Perceptron) and Classification Trees for the classification of the companies quoted on the Warsaw Stock Exchange. The author uses economic and financial indicators and the market ones for the years 2006-2007 in the yearly and quarterly perspective. The article presents the estimation of quality classification depending on the criterion of companies discriminations. One-element criteria (rate of return, debt repayment, profit) and complex criteria (financial standing) are used. The results of the classification are acceptable. The comparison of precision of classical classification (based on discriminant function) and non-classical methods of division are made. The best results are received for Classification Trees.