

**Bartłomiej Lach**

Uniwersytet Ekonomiczny w Poznaniu

# **SELEKCJA AKCJI Z WYKORZYSTANIEM MODELI DYSKRYMINACYJNYCH ORAZ OPTYMALIZACJI PODZIAŁU ZBIORU OBIEKTÓW**

## **Wprowadzenie**

Wybór odpowiedniej spółki oraz właściwego momentu zakupu jej akcji stanowi podstawowe zadanie stojące przed każdym podmiotem inwestującym środki w akcje. Korzystając z technik analizy rynku kapitałowego inwestor ma możliwość podejmowania decyzji inwestycyjnych pozwalających zwiększać szansę efektywnego inwestowania na giełdzie. Celem przeprowadzanego badania było ustalenie zasadności oraz skuteczności stosowania analizy dyskryminacyjnej jako narzędzia analizy fundamentalnej w procesie doboru spółek do portfela inwestycyjnego. Przeprowadzone badanie powinno umożliwić odpowiedź na pytanie, czy analiza dyskryminacyjna wsparta wykorzystaniem iteracyjnej metody optymalizacji podziału zbioru obiektów może stanowić narzędzie skutecznie wspomagające proces podejmowania decyzji inwestycyjnych. Zadaniem szacowanych modeli dyskryminacyjnych był dobór spółek do portfela inwestycyjnego w taki sposób, aby uzyskiwana przez portfel stopa zwrotu była wyższa niż średnia stopa wszystkich badanych spółek w danym okresie. Podczas konstrukcji modeli dyskryminacyjnych dodatkowo wykorzystano iteracyjną metodę optymalizacji podziału zbioru obiektów – metodę k-średnich. Została ona wykorzystana na etapie definiowania populacji, z których mogą pochodzić kolejne obserwacje. Metoda k-średnich została zastosowana w celu poprawy zdolności poprawnego klasyfikowania przez modele dyskryminacyjne, zwiększając tym samym szansę opracowania skutecznego narzędzia prognostycznego.

W badaniu poprawności klasyfikacji skonstruowanych modeli dyskryminacyjnych wykorzystano obiekty próby uczącej, na podstawie których skonstruowano modele dyskryminacyjne oraz obiekty próby testującej, których nie uwzględniano podczas budowania modeli.

## 1. Metodologiczne podstawy przeprowadzonego badania

Podstawowym narzędziem wykorzystanym w badaniu jest analiza dyskryminacyjna. Służy ona klasyfikowaniu obiektów na podstawie opisujących je cech diagnostycznych. Każdy obiekt opisywany jest przez zestaw cech wyrażonych za pomocą  $k$ -wymiarowego wektora zmiennych niezależnych. Główna idea analizy polega na rozdzieleniu zbioru wszystkich badanych obiektów, stanowiących próbę  $\Omega = \{Q_1, Q_2, \dots, Q_N\}$  na rozłączne populacje, z których mogły te obiekty pochodzić:  $\pi_0, \pi_1, \dots, \pi_{n-1}$ <sup>1</sup>. Populacje odpowiadają wariantom zmiennej zależnej, mierzonej na skali nominalnej (możliwe  $n$  wariantów zmiennej). Podzielony w ten sposób zbiór wszystkich dostępnych obiektów służy do wyznaczenia funkcji dyskryminacyjnej, stanowiącej następnie narzędzie „rozpoznające” obiekty nieznanego pochodzenia. Dodatkową korzyścią uzyskiwaną przez badacza stosującego analizę dyskryminacyjną jest wiedza odnośnie do stopnia wpływu poszczególnych zmiennych niezależnych na możliwość rozróżnienia obiektów pochodzących z różnych populacji. Informować o tym mogą współczynniki stojące przy poszczególnych zmiennych funkcji dyskryminacyjnych<sup>2</sup>. Konstruowana funkcja umożliwi klasyfikację obiektu na podstawie wartości jego cech diagnostycznych do jednej z ustanowionych populacji  $\pi_0, \pi_1, \dots, \pi_{n-1}$ .

Przed rozpoczęciem konstrukcji funkcji dyskryminacyjnej niezbędna jest weryfikacja odpowiednich założeń stawianych zmiennym diagnostycznym, opisującym obiekty badania. Wybór zastosowanych w modelu zmiennych powinien być merytorycznie uzasadniony, a wnoszone przez nie informacje powinny ujmować najbardziej istotne własności analizowanego zjawiska<sup>3</sup>. Zmienne diagnostyczne nie powinny być ze sobą skorelowane oraz powinny posiadać rozkład normalny. W praktyce, brak spełnienia założenia normalności często nie jest traktowany jako powód do zaprzestania dalszego badania.

Na podstawie funkcji gęstości populacji, z których mogą pochodzić obiekty badania oraz założeń bayesowskiej reguły klasyfikacyjnej<sup>4</sup> liniową funkcję dyskryminacyjną w przypadku populacji obiektów  $\pi_0$  i  $\pi_1$  można zapisać następująco<sup>5</sup>

<sup>1</sup> D. Hadasik: Upadłość przedsiębiorstw w Polsce i metody jej prognozowania. Akademia Ekonomiczna, Poznań 1998, s. 97.

<sup>2</sup> M. Walesiak: Metody analizy danych marketingowych. Wydawnictwo Naukowe PWN, Warszawa 1996, s. 103.

<sup>3</sup> E. Nowak: Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych. Państwowe Wydawnictwo Ekonomiczne, Warszawa 1990, s. 23.

<sup>4</sup> Więcej informacji o bayesowskiej regule klasyfikacyjnej można znaleźć w pracy D. Hadasik: Op. cit.

<sup>5</sup> Ibid., s. 115.

$$D(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + \mathbf{a}_0 \quad (1)$$

$$\mathbf{a} = \Sigma^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1),$$

$$\mathbf{a}_0 = -\frac{1}{2}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T \Sigma^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1),$$

gdzie:

$\mathbf{a}$  – wektor współczynników liniowej funkcji dyskryminacyjnej,

$\mathbf{a}_0$  – wyraz wolny liniowej funkcji dyskryminacyjnej,

$\mathbf{x}$  – wektor cech obiektu,

$\Sigma$  – macierz wariancji-kowariancji dla zmiennych opisujących obiektu,

$\boldsymbol{\mu}_0$  ( $\boldsymbol{\mu}_1$ ) – wektor wartości oczekiwanych zmiennych w populacjach  $\pi_0$  ( $\pi_1$ ).

W przypadku równej liczby obiektów próby uczącej pochodzących z obu populacji reguła klasyfikacyjna zakłada, że obiekt należy zaklasyfikować do populacji pierwszej ( $\pi_0$ ) w sytuacji, kiedy wartość funkcji dyskryminacyjnej dla obiektu jest większa od zera. Wartości ujemne funkcji dyskryminacyjnej wskazują na przynależność obiektu do populacji drugiej ( $\pi_1$ )<sup>6</sup>. W przypadku występowania większej liczby klas, z których mogą pochodzić obiekty badania, zasada postępowania jest inna. W takiej sytuacji wyznaczane są funkcje klasyfikacyjne dla każdej populacji z osobna. Klasyfikacja odbywa się na drodze wyznaczenia wartości wszystkich funkcji klasyfikacyjnych dla każdego obiektu z osobna. Analizowany obiekt zaklasyfikowany zostaje do populacji, dla której odpowiadająca jej funkcja klasyfikacyjna wskazała najwyższą wartość<sup>7</sup>.

Drugą ważną metodą wykorzystaną w badaniu była metoda k-średnich umożliwiająca optymalny podział zbioru obiektów. Metoda ta służy do skorygowania w sposób iteracyjny wstępnego i z góry znanego podziału zbioru obiektów. Zadaniem metody k-średnich jest przemieszczenie poszczególnych obiektów pomiędzy klasami w taki sposób, aby suma odległości wszystkich obiektów od środków ciężkości klas, do których przynależą była jak najmniejsza. W sytuacji, kiedy po kolejnej iteracji żaden obiekt nie zmieni swojego położenia, algorytm kończy swoje działanie, a aktualnie ustanowiony podział uznawany jest za optymalny. W jednej z najczęściej wykorzystywanych wersji metody k-średnich, opracowanej przez J.A. Hartingana, odległość obiektów mierzona jest za pomocą miary euklidesowej<sup>8</sup>.

<sup>6</sup> G.J. McLachlan: Discriminant Analysis and Statistical Pattern Recognition. John Wiley & Sons, Inc., Hoboken, New Jersey 2004.

<sup>7</sup> Elektroniczny podręcznik pakietu STATISTICA. Analiza dyskryminacyjna. www.statsoft.pl

<sup>8</sup> T. Grabiński: Metody taksonometrii. Akademia Ekonomiczna, Kraków 1992, s. 125.

## 2. Wykorzystany materiał badawczy

W badaniu wykorzystano dane dotyczące działalności spółek pochodzących z trzech branż: budowlanej, informatycznej oraz bankowej. Do konstrukcji oraz testowania modeli dyskryminacyjnych wykorzystano łącznie 61 spółek notowanych na Giełdzie Papierów Wartościowych (25 – branża budowlana, 20 – branża informatyczna, 16 – banki). Każdą ze spółek opisuje zestaw wskaźników finansowych wyznaczanych na podstawie kwartalnych sprawozdań finansowych począwszy od pierwszego kwartału 2008 roku. Ostatnim kwartałem uwzględnionym w badaniu był kwartał drugi 2012 roku. Próbę uczącą stanowiło czternaście kwartałów (ostatnim był drugi kwartał 2011 roku), a testującą – cztery (od trzeciego kwartału 2011 roku do drugiego kwartału 2012 roku).

Wskaźniki finansowe dobierane były tak, aby w możliwie pełny sposób opisywały sytuację finansowo-majątkową przedsiębiorstw. Kierując się tą zasadą w zbiorze zmiennych diagnostycznych znalazły się wskaźniki opisujące różne aspekty kondycji finansowo-majątkowej przedsiębiorstw (rentowność, zadłużenie, płynność finansowa, sprawność działania)<sup>9</sup>. Wstępny zbiór cech diagnostycznych został ograniczony ze względu na niespełnienie przez niektóre z nich wymagań stawianych przez analizę dyskryminacyjną. Ostatecznie zbiór cech diagnostycznych stanowiło 6 wskaźników finansowych: płynność bieżąca, wskaźnik zadłużenia kapitału własnego, udział kapitału własnego w aktywach trwałych, rentowność aktywów ogółem, wskaźnik rotacji aktywów ogółem oraz wskaźnik źródeł finansowania inwestycji. W celu zapewnienia porównywalności obserwacji pochodzących z różnych okresów zastosowana została standaryzacja wartości wskaźników finansowych w poszczególnych okresach. Skutkiem zastosowania normalizacji jest utrata informacji o bezwzględnej wielkości wskaźników w okresach na rzecz wartości wskazujących na poziom wskaźnika na tle innych spółek w danym okresie<sup>10</sup>.

Drugą grupą danych opisujących obiekty badania są stopy zwrotu akcji spółek uwzględnionych w badaniu. Stopy zwrotu wyliczane były dla trzymiesięcznych okresów po publikacji kwartalnych sprawozdań finansowych. Spółki notowane na GPW w Warszawie po zakończeniu każdego kwartału mają 45 dni na opublikowanie okresowego sprawozdania finansowego<sup>11</sup>. Sprawo-

<sup>9</sup> M. Jerzemowska: Analiza finansowa w zarządzaniu przedsiębiorstwem. Tom 1. Red. L. Bednarski i T. Waśniewski. FRR w Polsce, Warszawa 1996, s. 321.

<sup>10</sup> B. Lach: Metody i zastosowania badań operacyjnych w gospodarce, finansach i szkolnictwie wyższym. Red. W. Sikora. Uniwersytet Ekonomiczny, Poznań 2013 (w druku).

<sup>11</sup> Ustawa o rachunkowości z dnia 29 kwietnia 1994 r. Dz.U. 2009, nr 152, poz. 1223.

zdania finansowe poszczególnych przedsiębiorstw nie są zatem publikowane tego samego dnia. W badaniu przyjęto dla uproszczenia jeden wspólny termin publikacji sprawozdań finansowych przez wszystkie analizowane spółki. Dniem tym jest trzydziesty dzień po zakończeniu kwartału, którego dotyczyło sprawozdanie finansowe. Przykładowy okres obserwacji cen akcji po opublikowaniu sprawozdań finansowych za pierwszy kwartał 2008 roku rozpoczyna się 01.05.2008 roku, a kończy 30.07.2008 roku.

### 3. Warianty oraz założenia badania empirycznego

Celem stosowania modeli dyskryminacyjnych w badaniu jest typowanie spółek, które z punktu widzenia fundamentalnego dają szansę na wypracowania ponadprzeciętnych stóp zwrotu w okresach po publikacji kwartalnych sprawozdań finansowych. W badaniu empirycznym wykorzystano zebrane dane finansowe dotyczące 61 spółek pochodzących z trzech różnych branż. Zróżnicowany rodzaj działalności spółek w poszczególnych branżach może mieć wpływ na odmienne kształtowanie się niektórych wskaźników finansowych utrudniając tym samym ich porównywalność. Z tego powodu badanie przeprowadzono w dwóch wariantach. Pierwszy wariant zakłada brak rozróżnienia spółek ze względu na branżę, z której pochodzą. W przypadku pierwszego wariantu, przy konstrukcji oraz testowaniu modeli dyskryminacyjnych korzysta się z próby, do której należą spółki pochodzące z wszystkich trzech analizowanych branż. Drugi z wariantów uwzględnia przynależność obiektów do jednej z trzech badanych branż. Związane jest to z konstrukcją oddzielnych modeli dyskryminacyjnych dla każdej branży z osobna. Wariant drugi zakłada więc konieczność oddzielenia od siebie spółek prowadzących odmienny rodzaj działalności, który może wpływać na odmienne kształtowanie się wskaźników finansowych spółek w poszczególnych branżach.

Działalność przedsiębiorstw niektórych branż podlega silnemu oddziaływaniu czynników sezonowych. Przykładem tego może być branża budowlana, w której bardziej korzystnym okresem dla działalności przedsiębiorstw jest druga połowa roku<sup>12</sup>. Ze względu na możliwość występowania czynników sezonowych mogących wpływać na działalność przedsiębiorstw oraz na kształtowanie

---

<sup>12</sup> Wpływ czynnika sezonowego w branży budowlanej można zaobserwować na wykresie prezentującym dynamikę produkcji budowlano-montażowej w Polsce w ostatnich latach. Strona internetowa Głównego Urzędu Statystycznego. [www.stat.gov.pl/gus](http://www.stat.gov.pl/gus)

się niektórych wskaźników finansowych, w badaniu zdecydowano się na budowę odrębnych modeli dyskryminacyjnych, dla każdego z czterech kwartałów. Budowa oddzielnych modeli ma na celu spowodowanie lepszego dopasowania współczynników funkcji dyskryminacyjnych do wskaźników finansowych spółek w poszczególnych kwartałach. Możliwa jest bowiem sytuacja, w której o sukcesie spółki na giełdzie po opublikowaniu sprawozdania finansowego za pierwszy kwartał decydują w największym stopniu np. wskaźniki związane z zadłużeniem spółki, a po publikacji sprawozdań za kwartał trzeci – wskaźniki rentowności. Zastosowanie odrębnych modeli dyskryminacyjnych dla każdego kwartału z osobna powinno umożliwić wykrycie takich zależności. W efekcie, do typowania spółek, które z punktu widzenia fundamentalnego powinny uzyskiwać ponadprzeciętne stopy zwrotu w okresie po publikacji sprawozdań finansowych konieczne jest wykorzystanie jednego z czterech modeli. Podejście to dotyczy obu przyjętych wariantów badania.

Istotnym elementem badania jest wykorzystanie metody iteracyjnej optymalizacji podziału zbioru obiektów – metody k-średnich na etapie definiowania populacji, do których klasyfikowane są obiekty badania. Pamiętając o konstrukcji funkcji dyskryminacyjnej należy zauważyć, że jakość modelu oraz zdolność do prawidłowego przyporządkowywania obiektów do populacji zależy w dużej mierze od poziomu istotności różnic obiektów w poszczególnych populacjach. Ze względu na przyjęte cechy diagnostyczne obiekty pochodzące z obu populacji powinny istotnie różnić się między sobą, ułatwiając w ten sposób poprawną klasyfikację. W tym właśnie celu zastosowana została metoda k-średnich, która ma podzielić obiekty próby uczącej w taki sposób, aby te znajdujące się wewnątrz grupy były możliwie podobne oraz znacząco różniły się od obiektów spoza grupy.

Obiekty znajdujące się w poszczególnych grupach dla podziałów na 4, 5 oraz 6 grup zbadane zostały pod kątem uzyskiwanych stóp zwrotu na giełdzie w okresie po publikacji kwartalnych sprawozdań. Dla każdej grupy wyznaczana była średnia wartość zestandaryzowanych stóp zwrotu z akcji spółek należących do poszczególnych grup. Wartość ta określała, jak średnio biorąc prezentowały się stopy spółek tej grupy na tle wszystkich obiektów w danym kwartale. Grupa, której średnia wartość zestandaryzowanych stóp zwrotu w modelu dla kwartału była dodatnia i najwyższa stanowiła w badaniu populację spółek atrakcyjnych<sup>13</sup>.

---

<sup>13</sup> Warunkiem uznania grupy za atrakcyjną była również wystarczająco reprezentatywna liczba znajdujących się w niej obiektów, zbliżona do średniej liczebności obiektów w ustanowionych grupach.

Obiekty z pozostałych grup, wobec przyjętych założeń, nie leżą w polu zainteresowania inwestorów. Podział z wykorzystaniem metody k-średnich przeprowadzony został w przypadku obu wariantów badania. Pierwszemu wariantowi badania, w którym nie dokonywano rozróżnienia obiektów ze względu na branżę, z której pochodziły spółki odpowiada tabela 1. W drugim wariantcie badania, dla każdej branży konstruowane były oddzielne modele dyskryminacyjne. Ustanawianie populacji obiektów należało zatem wykonać dla każdej z trzech branż z osobna. W tabeli 2 przedstawiono wyniki postępowania dla branży informatycznej.

Tabela 1

Grupy obiektów pochodzących z trzech branż po optymalizacji podziału metodą k-średnich

Podział na k grup	Model dla kwartału								
	I		II		III		IV		
	Średnia, zestandaryzowana stopa zwrotu w grupie	$N_i$	Średnia, zestandaryzowana stopa zwrotu w grupie	$N_i$	Średnia, zestandaryzowana stopa zwrotu w grupie	$N_i$	Średnia, zestandaryzowana stopa zwrotu w grupie	$N_i$	
4	1	0,343	5	-0,041	55	-0,249	12	0,134	56
	2	-0,107	159	0,062	152	<b>0,237</b>	<b>39</b>	-0,768	12
	3	0,191	57	0,93	1	-0,042	119	-0,88	6
	4	0,367	12	-0,325	25	-0,411	3	0,042	157
	Razem:		233		233		173		231
5	1	-0,046	137	<b>0,118</b>	<b>104</b>	0,014	153	-0,725	4
	2	-0,267	30	0,93	1	-0,183	10	0,086	81
	3	0,191	57	-0,041	85	-0,411	3	-0,880	6
	4	0,347	5	-0,302	10	2,95	1	-0,166	96
	5	0,442	4	-0,306	33	-0,328	6	<b>0,377</b>	<b>44</b>
	Razem:		233		233		173		231
6	1	0,442	4	-0,189	19	0,047	18	-0,981	3
	2	0,347	5	-0,041	55	-0,175	7	0,377	44
	3	<b>0,191</b>	<b>57</b>	0,93	1	2,95	1	-0,880	6
	4	-0,466	17	-0,841	7	0,005	138	-0,135	92
	5	0,032	21	-0,098	1	-0,411	3	0,096	80
	6	-0,055	129	0,073	150	-0,328	6	-0,705	6
	Razem:		233		233		173		231

Tabela 2

Grupy obiektów branży informatycznej po optymalizacji podziału  
metodą k-średnich

Podział na k grup		Model dla kwartału							
		I		II		III		IV	
k:		Średnia, zestandaryzowana stopa zwrotu w grupie	N <sub>i</sub>	Średnia, zestandaryzowana stopa zwrotu w grupie	N <sub>i</sub>	Średnia, zestandaryzowana stopa zwrotu w grupie	N <sub>i</sub>	Średnia, zestandaryzowana stopa zwrotu w grupie	N <sub>i</sub>
4	1	-0,458	6	0,162	4	-0,149	6	0,145	28
	2	-0,192	30	-0,016	31	0,147	14	-0,052	35
	3	-0,264	6	-0,041	31	-0,133	23	1,132	3
	4	0,241	36	0,09	12	0,126	15	-0,471	12
	Razem:			78		78		58	
5	1	-0,359	21	-0,067	18	0,147	14	-0,164	2
	2	<b>0,254</b>	<b>27</b>	0,121	9	-0,149	6	0,084	32
	3	-0,458	6	-0,146	27	0,148	14	-0,929	7
	4	0,147	2	<b>0,157</b>	<b>23</b>	-0,208	1	0,013	34
	5	0,144	22	0,446	1	-0,132	23	1,13	3
	Razem:			78		78		58	
6	1	-0,359	21	0,539	5	0,052	13	-0,016	2
	2	0,097	5	0,446	1	0,405	3	-0,327	2
	3	0,114	18	-0,073	14	-0,105	12	0,427	2
	4	0,180	18	-0,095	24	-0,165	12	-0,929	7
	5	0,105	14	-0,021	23	<b>0,187</b>	<b>12</b>	0,0129	34
	6	0,147	2	0,061	11	-0,149	6	<b>0,191</b>	<b>31</b>
	Razem:			78		78		58	

Średnia zestandaryzowanych stóp zwrotu wyższa od zera wskazuje na grupy obiektów, które uzyskiwały w poszczególnych kwartałach stopy zwrotu wyższe niż portfel odniesienia. W przypadku pierwszego wariantu badania portfel odniesienia składał się z akcji wszystkich badanych spółek należących do trzech analizowanych branż w poszczególnych kwartałach od 2008 do 2011 roku. W przypadku drugiego wariantu, portfel odniesienia składał się ze wszystkich akcji spółek danej branży uwzględnionych w badaniu. Należy podkreślić, że dodatnia wartość średniej zestandaryzowanych stóp zwrotu nie musi odpowiadać zyskowi z portfela. Stopy ponadprzeciętne w niektórych kwartałach mogą oznaczać tylko ograniczanie strat w stosunku do portfela odniesienia.



Wysokie i dodatnie wartości średnich zestandaryzowanych stóp zwrotu wskazują na grupy, których obiekty średnio biorąc uzyskiwały stopy zwrotu wyższe niż średnia stopa w odpowiednim kwartale. Wyróżnione grupy stanowią zatem populacje w modelach dyskryminacyjnych, do których klasyfikowane są obiekty przejawiające zdolności do uzyskiwania ponadprzeciętnych stóp zwrotu. Warto przypomnieć, że w przypadku modeli dyskryminacyjnych z liczbą populacji większą od 2, do klasyfikowania obiektów wykorzystuje się funkcje klasyfikacyjne Fishera. W przypadku modelu dla pierwszego kwartału branży informatycznej (tabela 2), wyznaczonych zostało zatem 5 funkcji klasyfikacyjnych, z których każda odpowiada jednej z 5 grup obiektów, spośród których znajduje się grupa obiektów uznanych za atrakcyjne (grupa 2). Obiekt nieznanego pochodzenia zostanie zaklasyfikowany do tej grupy jeżeli z wyliczonych dla niego wartości wszystkich pięciu funkcji klasyfikacyjnych, największą okaże się wartość funkcji nr 2. Klasyfikacja obiektu do którejkolwiek z pozostałych wskazuje na brak istotnej siły fundamentalnej spółki w świetle przyjętych założeń.

#### 4. Konstrukcja i weryfikacja modeli dyskryminacyjnych

Na podstawie ustanowionego podziału obiektów próby uczącej skonstruowane zostały modele dyskryminacyjne. W przypadku pierwszego wariantu badania były to cztery modele (odpowiadające kolejnym kwartałom) dla wszystkich spółek bez względu na branżę z której pochodziły. Drugi wariant zakładał rozróżnienie spółek ze względu na branżę. Konstruowanych było w tym przypadku dwanaście modeli dyskryminacyjnych odpowiadających trzem branżom, po cztery dla każdego kwartału. W tabeli 3 przedstawiono współczynniki funkcji klasyfikacyjnych Fishera odpowiadających populacjom spółek atrakcyjnych w modelach dla branży informatycznej. Należy podkreślić, że w tabeli znajdują się tylko współczynniki funkcji klasyfikujących obiekty, od których oczekuje się ponadprzeciętnych stóp zwrotu w trzymiesięcznym okresie po publikacji sprawozdań.

Tabela 3

Współczynniki funkcji klasyfikacyjnych Fishera odpowiadających populacjom spółek atrakcyjnych w modelach dla branży informatycznej

Zmienna w modelu	Model dla kwartału			
	I	II	III	IV
Płynność bieżąca	-0,375	0,575	0,523	-0,406
Wskaźnik zadłużenia kapitału własnego	1,680	-2,984	0,215	-0,931
Udział kapitału własnego w aktywach trwałych	-0,839	-0,448	-1,412	-2,119
Rentowność aktywów ogółem	11,787	1,107	0,036	0,146
Wskaźnik rotacji aktywów ogółem	0,967	-1,606	-2,224	-0,496
Wskaźnik źródeł finansowania inwestycji	-2,246	0,847	0,637	-0,316
Stała w modelu	-7,992	-2,223	-2,381	-2,400

Współczynniki funkcji odpowiadające zmiennym w modelach informują o kierunku i sile wpływu poszczególnych cech diagnostycznych na przynależność obiektów do grupy spółek, od których oczekuje się ponadprzeciętnych stóp zwrotu. Pamiętając o przeprowadzonej na wstępie standaryzacji wartości wskaźników finansowych możliwa jest interpretacja współczynników funkcji klasyfikacyjnych w modelach dla poszczególnych kwartałów. Analizując tabelę 3 można stwierdzić, że spółki branży informatycznej uzyskujące w pierwszym kwartale roku możliwie wysoką i ponadprzeciętną wartość wskaźnika rentowność aktywów (wartość zestandaryzowana większa od zera) zwiększają swoją szansę zaklasyfikowania do grupy obiektów atrakcyjnych. Ponadto, szansa ta wzrastała również dla spółek, które finansowały swoją działalność długiem w większym stopniu niż średni poziom na rynku w okresie. W pierwszym kwartale na wzrost wartości funkcji klasyfikacyjnej odpowiadającej populacji obiektów leżących w polu zainteresowania inwestorów ma również wpływ możliwie niski i nieprzekraczający średniego poziomu w okresie wskaźnik źródeł finansowania inwestycji. Warto również zwrócić uwagę na to, że w poszczególnych kwartałach różne wskaźniki finansowe mogą w różny sposób oraz z różnym natężeniem wpływać na przynależność obiektu do grupy obiektów atrakcyjnych inwestycyjnie. Uzasadnione wydaje się zatem podejście, w którym ze względu na występowanie czynników sezonowych oddziałujących na działalność przedsiębiorstw skonstruowane zostały odrębne modele dyskryminacyjne dla kwartałów.

Zastosowanie metody iteracyjnej optymalizacji podziału zbioru obiektów spowodowało utworzenie grup obiektów wyraźnie różniących się między sobą ze względu na zestaw opisujących je cech diagnostycznych. Z tego powodu odsetek poprawnie zaklasyfikowanych obiektów próby uczącej w modelach dla kwartałów był bardzo wysoki, a globalne współczynniki klasyfikacji w modelach oscylowały w granicach 95%. Analiza dyskryminacyjna nie jest zatem bezpośrednio odpowiedzialna za wysoki poziom poprawności klasyfikacyjnej. Stanowi ona jednak ważne narzędzie podczas procesu klasyfikacji oraz pozwala wyciągnąć wnioski wynikające z interpretacji współczynników funkcji klasyfikacyjnych. Bardzo wysoki odsetek poprawnie zaklasyfikowanych obiektów próby uczącej zarówno w przypadku pierwszego, jak i drugiego wariantu badania nie gwarantuje sukcesu podczas testowania modeli na danych próby testującej. Modele zweryfikowane zostaną na obiektach próby testującej podczas symulowania strategii inwestycyjnej w ostatniej części badania.

## **5. Symulacja strategii inwestycyjnej na podstawie opracowanych modeli**

Ostatnim etapem przeprowadzonego badania jest symulacja strategii inwestycyjnych skonstruowanych na podstawie modeli dyskryminacyjnych zbudowanych dla dwóch wariantów badania. Strategie zakładają zakup do portfela inwestycyjnego akcji spółek typowanych przez modele dyskryminacyjne w poszczególnych kwartałach. W przypadku pierwszego wariantu badania, w każdym kwartale stosuje się jeden, wspólny dla wszystkich spółek model dyskryminacyjny. Wskazane przez niego spółki w kwartale stanowią portfel inwestora. W drugim wariantcie, w każdym kwartale spółki typowane są za pomocą trzech modeli – dla każdej branży z osobna. Wskazane przez modele spółki znajdują się w portfelu przez trzy miesiące do czasu publikacji sprawozdań finansowych za kolejny kwartał. Wartościowy udział akcji w portfelu jest równy dla wszystkich spółek, a stopa zwrotu z portfela jest równa średniej stóp zwrotu wszystkich walorów znajdujących się w portfelu w danym kwartale. Symulacja przeprowadzona została na danych próby uczącej oraz testującej (wyróżnione okresy 2011 roku oraz 2012).

Tabela 4

Wyniki strategii inwestycyjnej opartej na modelach dyskryminacyjnych bez rozróżnienia branż, z których pochodzą spółki

Rok	Portfel	Kwartał				Średnia
		I	II	III	IV	
2008	Strategia	-2,8%	0,6%	-21,0%	2,7%	<b>-5,1%</b>
	Odniesienia	-11,8%	-6,4%	-14,6%	-0,2%	-8,2%
2009	Strategia	20,8%	13,8%	10,0%	4,4%	<b>12,2%</b>
	Odniesienia	9,9%	10,3%	1,0%	7,9%	7,3%
2010	Strategia	-4,1%	6,9%	2,2%	1,3%	<b>1,6%</b>
	Odniesienia	-4,8%	6,1%	0,6%	2,9%	1,2%
2011	Strategia	-13,1%	-25,0%	-16,9%	-7,6%	<b>-15,7%</b>
	Odniesienia	-9,4%	-22,8%	-8,6%	2,2%	-9,6%
2012	Strategia	-14,7%	22,2%			<b>3,7%</b>
	Odniesienia	-19,9%	11,4%			-4,3%
<b>Średnia</b>	Strategia	<b>-2,8%</b>	<b>3,7%</b>	<b>-6,4%</b>	<b>0,2%</b>	<b>-0,7%</b> (-2,7%)
	Odniesienia	-7,2%	-0,3%	-5,4%	3,2%	<b>-1,3%</b> (-2,4%)

Tabela 5

Wyniki strategii inwestycyjnej opartej na modelach dyskryminacyjnych z rozróżnieniem branż, z których pochodzą spółki

Rok	Portfel	Kwartał				Średnia
		I	II	III	IV	
2008	Strategia	-9,7%	-2,0%	-13,3%	7,4%	<b>-4,4%</b>
	Odniesienia	-11,8%	-6,4%	-14,6%	-0,2%	-8,2%
2009	Strategia	9,8%	11,9%	0,9%	10,9%	<b>8,4%</b>
	Odniesienia	9,9%	10,3%	1,0%	7,9%	7,3%
2010	Strategia	-5,7%	13,2%	9,4%	1,9%	<b>4,7%</b>
	Odniesienia	-4,8%	6,1%	0,6%	2,9%	1,2%
2011	Strategia	-4,6%	-17,6%	-8,4%	17,9%	<b>-3,2%</b>
	Odniesienia	-9,4%	-22,8%	-8,6%	2,2%	-9,6%
2012	Strategia	-13,9%	12,9%			<b>-0,5%</b>
	Odniesienia	-19,9%	11,4%			-4,3%
<b>Średnia</b>	Strategia	<b>-4,8%</b>	<b>3,7%</b>	<b>-2,9%</b>	<b>9,5%</b>	<b>1,0%</b> (-2,7%)
	Odniesienia	-7,2%	-0,3%	-5,4%	3,2%	<b>1,4%</b> (-2,4%)

W zdecydowanej większości okresów stopy zwrotu uzyskiwane przez inwestora stosującego strategię były wyższe niż stopy z portfela odniesienia, w którego skład wchodziło 61 spółek. Należy jednak zauważyć, że tylko w przypadku drugiego wariantu badania udało się średnio biorąc uzyskiwać wyższą stopę we wszystkich czterech kwartałach. Również symulacja przeprowadzona na spółkach próby testującej wskazuje, że w przypadku odrębnych modeli dla poszczególnych kwartałów, w każdym z testowanych okresów portfel strategii uzyskał stopę zwrotu wyższą niż odpowiadający portfel odniesienia. W prawym narożniku tabel 4 oraz 5 znajdują się wartości średnich stóp zwrotu uzyskanych przez portfele strategii oraz portfele odniesienia (wartości w nawiasach) dla wszystkich badanych lat oraz kwartałów. W obu wariantach badania średnie stopy zwrotu w kolejnych latach, jak i kwartałach były wyższe w przypadku portfeli strategii. Porównując wyniki uzyskane w obu wariantach należy stwierdzić, że wyższe średnie stopy zwrotu odpowiadały podejściu, w którym konstruowano oddzielne modele dla branż.

## Podsumowanie

Wyniki przeprowadzonego badania wskazują na możliwość stosowania analizy dyskryminacyjnej jako narzędzia analizy fundamentalnej wspomagającej proces podejmowania decyzji inwestycyjnych. Wskazują na to przede wszystkim wyniki uzyskanego w przypadku drugiego z dwóch przyjętych wariantów badania, w którym zbiór spółek atrakcyjnych stanowiły spółki typowane przez odrębne modele dyskryminacyjne skonstruowane dla poszczególnych branż. Przeprowadzona dla tego wariantu symulacja strategii inwestycyjnej wskazała na możliwość uzyskiwania wyższych stóp zwrotu przez portfel strategii opartej na modelach dyskryminacyjnych niż portfel odniesienia. Konstruowane odrębne modele dyskryminacyjne dla przedsiębiorstw prowadzących podobny rodzaj działalności gospodarczej umożliwiło skuteczniejsze typowanie spółek uzyskujących ponadprzeciętne stopy zwrotu w okresie po publikacji kwartalnych sprawozdań finansowych. Należy pamiętać, że wyższe stopy zwrotu nie musiały zawsze oznaczać zysku. W niektórych kwartałach mogły wskazywać jedynie na możliwość ograniczania strat.

Należy podkreślić istotne znaczenie zastosowania metody iteracyjnej optymalizacji podziału zbioru obiektów, która wykorzystana była na etapie definiowania populacji, z których mogły pochodzić obiekty badania. O skuteczności strategii inwestycyjnej skonstruowanej dla drugiego wariantu badania świadczą

przede wszystkim wyniki testów przeprowadzonych na danych, z których nie korzystano podczas oszacowywania parametrów modeli. Z tego powodu, przed wykorzystaniem w praktyce strategii inwestycyjnej wskazane jest jej dokładne przetestowanie na odpowiednio dużej liczbie obiektów próby testującej.

## Literatura

- Elektroniczny podręcznik pakietu STATISTICA. Analiza dyskryminacyjna. [www.statsoft.pl](http://www.statsoft.pl)
- Grabiński T.: Metody aksonometrii. Akademia Ekonomiczna, Kraków 1992.
- Hadasik D.: Upadłość przedsiębiorstw w Polsce i metody jej prognozowania. Akademia Ekonomiczna, Poznań 1998.
- McLachlan G.J.: Discriminant Analysis and Statistical Pattern Recognition. John Wiley & Sons, Inc., Hoboken, New Jersey 2004.
- Jerzemska M.: Analiza finansowa w zarządzaniu przedsiębiorstwem. Tom 1. Red. L. Bednarski i T. Waśniewski. FRR w Polsce, Warszawa 1996.
- Nowak E.: Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych. Państwowe Wydawnictwo Ekonomiczne, Warszawa 1990.
- Sierpińska M., Jachna T.: Ocena przedsiębiorstwa według standardów światowych. Wydawnictwo Naukowe PWN, Warszawa 2004.
- Ustawa o rachunkowości z dnia 29 kwietnia 1994 r. Dz.U. 2009, nr 152, poz. 1223.
- Walesiak M.: Metody analizy danych marketingowych. Wydawnictwo Naukowe PWN, Warszawa 1996.

## STOCK SELECTION USING DISCRIMINANT ANALYSIS AND CLUSTER ANALYSIS

### Summary

This article consists primarily of discriminant analysis, which could be used as a tool to guide the investors to choose appropriate shares. Author used data on the financial situation of various companies from three various branches. The objects in the examination were described through set of financial indicators. The purpose of the discriminative models was to select the companies that had a higher return rate than average in a three month period after publishing quarterly financial reports. The use of the k-means method amends quality of the discriminant model improving the chance to figure out an effective prognostic tool.