

Krzysztof Gajowniczek, Tomasz Ząbkowski, Michał Gostkowski

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

ZASTOSOWANIE DRZEW KLASYFIKACYJNYCH DO BADANIA KONDYCJI FINANSOWEJ PRZEDSIĘBIORSTW SEKTORA ROLNO-SPOŻYWCZEGO

*APPLICATION OF THE CLASSIFICATION TREES FOR THE REASERCH
OF FINANCIAL STANDING OF THE ENTERPRISES IN AGRI-FOOD SECTOR*

Słowa kluczowe: drzewa klasyfikacyjne, kondycja finansowa przedsiębiorstw

Key words: classification trees, financial standing of the enterprises

Abstrakt. Celem pracy była empiryczna weryfikacja przydatności wykorzystania modeli klasyfikacyjnych do oceny kondycji finansowej przedsiębiorstw sektora rolno-spożywczego oraz na ich podstawie wyznaczenie determinant poziomu kondycji finansowej badanych przedsiębiorstw. Badania oparto na wynikach ankiet przeprowadzonych w latach 2009-2011 w 501 przedsiębiorstwach. Sytuacja finansowa jest jedną z najistotniejszych kwestii dla zarządzania przedsiębiorstwem. Problemy finansowe przedsiębiorstwa mogą przyczynić się do niewypłacalności, dlatego analiza kondycji finansowej jest niezbędna nie tylko do bieżącego zarządzania, ale także do uchronienia przed negatywnymi skutkami spadku koniunktury. Dostrzeżone odpowiednio wcześniej sygnały ostrzegawcze i podjęte na ich podstawie decyzje mogą zapobiec upadłości przedsiębiorstwa.

Wstęp

Dynamiczny rozwój gospodarczy przedsiębiorstw wiąże się ze wzrostem stopnia niepewności i ryzykiem popełnienia błędów przy podejmowaniu decyzji na szczeblu zarządczym. Dlatego dobór odpowiednich instrumentów finansowych, umożliwi monitorowanie działalności gospodarczej z punktu widzenia efektywnego zarządzania finansami w zakresie inwestowania w aktywa przedsiębiorstwa.

Ogólnosiwiatowy kryzys gospodarczy pokazał jak istotne jest ciągle monitorowanie sytuacji finansowej przedsiębiorstwa. Tylko dzięki wczesnemu wykryciu nieprawidłowości, firma może podjąć działania zapobiegawcze i uchronić się przed upadłością. Niezwykle istotne, zatem staje się wykorzystanie sprawdzonych narzędzi analitycznych, pozwalających przewidzieć zagrożenia i w efekcie odpowiednio wcześniej podjąć działania naprawcze.

Ocena kondycji finansowej przedsiębiorstwa stanowi istotny instrument zarządzania przedsiębiorstwem. Jest ona wykorzystywana zarówno jako niezależny instrument wspomagający zarządzanie, jak i element systemów rachunkowości zarządczej lub controllingu. Najczęściej, do oceny wykorzystuje się analizę sprawozdań finansowych i analizę wskaźnikową [Antonowicz 2007]. Jednak coraz częściej menadżerowie sięgają po inne metody będące w stanie z wyprzedzeniem zidentyfikować symptomy pogarszającej się sytuacji przedsiębiorstwa, dlatego w ostatnim czasie wzrosło znaczenie zastosowania predykcyjnych modeli statystycznych [Hamrol i in. 2004, Chrzanowska i in. 2009].

Zastosowanie predykcyjnych modeli statystycznych do oceny sytuacji finansowej przedsiębiorstwa ma oczywiście swoje ograniczenia. Jedną z głównych wad zastosowania modeli statystycznych jest brak uwzględnienia specyficznych i wyjątkowych oraz niewymiernych finansowo wielkości, takich jak: szanse rozwojowe, stan koniunktury, pozycja przedsiębiorstwa na rynku lub nastroje wśród pracowników. Jednak ze względu na łatwość zastosowania, niewielką pracochłonność, a przede wszystkim stosunkowo dużą wiarygodność, modele statystyczne są szczególnie użyteczne, gdy celem badań jest podział badanych przedsiębiorstw na stabilne finansowo oraz takie, które znajdują się w trudnej kondycji. Wyodrębnione taką drogą przedsiębiorstwa

znajdujące się w korzystnej sytuacji finansowej mogą stanowić wzorzec w procesie promowania sprawdzonych rozwiązań.

Ocena kondycji finansowej przedsiębiorstwa powinna zatem przyczyniać się do optymalizacji zysków i nadwyżki finansowej. Powinna ona służyć zarządzaniu, zarówno bieżącemu, jak i strategicznemu, a także dać jednostce obraz, co o jej sytuacji finansowej mogą myśleć kontrahenci oraz wierzyciele. Systematyczna i skrupulatna analiza pozwoli przedsiębiorstwu wykryć ewentualne zagrożenia w kontynuacji działalności w celu ich zniwelowania w początkowym stadium.

Celem badań była empiryczna weryfikacja przydatności wykorzystania drzew klasyfikacyjnych do oceny kondycji finansowej przedsiębiorstw sektora rolno-spożywczego oraz na ich podstawie wyznaczenie determinant poziomu kondycji finansowej badanych przedsiębiorstw.

Material i metodyka badań

Badania oparto na wynikach ankiet przeprowadzonych w latach 2009-2011 wśród przedsiębiorstw z sektora rolno-spożywczego. Ankiety zostały rozesłane do 10 tys. losowo wybranych przedsiębiorstw przetwórstwa mięsa, mleka, zbóż oraz przetwórstwa owoców i warzyw, odpowiedzi uzyskano z 507 przedsiębiorstw. Do dalszej analizy wykorzystano 501 ankiet niezawierających braków w danych w zestawie odpowiedzi. Ankieta wystosowana do przedsiębiorstw zawierała 58 pytań o zróżnicowanej formie, w której występowały głównie pytania zamknięte. Wśród pytań można wyznaczyć następujące obszary tematyczne: (1) ogólne informacje o przedsiębiorstwie, (2) organizacja i zarządzanie logistyką, (3) zapasy, (4) magazynowanie i magazyny, (5) opakowania i logistyka zwrotna, (6) transport, (7) zarządzanie informacją i informatyka.

Badane przedsiębiorstwa zajmowały się produkcją wyrobów piekarskich i mącznych (42%), przetwórstwem mięsa (23%), pozostałych artykułów spożywczych (9%), wytwarzaniem produktów przemiału zbóż i skrobi (8%), przetwórstwem owoców i warzyw (6%), wytwarzaniem wyrobów mleczarskich (5%), produkcją gotowych pasz dla zwierząt (3%), napojów (3%), olejów i tłuszczów (1%). W badanej próbie struktura przedsiębiorstw według liczby zatrudnionych osób przedstawiała się następująco: mikro (12%), małe (62%), średnie (20%) oraz duże (6%).

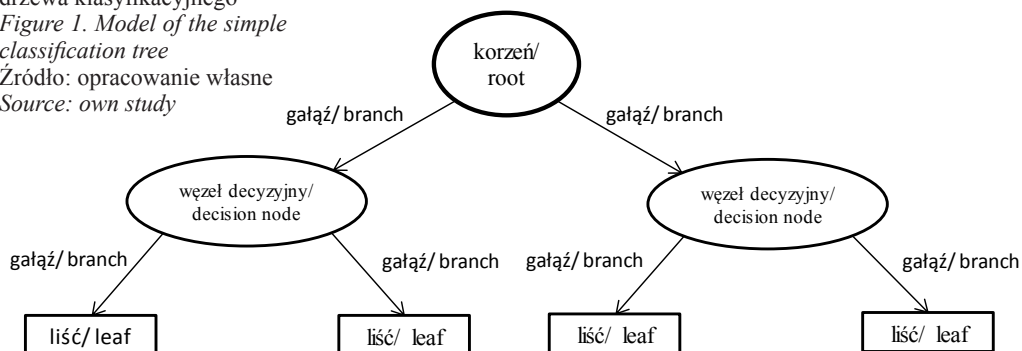
W przeprowadzonych badaniach wykorzystano predykcyjne modele statystyczne oparte o drzewa klasyfikacyjne, z uwagi na możliwość przedstawienia otrzymanych wyników w formie reguł decyzyjnych oraz wyboru zmiennych niezależnych najbardziej wpływających na poziom kondycji finansowej przedsiębiorstw.

Konstrukcja binarnych drzew klasyfikacyjnych polega na sekwencyjnym dzieleniu podzbiorów przestrzeni próby X , na dwa rozłączne i dopełniające się podzbiory. W każdym kroku podział dokonywany jest tak, aby uzyskane podzbiory były możliwie jednorodne. Opisany powyżej proces można przedstawić graficznie za pomocą grafu spójnego bez cykli, czyli drzewa [Breiman i in. 1984].

Rysunek 1. Model prostego drzewa klasyfikacyjnego

Figure 1. Model of the simple classification tree

Źródło: opracowanie własne
Source: own study



Niech T oznacza drzewo klasyfikacyjne, natomiast węzły decyzyjne (wierzchołki) utożsamia się z podzbiorami przestrzeni X i oznacza się literą t . Węzły odpowiadające podzbiорom końcowym nazywa się liśćmi, a zbiór wszystkich liści drzewa T oznacza się przez \tilde{T} . Przykładową sytuację przedstawiono na rysunku 1, ilustrującym drzewo o 3 węzłach decyzyjnych oraz 4 liściach.

Cechą charakterystyczną drzewa klasyfikacyjnego są związane z nim dwie funkcje. Określona na zbiorze liści \tilde{T} funkcja I przyporządkowująca każdemu z nich etykietę ze zbioru klas G oraz określona dla każdego węzła decyzyjnego funkcja podziału s . Ta ostatnia każdemu elementowi

węzła t przyporządkowuje jedną z dwóch wartości – prawdę t_t lub fałsz t_f oraz prawdopodobień-

stwa przynależności do nowych węzłów $\hat{p}_t = \frac{n(t_t)}{n(t)}$, $\hat{p}_f = \frac{n(t_f)}{n(t)}$. Wynika z tego, że każdy

model zbudowany na podstawie drzewa klasyfikacyjnego T ma postać: $\hat{d}_T = \sum_{t \in \tilde{T}} I(t)I(x \in T)$,

gdzie I jest funkcją wskaźnikową. Ponadto, dla dowolnego obiektu (x_i, y_i) , $i = 1, \dots, N$ określa

się prawdopodobieństwa $p(t) = P(x \in t)$, $p(t/j) = P(x \in t/y = j)$, gdzie ze wzoru na prawdopo-

dobieństwo całkowite $p(t) = \sum_{j=1}^J \pi_j p(t/j)$ wynika, że π_j jest prawdopodobieństwem *a priori* tego, że obiekt należy do j -tej klasy.

Ważnym zagadnieniem jest wybór odpowiedniej funkcji podziału s i związanej z nią miary niejednorodności $i(t)$ elementów w węzle. Miarę niejednorodności w węzle t można zdefiniować jako $i(t) = Q(p(1/t), \dots, p(J/t))$, gdzie w omawianym badaniu postacią funkcji Q jest indeks Giniego: $Q(p_j) = 1 - \sum_{j=1}^J p_j^2$.

Celem modelu jest przewidzenie klasy wyróżnionej (pozytywnej), w związku z czym poprawne decyzje to: prawidłowe wskazanie klasy wyróżnionej (ang. *True Positive* – TP) oraz klasy niewyróżnionej (negatywnej) (ang. *True Negative* – TN). Błąd zostanie popełniony w przypadku, gdy model błędnie wskaże wyróżnioną klasę (ang. *False Positive* – FP) lub nie, gdy powinien ją wskazać (ang. *False Negative* – FN). Wyniki działania reguły sumarycznie można przedstawić za pomocą następującej macierzy klasyfikacji [Stehman 1997] (tab. 1). Suma liczb w pierwszym wierszu wskazuje prawdziwą liczbę przypadków pozytywnych (tab. 1), natomiast suma liczb w drugim wierszu – przypadków negatywnych. Pierwsza kolumna wskazuje liczbę zaklasyfikowanych przypadków przez regułę do kategorii obserwacji pozytywnych, zaś druga kolumna – liczbę zaklasyfikowanych obserwacji do kategorii negatywnej. Najlepszy model to taki, który maksymalizuje całkowitą liczbę poprawnych klasyfikacji:

$$\frac{TP + TN}{TP + FP + TN + FN} \rightarrow \max$$

Tabela 1. Macierz klasyfikacji
Table 1. Classification matrix

		Wartość przewidywana/ <i>Predicted value</i>	
		pozytywna/ <i>positive</i> (P)	negatywna/ <i>negative</i> (N)
Wartość rzeczywista/ <i>Real value</i>	pozytywna/ <i>positive</i> (P)	prawdziwe pozytywne/ <i>true positive</i> (TP)	fałszywe negatywne/ <i>false negative</i> (FN)
	negatywna/ <i>negative</i> (N)	fałszywe pozytywne/ <i>false positive</i> (FP)	prawdziwe negatywne/ <i>true negative</i> (TN)

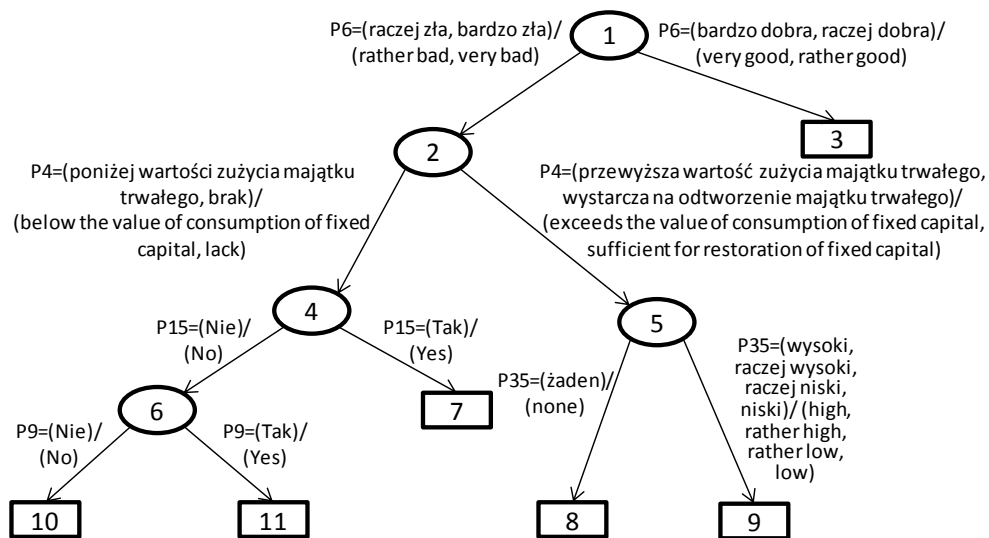
Źródło: opracowanie własne
Source: own study

W przeprowadzonym badaniu na etapie podziału oryginalnego zbioru danych nie użyto standardowego podejścia, które wydziela zbiór uczący oraz zbiór testowy, natomiast zaproponowane zostało podejście wykorzystujące 10-krotny sprawdzian krzyżowy. Generowany jest zbiór uczący zawierający 90% wszystkich obserwacji oraz zbiór testowy zawierający pozostałe obserwacje. Na tak podzielonym zbiorze jest budowany model, którego jakość jest następnie sprawdzana na zbiorze testowym. Następnie oryginalny zbiór kolejny raz jest dzielony w takich samych proporcjach na zbiór uczący oraz testowy, z tym, że za każdym razem przez losowanie do obydwu zbiorów wchodzi różne obserwacje. Cała procedura jest powtarzana 10 razy, po czym w celu obliczenia ostatecznego błędu rozpoznawania, uzyskane wyniki cząstkowe uśredniają się.

Wyniki badań

Opierając się na wynikach badań, spośród 57 potencjalnych cech objaśniających algorytm drzew klasyfikacyjnych za najbardziej istotne uznano następujące zmienne: P6 – subiektywna ocena pozycji rynkowej przedsiębiorstwa przypisana przez osobę wypełniającą ankietę, P4 – subiektywna ocena skali procesów inwestycyjnych zachodzących w firmie, P35 – stopień standaryzacji wykorzystywanych opakowań, P15 – planowanie inwestycji w powierzchnie magazynowe, P9 – posiadanie odrębnego działu transportowego. Konstruując drzewo klasyfikacyjne (rys. 2) zostały wygenerowane następujące reguły decyzyjne w wynikowych liściach:

- liść nr 3: jeżeli P6 jest „bardzo dobra” lub „raczej dobra” to z prawdopodobieństwem 95% kondycja przedsiębiorstwa jest dobra,
- liść nr 7: jeżeli P6 jest „raczej zła” lub „bardzo zła” oraz P4 „jest poniżej wartości zużycia majątku trwałego” lub „brak” oraz P15 jest „tak” to z prawdopodobieństwem 80% kondycja przedsiębiorstwa jest dobra,
- liść nr 8: jeżeli P6 jest „raczej zła” lub „bardzo zła” oraz P4 „przewyższa wartość zużycia majątku trwałego” lub „wystarcza na odtworzenie majątku trwałego” oraz P35 jest „żaden” to z prawdopodobieństwem 63% kondycja przedsiębiorstwa jest zła,



Rysunek 2. Struktura finalnego drzewa klasyfikacyjnego

Figure 2. Structure of the final classification tree

Źródło: opracowanie własne

Source: own study

- liść nr 9: jeżeli P6 jest „raczej zła” lub „bardzo zła” oraz P4 „przewyższa wartość zużycia majątku trwałego” lub „wystarcza na odtworzenie majątku trwałego” oraz P35 jest „wysoki”, „raczej wysoki”, „raczej niski”, „niski” to z prawdopodobieństwem 80% kondycja przedsiębiorstwa jest dobra,
- liść nr 10: jeżeli P6 jest „raczej zła” lub „bardzo zła” oraz P4 „jest poniżej wartości zużycia majątku trwałego” lub „brak” oraz P15 jest „nie” oraz P9 jest „nie” to z prawdopodobieństwem 74% kondycja przedsiębiorstwa jest zła.
- liść nr 11: jeżeli P6 jest „raczej zła” lub „bardzo zła” oraz P4 „jest poniżej wartości zużycia majątku trwałego” lub „brak” oraz P15 jest „nie” oraz P9 jest „tak” to z prawdopodobieństwem 61% kondycja przedsiębiorstwa jest dobra.

Na podstawie wyżej zdefiniowanych reguł decyzyjnych, dokonano klasyfikacji wszystkich przedsiębiorstw na wcześniej zdefiniowane dwie grupy.

Analizując wartości umieszczone w tabeli 2 można zauważyć, że klasyfikacja przedsiębiorstw na grupy „dobra” oraz „zła” jest w 88,6% trafna ((388+56)/501). Na tej podstawie można twierdzić, że zbudowane drzewo klasyfikacyjne w dobrym stopniu odwzorowuje zależności znajdujące się w zbiorze danych. Niepokój jedynie może budzić fakt, że w 21 przypadkach model niepoprawnie zaklasyfikował podmioty jako przedsiębiorstwa mające dobrą kondycję finansową, pomimo że takiej one nie miały.

Podsumowanie

Wielość oraz różnorodność wskaźników stosowanych w tradycyjnej analizie ekonomiczno-finansowej oraz trudności w ich interpretacji skłaniają badaczy do poszukiwania metod umożliwiających w sposób jednoznaczny ocenienie sytuacji finansowej przedsiębiorstwa. Modele klasyfikacyjne spełniają to kryterium, co powoduje, że mają szerokie zastosowanie w praktyce. Charakteryzuje je obiektywizm, wysoka skuteczność predykcji oraz prostota wynikająca z ograniczenia analizy do kilku najważniejszych cech. Otrzymany model może stanowić istotne uzupełnienie narzędzi wykorzystywanych przez menadżerów do monitorowania sytuacji przedsiębiorstwa. Badanie sprawozdania finansowego, czy analiza wskaźnikowa w połączeniu ze stworzeniem predykcyjnego modelu statystycznego pozwala wspomagać procesy decyzyjne, np. w sferze działań naprawczych, czy może też stanowić narzędzie do oceny stabilności sytuacji finansowej przedsiębiorstwa.

Zbudowane modele pozwalają nie tylko dokonać oceny przyszłego zagrożenia upadłością, ale także służą analizie porównawczej sytuacji finansowej różnych podmiotów. Stanowią one narzędzie wczesnego ostrzeżenia o powstawaniu nowych zagrożeń lub narastaniu istniejących trudnościach finansowych, dlatego warto je stosować w analizie kondycji finansowej podmiotów szczególnie w okresach osłabienia koniunktury gospodarczej.

Należy nadmienić, że uzyskane za pomocą modeli klasyfikacyjnych wyniki nie powinny być traktowane jako rozstrzygające. Odpowiednio stosowane mogą w sposób istotny wspomagać proces oceny rzeczywistej sytuacji finansowej przedsiębiorstw.

Podjęty temat stanowi początek badań nad budową wiarygodnego systemu wczesnego ostrzeżenia przed zagrożeniem bankrutem w polskich przedsiębiorstwach sektora rolno-spożywczego. Realizacja tak nakreślonego problemu badawczego pozwoli dokonać pogłębionej diagnozy przedsiębiorstw przemysłu rolno-spożywczego, ocenić ich stabilność finansową oraz wskazać kierunki potencjalnych działań poprawiających ich funkcjonowanie.

Tabela 2. Wyniki klasyfikacji
Table 2. Classification results

		Wartość przewidywana/Predicted value	
		dobra/ <i>good</i>	zła/ <i>bad</i>
Wartość rzeczywista/ <i>Real value</i>	dobra/ <i>good</i>	388	36
	zła/ <i>bad</i>	21	56

Źródło: opracowanie własne
Source: own study

Literatura

- Antonowicz P. 2007: *Metody oceny i prognoza kondycji ekonomiczno-finansowej przedsiębiorstw*, Wyd. Ośrodek Doradztwa i Doskonalenia Kadr, Gdańsk.
- Breiman L., Friedman J., Olshen, Stone R. 1984: *Classification and Regression Tress*. Chapman & Hall.
- Chrzanowska M., Alfaro E., Witkowska D. 2009: *The individual borrowers recognition: Single and ensemble trees*, Expert Systems with Applications, 36, 6409-6414.
- Hamrol M., Czajka B., Piechocki M. 2004: *Upadłość przedsiębiorstwa – model analizy dyskryminacyjnej*, Przegląd Organizacji, 6.
- Stehman S.V. 1997: *Selecting and interpreting measures of thematic classification accuracy*. Remote Sensing of Environment, 62(1), 77-89.

Summary

Financial situation is one of the most important issue for the managers. Financial problems of the enterprises could threaten insolvency that is why the analysis of financial standing is essential not only for current management, but also for protection before side effects of economic downturn. Early enough recognition of warning signals and good decisions can prevent the liquidation of the companies. The aim of this article is the empirical verification of the suitability of the use of predictive classification models to assess the financial condition of companies in the agri-food sector and based on their; the appointment of the determinants affecting the level of the financial condition of the researched companies.

Adres do korespondencji
mgr Krzysztof Gajowniczek, dr inż. Tomasz Ząbkowski, mgr Michał Gostkowski
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Wydział Zastosowań Informatyki i Matematyki
ul. Nowoursynowska 159, 02-776 Warszawa
e-mail: tomasz_zabkowski@sggw.pl, tel. (22) 59 37 290
e-mail: krzysztof_gajowniczek@sggw.pl, tel. 506 746 850
e-mail: michal_gostkowski@sggw.pl, tel. 889 632 764