



ROZMYTE METODY KLASYFIKACJI W ANALIZIE SEGMENTÓW RYNKOWYCH NA PRZYKŁADZIE RYNKU MOTORYZACYJNEGO

*Bartłomiej Jefmański, Katedra Ekonometrii i Informatyki,
Uniwersytet Ekonomiczny we Wrocławiu*

Wprowadzenie

Europejski rynek samochodów osobowych jest coraz mocniej zróżnicowany pod względem takich parametrów, jak: wymiary samochodów, ich wyposażenie standardowe, a co za tym idzie cena, która jest w dalszym ciągu jednym z najważniejszych kryteriów wyboru nowego samochodu. Różnorodność oferty producentów samochodów osobowych wymusiła stosowanie podziału tego rynku na segmenty. Jak się jednak okazuje, w przypadku niektórych modeli samochodów może pojawić się problem z jednoznacznym przyporządkowaniem do któregoś z segmentów. Rozwiązaniem może być zastosowanie formalnych metod statystycznych umożliwiających obiektywne przyporządkowanie samochodu do segmentu rynkowego.

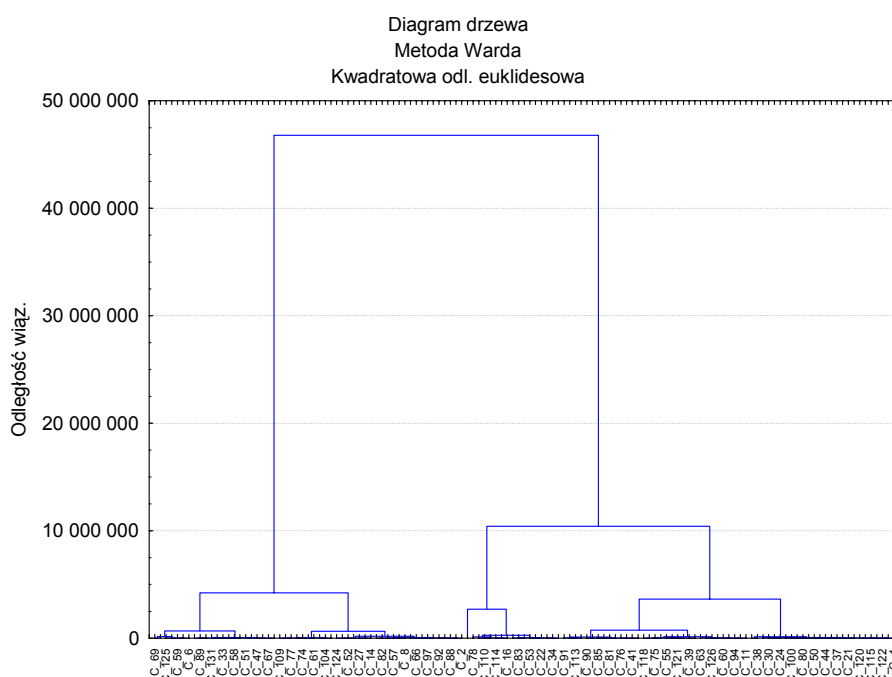
Metody klasyfikacji w segmentacji rynku

Klasyczne podejście w zagadnieniu segmentacji rynku opiera się głównie na wykorzystaniu metod klasyfikacji. Ze względu na różnorodność tych metod w literaturze przedmiotu wprowadza się różne ich podziały. Jeden z nich zakłada wyodrębnienie trzech grup:

- ◆ metody hierarchiczne (aglomeracyjne i deglomeracyjne),
- ◆ metody podziału (obszarowe, gęstościowe i optymalizujące wstępny podział zbioru obiektów),
- ◆ metody prezentacji graficznej (skalowanie wielowymiarowe, analiza korespondencji).

Szczegółowy opis wymienionych metod zawarty został w wielu polskich i zagranicznych publikacjach traktujących o metodach klasyfikacji stosowanych w badaniach marketingowych, w tym badaniach segmentacyjnych. Zwróćmy uwagę jedynie na pierwszą grupę metod klasyfikacji, która wydaje się szczególnie użyteczna w procesie wyodrębniania segmentów rynkowych. Różni się ona od pozostałych m.in. tym, że wynikiem zastosowania metod tej grupy jest drzewo zwane dendrogramem. Taka graficzna forma prezentacji wyników klasyfikacji jest ważnym atutem tej grupy metod, ponieważ pozwala określić wzajemne usytuowanie poszczególnych klas oraz obiektów w nich zawartych. Przykład

dendrogramu dla rynku samochodów osobowych otrzymanego za pomocą metody Warda, będącej jedną z metod hierarchicznych, aglomeracyjnych zaprezentowano na rys. 1.



Rys. 1. Dendrogram otrzymany za pomocą metody Warda; źródło: opracowanie własne z wykorzystaniem modułu analizy skupień programu *STATISTICA*.

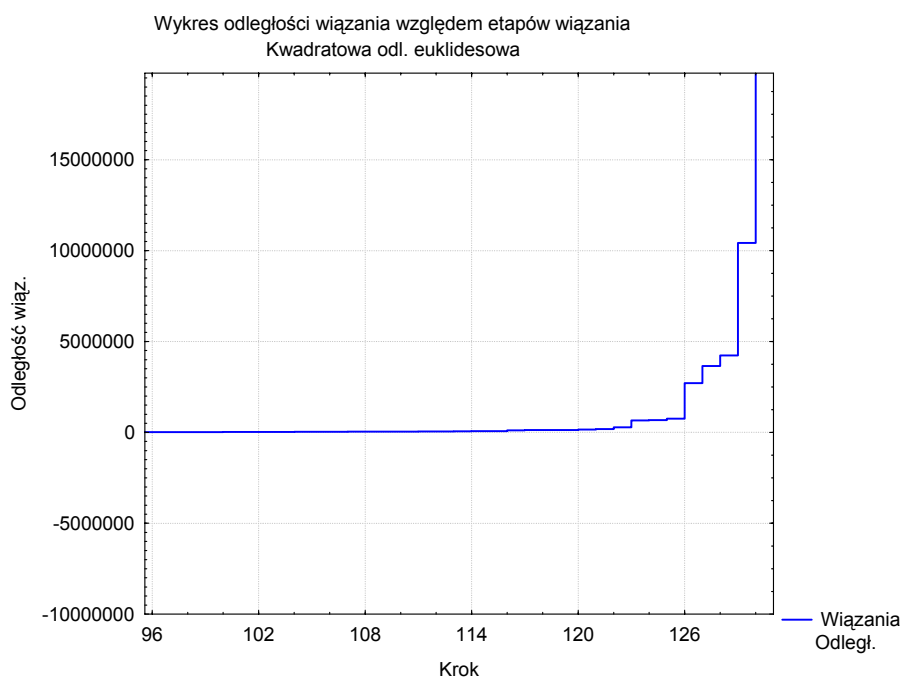
W pierwszym kroku analizy każdy z samochodów tworzy osobne, jednoelementowe skupienie. Następnie, przesuwając się w górę dendrogramu, samochody podobne pod względem przyjętych parametrów łączone są w skupienia. Połączenie dwóch lub więcej skupień reprezentowane jest przez węzły. Analizując powyższy dendrogram można zauważyć początkowo sześć skupień, które następnie łączą się, dając trzy, a w efekcie końcowym dwa skupienia.

Niezależnie od tego, którą grupę metod klasyfikacji wykorzystamy w procesie segmentacji rynku, zawsze pojawia się problem ustalenia liczby skupień. W przypadku metod hierarchicznych do jej ustalenia można zastosować krzywą aglomeracji [zob. rys. 2.].

Liczbę skupień można ustalić również w sposób bardziej formalny za pomocą odpowiednich metod statystycznych [zob. np. Gatnar, Walesiak 2004, s. 338-341]. Warto podkreślić, że w praktyce badań marketingowych często stosowane jest podejście dwuetapowe. Polega ono na tym, że w pierwszym etapie za pomocą metod hierarchicznych ustala się wstępną liczbę skupień, a następnie w drugim etapie, za pomocą np. metody k -średnich, dokonuje się właściwej klasyfikacji obiektów.

Powyżej została zaprezentowana jedna z wielu metod klasyfikacji, którą można zastosować w procedurze segmentacji rynku. Wszystkie one mają jednak cechę wspólną, która w niektórych przypadkach może okazać się ich słabością. Chodzi bowiem o jednoznaczne przyporządkowanie obiektów do wyodrębnionych skupień. W przypadku rynku samochodowego pozycjonowanie modelu samochodu w określonym segmencie nie zawsze w pełni

oddaje faktyczny stan rzeczy. Nie istnieją bowiem w tym przypadku sztywne granice segmentów. Poszczególne segmenty raczej się zazębiają, powodując, że pewne modele znajdują się na pograniczu, a ich jednoznaczne przyporządkowanie jest mocno utrudnione.



Rys. 2. Krzywa aglomeracji; źródło: opracowanie własne z wykorzystaniem modułu analizy skupień programu *STATISTICA 8*.

Zastosowanie rozmytych metod klasyfikacji pozwala odejść od orzekania, czy obiekt należy do określonej klasy czy nie. Umożliwia określenie stopnia przynależności obiektu do różnych klas. Orzekamy nie o tym, czy obiekt należy czy też nie do danej klasy, ale określamy jego stopień przynależności do różnych klas. Takie podejście w większym stopniu odzwierciedla rzeczywistość oraz może uchronić badacza przed utratą pewnych informacji w stosunku do podejścia, w którym obiekty przyporządkowuje się tylko do jednej klasy.

Podstawy teoretyczne rozmytej metody segmentacji rynku

Do klasycznych rozmytych metod klasyfikacji zalicza się m.in. takie metody jak: rozmyta metoda k -średnich (*fuzzy c-means – FCM*), algorytm Gustafsona-Kessela, algorytm Gatha-Geva. Na potrzeby niniejszego opracowania zastosowano rozmytą metodę k -średnich.

FCM to metoda iteracyjna, której idea jest bardzo zbliżona do klasycznej metody k -średnich. Jej stosowanie nie wymaga czynienia założeń co do charakteru materiału empirycznego poddanego analizie. Celem metody jest znalezienie takich środków ciężkości klas, które minimalizują funkcję:

$$J_m = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d_{ij}^2$$



gdzie:

μ_{ij} – stopień przynależności j -tego obiektu do i -tej klasy rozmytej,

d_{ij} – odległość euklidesowa między środkiem ciężkości i -tej klasy rozmytej a j -tym obiektem,

m – parametr rozmycia (*fuzzification parametr*), przy czym $m > 1$.

Algorytm rozmytej metody k -średnich zbudowany jest z następujących kroków:

Krok 1. Losowe zainicjowanie macierzy przynależności $\mathbf{U} = [\mu_{ij}]$ przy czym: $\sum_{i=1}^c \mu_{ij}$,
 $\forall_j = 1, 2, \dots, n$.

Krok 2. Obliczenie środków ciężkości klas zgodnie ze wzorem:

$$c_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m},$$

przy czym:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}}}.$$

Krok 3. Obliczenie nowej macierzy przynależności \mathbf{U}_{nowe} . Jeżeli $\|\mathbf{U}_{nowe} - \mathbf{U}\| > \varepsilon$, gdzie $\|\mathbf{U}_{nowe} - \mathbf{U}\|$ to odległość euklidesowa, a ε to przyjęty próg zbieżności, wówczas należy przyjąć $\mathbf{U} = \mathbf{U}_{nowe}$ i przejść do kroku 2. Postępowanie kończy się w sytuacji, gdy $\|\mathbf{U}_{nowe} - \mathbf{U}\| < \varepsilon$ lub zostanie osiągnięta zadana liczba iteracji k .

Przed rozpoczęciem obliczeń należy jednak określić stosowaną miarę odległości, liczbę klas oraz wartość parametru rozmycia m .

Parametr m określa stopień rozmycia wyników klasyfikacji. Wartość parametru winna być $m > 1$, przy czym wartości bliskie jedności skutkować będą otrzymaniem wyników zbliżonych do tych, które zostałyby uzyskane przy zastosowaniu metod klasycznych. Wzrost wartości parametru m powoduje, że stopnie przynależności obiektów do poszczególnych klas będą przyjmowały wartości zbliżone do odwrotności liczby klas.



Wyniki rozmytej segmentacji rynku

W analizie uwzględniono 131 modeli samochodów będących w sprzedaży na rynku polskim według stanu na dzień 01.05.2008 r. Każdy z modeli scharakteryzowano za pomocą trzech zmiennych: długość, szerokość i rozstaw osi. Są to trzy podstawowe parametry stosowane przy pozycjonowaniu samochodów na rynku europejskim. Dane techniczne samochodów zaczerpnięto z dodatku do miesięcznika „Auto Moto” 2008, nr 3 (142). W przypadku modeli samochodów o kilku generacjach w analizie uwzględniono tylko te najnowsze (np. trzecia generacja Opla Astry). Ponadto w analizie nie uwzględniono wersji kombi poszczególnych modeli samochodów oraz samochodów typu: mini van (np. Ford C-Max), van (np. Renault Espace), kombi van (np. Mazda 5), crossover (np. Nissan Qashquai) oraz samochodów klasy SUV (np. Toyota RAV-4).

Tabela 1. Najbardziej „rozmyte” obiekty w poszczególnych klasach – podział na sześć klas.

	Stopnie przynależności do klas					
	μ_{1j}	μ_{2j}	μ_{3j}	μ_{4j}	μ_{5j}	μ_{6j}
Klasa pierwsza (C – „klasa niższa średnia”)						
Fiat Bravo	0,8729543	0,1269617	0,0000001	0,0000095	0,0000005	0,0000738
Ford Focus	0,6249366	0,3744294	0,0000002	0,0000355	0,0000043	0,0005939
Klasa druga (D – „klasa średnia”)						
BMW 3	0,0015951	0,6602226	0,0000001	0,0000048	0,0001554	0,3380221
Peugeot 307	0,2739943	0,7258655	0,0000000	0,0000075	0,0000007	0,0001320
Klasa trzecia (A – „mini”)						
Opel Agila	0,0009486	0,0000840	0,8741533	0,1247950	0,0000027	0,0000164
Citroën C2	0,0001224	0,0000093	0,8885162	0,1113506	0,0000002	0,0000013
Klasa czwarta (B – „małe”)						
Peugeot 207	0,2330792	0,0003562	0,0000322	0,7665183	0,0000007	0,0000134
Suzuki Swift	0,0002426	0,0000140	0,2021452	0,7975961	0,0000002	0,0000019
Klasa piąta (F – „luksusowe”)						
Lexus GS	0,0000059	0,0001975	0,0000000	0,0000002	0,8685791	0,1312174
Klasa szósta (E – „klasa wyższa średnia”)						
Brak samochodów o stopniu przynależności do szóstej klasy mniejszym niż $\mu_6 = 0,9$						

Źródło: obliczenia własne z wykorzystaniem programu R.



Rozmytą metodę k-średnich zastosowano dla odległości euklidesowej oraz parametru rozmycia $m = 1,25$. Zbadano stopnie przynależności samochodów do sześciu klas². Samochody, których przynależność do wyodrębnionych klas była najbardziej „rozmyta” ($\mu_{ij} < 0,9$), wyszczególniono w tabeli 1. W celu wskazania samochodu charakterystycznego dla każdej z wyodrębnionych klas obliczono środki ciężkości poszczególnych klas, a następnie porównano je z parametrami samochodów należących do tych klas. Wyniki zestawiono w tabeli 2.

Tabela 2. Środki klas – podział na sześć klas.

Klasa	Długość	Szerokość	Rozstaw osi	„Reprezentant” klasy
1	4889,107	1844,327	2840,169	Toyota Auris
2	4671,614	1791,455	2716,143	Mitsubischi Lancer
3	5072,924	1894,329	3033,632	Kia Picanto
4	4378,052	1761,457	2609,563	Ford Fiesta
5	4075,061	1721,302	2537,675	Jaguar XF
6	3608,381	1630,814	2385,456	Jaguar X-Type

Źródło: obliczenia własne z wykorzystaniem programu R.

Porównując wyniki zestawione w tabelach 1 i 2 z panującym na rynku europejskim podziałem na sześć segmentów, można stwierdzić, że poszczególne klasy odpowiadają następującym segmentom: klasa pierwsza – segment C, klasa druga – segment D, klasa trzecia – segment A, klasa czwarta – segment B, klasa piąta – segment F, klasa szósta – segment E.

W większości segmentów można wskazać modele samochodów, których przynależność nie jest jednoznaczna. Wyjątkiem jest segment E, gdzie wartość stopnia przynależności dla żadnego modelu nie ukształtowała się poniżej przyjętego progu równego 0,9.

W segmencie aut najmniejszych dwa modele: Opel Agila i Citroën C2 mają stopień przynależności mniejszy niż 0,9. Jednak stopnie ich przynależności do segmentu B są na tyle małe, że można uznać, że segment A jest segmentem jednorodnym i nie ma w nim modeli mogących przysporzyć problemów z właściwym zaklasyfikowaniem ich do segmentu.

W segmencie B sytuacja jest już nieco bardziej złożona. Osoby zainteresowane kupnem samochodu z segmentu C powinny zainteresować się modelem Peugeota 207. Jego zaklasyfikowanie do segmentu B nie jest jednoznaczne, a wartość stopnia przynależności do segmentu C podpowiada, że może to być interesująca propozycja zakupu. Na podstawowy model tego samochodu należy przeznaczyć 40 800 zł. Najtańszy samochód z segmentu C kosztuje 40 490 zł (Fiat Linea). Ceny są zatem porównywalne, co czyni z oferty zakupu Peugeota 207 atrakcyjną propozycję.

W przypadku Suzuki Swift sytuacja jest odwrotna. Jego wysoki stopień przynależności do segmentu aut najmniejszych powoduje, że cena 39 900 zł za podstawowy model tego

² W analizie przyjęto podział na sześć klas wzorem stosowanego na europejskim rynku samochodów osobowych podziału na sześć segmentów: A – „mini”, B – „małe”, C – „klasa niższa średnia”, D – „klasa średnia”, E – „klasa wyższa średnia”, F – „luksusowe”.



samochodu jest znacznie wyższa niż ceny większości aut z segmentu A. Zakup tego samochodu należało by zatem poprzedzić dokładną analizą i porównaniem standardowego wyposażenia tego pojazdu z podstawowym wyposażeniem modeli pozycjonowanych w segmencie aut najmniejszych.

W segmencie C należy zwrócić szczególną uwagę na model Forda Focusa. Rozmiary tego samochodu pozwalają mu aspirować do segmentu D, czyli do klasy średniej. Cena najtańszego modelu z segmentu D wynosi 60 500 zł. Porównując ją z ceną najtańszego Forda Focusa, która wynosi 52 600 zł, można stwierdzić, że stanowi on atrakcyjną ofertę na polskim rynku samochodów osobowych.

Również w segmencie D znalazły się modele, których przynależność do segmentu nie jest jednoznaczna. Wartości stopni przynależności do segmentu D poniżej poziomu 0,9 zaobserwowano dla dwóch modeli samochodów: BMW 3 oraz Peugeot 307. Model BMW 3 posiada dość dużą wartość stopnia przynależności do segmentu E. Ceny najtańszego modelu z segmentu E, którym jest Volvo S60, zaczynają się od kwoty 111 900 zł. Cena najtańszej wersji BMW 3 kształtuje się na podobnym poziomie, bo 113 000 zł. Zatem BMW 3 to jeden z najdroższych modeli w segmencie D (wyprzedza go jedynie Audi A4, którego ceny zaczynają się od 118 590 zł) i jeden z najtańszych w segmencie E (tańsze jest jedynie Volvo S60, za które trzeba zapłacić 1 100 zł mniej).

Ciekawą propozycją zakupu może stanowić również Peugeot 307. Model pozycjonowany jest w segmencie C, podczas gdy wyniki analizy wskazują, że został on zaklasyfikowany do segmentu D. Brak jednoznacznej przynależności tego modelu do segmentu C lub D, co znajduje odzwierciedlenie w jego wartościach stopni przynależności, może budzić zainteresowanie wśród potencjalnych nabywców. Zauważmy, że najtańsza wersja Peugeota 307 kosztuje 69 000 zł. W segmencie C jest to jeden z droższych modeli samochodów. Natomiast w segmencie D, do którego został zaklasyfikowany na podstawie zastosowanej w artykule metody, byłby to jeden z najtańszych modeli (jedynie Citroen C5 jest tańszy o 8 500 zł).

W segmencie F niewielkie wątpliwości budzi jedynie Lexus GS. Został on zaklasyfikowany do segmentu aut największych, ale posiada również niewielki stopień przynależności do klasy szóstej. W prasie motoryzacyjnej model ten pozycjonowany jest w segmencie E. Jednak wyniki analizy pokazują, że model ten powinien być pozycjonowany w najwyższej klasie. Wówczas będzie on najtańszym modelem w segmencie F, co czyni z niego atrakcyjną cenowo ofertę zakupu (podstawowa cena modelu w wysokości 203 200 zł jest niższa o 88 900 zł od ceny najtańszego modelu segmentu F – Audi A8).

Analizując uzyskane wyniki, należy pamiętać, że w badaniu uwzględniono jedynie trzy parametry charakteryzujące modele samochodów osobowych. Bardziej szczegółowa analiza wymagałaby uwzględnienia również wyposażenia standardowego oferowanego dla poszczególnych modeli samochodów, okresu gwarancji itp. Nie bez znaczenia pozostają również oferty promocyjne stosowane przez dealerów głównie w postaci upustów cenowych na wybrane modele i wyposażenie. Jednak wyniki zaprezentowane w artykule pozwoliły wskazać modele samochodów, którym konsument powinien się dokładniej przyjrzeć, podejmując decyzję o zakupie samochodu. Dla wskazanych w artykule modeli samochodów



można przeprowadzić pogłębioną analizę porównawczą, uwzględniając wyszczególnione powyżej dodatkowe zmienne decydujące o atrakcyjności danego modelu samochodu.

Literatura

1. Cox E.: *Fuzzy modeling and genetic algorithms for data mining and exploration*. San Francisco: Morgan Kaufmann Publishers 2005, s. 230-238.
2. Gatnar E., Walesiak M. (red.): *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo AE we Wrocławiu 2004, s. 322-341.
3. Höppner F., Klawonn F., Kruse R., Runkler T.: *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Chichester: Wiley & Sons 1999, s. 37-49.
4. Lasek M.: *Data Mining. Zastosowania w analizach i ocenach klientów bankowych*. Warszawa: Oficyna Wydawnicza „Zarządzanie i finanse” 2002, s. 146.
5. Nascimento S., Mirkin B., Moura-Pires F.: *A fuzzy clustering model of data and fuzzy c-means*. W: Materiały konferencyjne IEEE International Conference on Fuzzy Systems: Soft Computing in the Information Age, Piscataway 2000, Vol. 1, s. 302-303.