

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

**RESEARCH PAPERS**

of Wrocław University of Economics

**236**

# **Badania marketingowe – metody, nowe podejścia i konteksty badawcze**



pod redakcją

**Krystyny Mazurek-Łopacińskiej**

**Magdaleny Sobocińskiej**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2011

Recenzenci: Anna Dąbrowska, Lechosław Garbarski, Józef Garczarczyk

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna na stronie [www.ibuk.pl](http://www.ibuk.pl)

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl> oraz w The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com) a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon [http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się na stronie internetowej Wydawnictwa [www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Kopiowanie i powielanie w jakiegokolwiek formie wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2011

**ISSN 1899-3192**

**ISBN 978-83-7695-248-2**

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

Wstęp.....	9
------------	---

---

### Część 1. Rozwój badań marketingowych – kierunki, koncepcje, wymiary

---

<b>Krystyna Mazurek-Łopacińska, Magdalena Sobocińska:</b> Rozwój badań marketingowych – w kierunku nowych podejść i kontekstów badawczych związanych z funkcjonowaniem przedsiębiorstwa.....	13
<b>Dariusz Oczachowski:</b> Kultura kognitywna organizacji a prowadzenie badań marketingowych .....	23
<b>Tomasz Heryszek:</b> Myślenie kognitywne czy afektywne? Między danymi twardymi a przeczuciem – dylematy współczesnego przedsiębiorcy .....	33
<b>Jan W. Wiktor:</b> Koncepcja i sposób pomiaru internacjonalizacji przedsiębiorstwa w świetle doświadczeń projektu „Strategie marketingowe przedsiębiorstw na rynkach międzynarodowych” .....	42
<b>Aleksandra Nizielska:</b> Dystans kulturowy w procesie internacjonalizacji przedsiębiorstw – metodyczny aspekt badań.....	51
<b>Magdalena Soboń, Iga Rudawska, Sylwia Bąkowska:</b> Metodyka badania zachowań konsumenckich w międzynarodowej sieci badawczej.....	61
<b>Bogusław Bemberek:</b> Rola wywiadu gospodarczego w zarządzaniu wiedzą w klastrze .....	71

---

### Część 2. Nowe podejścia badawcze wynikające z rozwoju technologii informacyjnych i komunikacyjnych

---

<b>Mariusz Kuziak:</b> Wyzwania wobec badań użytkowników Internetu.....	83
<b>Karol Łopaciński:</b> Model tematycznej hurtowni danych na potrzeby badania przebiegu i efektów kampanii e-mailingowych.....	95
<b>Magdalena Jaciow:</b> Fora internetowe jako źródło informacji – możliwości i ograniczenia wykorzystania w badaniach zachowań nabywczych .....	113
<b>Beata Kolny:</b> Wykorzystanie netnografii do badania usług zagospodarowujących czas wolny .....	123
<b>Radosław Szulc, Piotr Ciszewski:</b> Wykorzystanie technologii łączności bezprzewodowej w badaniach marketingowych na rynku produktów <i>beauty care</i> .....	132

---

### Część 3. Metody ilościowe – nowe podejścia i zastosowania oraz triangulacja metod

---

<b>Adam Sagan:</b> Asymetryczne metody wielowymiarowe w badaniach marketingowych .....	145
<b>Mariusz Łapczyński:</b> Łączenie metod i narzędzi w budowie modeli predykcyjnych.....	155
<b>Józef Garczarczyk, Robert Skikiewicz:</b> Zastosowanie metody grupowania dwustopniowego w segmentacji klientów indywidualnych na rynku usług bankowych.....	164
<b>Anna Bryja:</b> Zmienne jakościowe w segmentacji rynku: miary powiązań a wyniki grupowania.....	175
<b>Bartłomiej Jefmański:</b> Nowe podejście w pomiarze opinii respondentów z zastosowaniem skal porządkowych i elementów teorii zbiorów rozmytych – charakterystyka wybranych aspektów metodologicznych.....	184
<b>Grzegorz Maciejewski:</b> Wykorzystanie analizy czynnikowej w badaniach konsumenckiego ryzyka .....	192
<b>Paweł Chlipała:</b> Zastosowanie eksperymentu w badaniach społecznie odpowiedzialnej konsumpcji – refleksje nad metodą, wyniki .....	203

---

### Część 4. Badania jakościowe i ich wykorzystanie w rozwiązywaniu problemów badawczych i decyzyjnych

---

<b>Marcin Komor:</b> Znaczenie i rozwój metod jakościowych w badaniach empirycznych w marketingu.....	215
<b>Sylwia Wrona:</b> Dobór próby w jakościowych badaniach marketingowych – problemy prawidłowej selekcji i rekrutacji uczestników .....	225
<b>Mateusz Rak, Joanna Nogiec:</b> Wykorzystanie wyników badań jakościowych do identyfikacji populacji w badaniach ilościowych.....	234
<b>Iwona Olejnik:</b> Metoda obserwacji – zastosowania w badaniach marketingowych.....	242
<b>Zbigniew Piskorz:</b> Rozpoznawanie okazji przedsiębiorczych – rezultaty badań jakościowych.....	250
<b>Zbigniew Spyra:</b> Zastosowanie podejścia etnograficznego we współczesnych badaniach marketingowych w sferze kultury .....	260
<b>Wanda Patrzalek:</b> Przebieg i zaburzenia procesów wymiany informacji oraz komunikacji wewnątrz gospodarstwa domowego w ogólnopolskich badaniach fokusowych .....	271

<b>Joanna Wardzała-Kordyś:</b> Wiedza i opinie o procesach reklamacji produktów wadliwych w kontekście badań gospodarstw domowych .....	280
<b>Agnieszka Dejnaka:</b> Komunikacja pomiędzy członkami gospodarstwa domowego przy użyciu nowoczesnych narzędzi wymiany informacji .....	290
<b>Jolanta Tkaczyk:</b> Rola opowieści w badaniach marketingowych .....	301

## Summaries

---

### Part 1. Development of marketing research – trends, concepts, dimensions

---

<b>Krystyna Mazurek-Łopacińska, Magdalena Sobocińska:</b> Development of marketing research – towards new approaches and contexts of research related to the functioning of a company .....	22
<b>Dariusz Oczachowski:</b> Cognitive culture of an organization and conducting marketing research .....	32
<b>Tomasz Heryszek:</b> Cognitive or affective thinking? Between hard data and foreboding – modern business dilemmas .....	41
<b>Jan W. Wiktor:</b> The concept and measurement method of the companies' internationalization in the light of the research project "Marketing strategies of companies on international markets" .....	50
<b>Aleksandra Nizielska:</b> Cultural distance in the process of companies' internationalization – methodological aspect of research .....	60
<b>Magdalena Soboń, Iga Rudawska, Sylwia Bąkowska:</b> Consumer behaviour research methods in the International Research Network .....	70
<b>Bogusław Bembenek:</b> The role of economic intelligence in knowledge management of cluster .....	80

---

### Part 2. New research approaches arising from the development of information and communication technologies

---

<b>Mariusz Kuziak:</b> Challenges to Internet audience measurement .....	94
<b>Karol Łopaciński:</b> Model of data mart prepared due to research regarding proceeding of e-mail campaigns and its effects .....	112
<b>Magdalena Jaciow:</b> Online forums as a source of information – possibilities and limitations of use in purchasing behaviour research .....	122
<b>Beata Kolny:</b> The application of netnographic surveys to research on leisure time services .....	131
<b>Radosław Szulc, Piotr Ciszewski:</b> The implementation of wireless technology in marketing research within the beauty industry .....	142

---

### Part 3. Quantitative methods – new approaches, applications and triangulation methods

---

<b>Adam Sagan:</b> Asymmetric multivariate methods in marketing research .....	154
<b>Mariusz Łapczyński:</b> Combining methods and tools in building predictive models.....	163
<b>Józef Garczarczyk, Robert Skikiewicz:</b> Applying a Two Step Cluster method in the segmentation of individual customers for the banking services market .....	174
<b>Anna Bryja:</b> Qualitative variables in market segmentation: Similarity coefficients and clustering results.....	183
<b>Bartłomiej Jefmański:</b> A new approach in respondents' opinion measurement using ordinal scales and elements of fuzzy sets theory – characteristics of selected methodological aspects .....	191
<b>Grzegorz Maciejewski:</b> The use of factor analysis in consumer risk research	202
<b>Paweł Chlipała:</b> Using an experiment in the research of socially responsible consumption – reflections on the method, the results.....	212

---

### Part 4. Qualitative research and its use in problems solving research and decision

---

<b>Marcin Komor:</b> The importance and development of quality methods in empirical research in marketing .....	224
<b>Sylwia Wrona:</b> Selection of a sample in qualitative marketing research – issues related to the accurate selection and recruitment of participants.....	233
<b>Mateusz Rak, Joanna Nogiec:</b> Using the results of qualitative research to the identification of the population in quantitative research.....	241
<b>Iwona Olejnik:</b> The method of observation – application in marketing research .....	249
<b>Zbigniew Piskorz:</b> Entrepreneurial opportunity recognition – results of qualitative research .....	259
<b>Zbigniew Spyra:</b> Application of ethnographic approach to contemporary marketing research in culture.....	270
<b>Wanda Patrzalek:</b> Proceeding and abnormal processes of information exchange and communication within the household in nationwide focus studies .....	279
<b>Joanna Wardzała-Kordyś:</b> Knowledge and opinions about complaints to the defective products in the context of household surveys .....	289
<b>Agnieszka Dejnaka:</b> Communication between members of the household by using modern information exchange tools.....	300
<b>Jolanta Tkaczyk:</b> The role of storytelling in marketing research .....	310

**Mariusz Łapczyński**

Uniwersytet Ekonomiczny w Krakowie

---

## ŁĄCZENIE METOD I NARZĘDZI W BUDOWIE MODELI PREDYKCYJNYCH

---

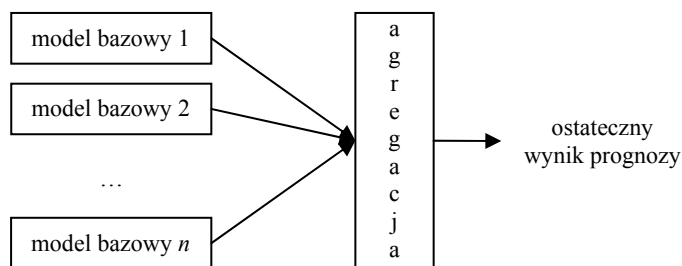
**Streszczenie:** Celem artykułu jest charakterystyka dwóch głównych strategii stosowanych w łączeniu narzędzi analitycznych podczas budowy modeli predykcyjnych. Pierwsza z nich to podejście wielomodelowe (*ensemble models*), w którym o ostatecznym wyniku analizy decydują zagregowane rozwiązania cząstkowe. Druga strategia wykorzystuje modele hybrydowe (*hybrid models*), a wynik analizy otrzymuje się przez połączenie co najmniej dwóch narzędzi analitycznych. Procedura ta jest zazwyczaj dwuetapowa, a wyniki jednej analizy są w kolejnym kroku podstawą do przeprowadzenia następnej. Studia literaturowe zawarte w artykule dotyczą złożonych modeli predykcyjnych wykorzystanych przede wszystkim w obszarze badań marketingowych.

**Słowa kluczowe:** modele hybrydowe, podejście wielomodelowe, modele predykcyjne.

### 1. Podejście wielomodelowe a modele hybrydowe

Łączenie modeli i narzędzi analitycznych jest obecnie powszechną praktyką podczas budowy modeli prognostycznych w wielu dziedzinach i obszarach badawczych. Większa czasochłonność związana z analizą danych jest bardzo często kompensowana większą trafnością predykcji, redukcją problemu niezbilansowanych prób lub dostarczeniem czytelnych wzorców ze zbiorów danych o złożonej strukturze. W literaturze przedmiotu wspomina się w tym kontekście o modelach hybrydowych (*hybrid models*) lub podejściu wielomodelowym (*ensemble models*). Przez pojęcie hybrydy należy rozumieć sekwencyjne wykorzystanie kilku modeli, z których co najmniej jeden jest modelem opisowym i co najmniej jeden jest modelem prognostycznym. Do budowy jednego z nich wykorzystuje się klasyczne narzędzia statystyczne, takie jak np. metodę  $k$ -średnich,  $k$ -medoidów, regresję logistyczną, natomiast do budowy drugiego – narzędzia do eksploracji danych, takie jak np. sieci Kohonena, drzewa klasyfikacyjne, losowy las, wzmacniane drzewa klasyfikacyjne, reguły asocjacyjne i sekwencyjne. Zdaniem autora hybrydyzacja oznacza łączenie modeli wzorcowych (z nauczycielem, nadzorowanych) z modelami bezwzorcowymi (bez nauczyciela, nienadzorowanymi) oraz łączenie klasycznych narzędzi statystycznych z narzędziami analitycznymi wywodzącymi się z obszaru *data mining*.

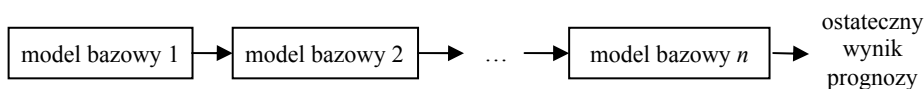
Nieco inną definicję modeli hybrydowych zaproponował E. Gatnar [2008, s. 68 i n.], traktując je jako te powstałe w wyniku połączenia architektury równoległej i szeregowej w podejściu wielomodelowym. Równoległe łączenie modeli polega, ogólnie rzecz ujmując, na budowie wielu modeli bazowych, których wyniki są agregowane w drodze głosowania lub uśredniania prawdopodobieństw przynależności obiektów do klas (rys. 1).



**Rys. 1.** Schemat architektury równoległej w podejściu wielomodelowym

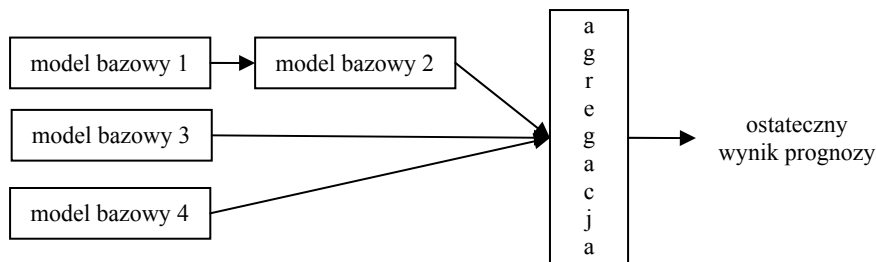
Źródło: opracowanie własne na podstawie [Gatnar 2008, s. 68].

Szeregowe łączenie modeli bazowych sprowadza się do wykorzystywania wyników jednego modelu bazowego do budowy kolejnego. Przykładem takiego podejścia jest algorytm AdaBoost oraz grupa tzw. wzmacnianych drzew klasyfikacyjnych lub regresyjnych (*gradient tree boosting*). W kolejnych modelach bazowych zwiększa się wagę przypadków błędnie sklasyfikowanych i jednocześnie zmniejsza się wagę przypadków poprawnie sklasyfikowanych. Schemat takiej architektury przedstawiono na rys. 2.



**Rys. 2.** Schemat architektury równoległej w podejściu wielomodelowym

Źródło: opracowanie własne na podstawie [Gatnar 2008, s. 69].



**Rys. 3.** Schemat architektury równoległej w podejściu wielomodelowym

Źródło: opracowanie własne na podstawie [Gatnar 2008, s. 69].



Stąd łatwo przedstawić schemat struktury hybrydowej, który mógłby wyglądać tak, jak to przedstawiono na rys. 3. Jest przy tym istotne, że w podejściu wielomodelowym mówi się najczęściej o jednej metodzie analitycznej wykorzystywanej wielokrotnie. Poza wspomnianymi wcześniej drzewami wzmacnianymi za klasyczne podejście w łączeniu modeli należy uznać losowy las [Breiman 2001].

Celem niniejszego artykułu jest krótka prezentacja wybranych złożonych modeli predykcyjnych. Ze względu na liczne przykłady pojawiające się w światowej literaturze nie jest możliwa pełna prezentacja różnorodnych podejść badawczych i analitycznych nawet wówczas, gdy wybór zostaje ograniczony analiz rynkowych.

## 2. Łączenie modeli wzorcowych

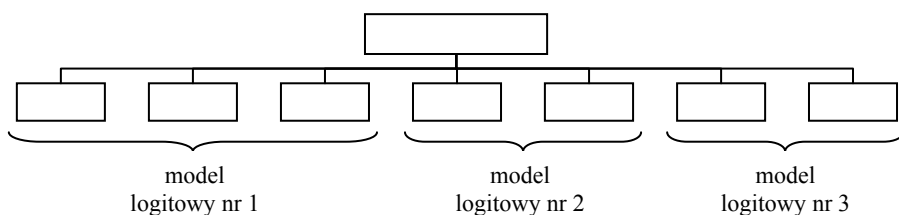
Jednym z popularniejszych sposobów łączenia narzędzi do budowy modeli wzorcowych jest sekwencyjne użycie drzew klasyfikacyjnych oraz regresji logistycznej. Drzewa klasyfikacyjne pozwalają wyodrębnić względnie jednorodne podzbiory, które są następnie wykorzystane do budowy modeli logitowych. Za jedną z pierwszych prób budowy takiego modelu można uznać połączenie drzew klasyfikacyjnych CHAID z dwumianowym modelem logitowym [Lindahl, Winship 1994]. Zmienna zależna, która pierwotnie była zmienną ciągłą (wysokość darowizny przekazanej na rzecz Northwestern University), została zdychotomizowana do dwóch kategorii: darowizny większej niż 10 tys. dolarów oraz darowizny nieprzekraczającej tej kwoty. Zbiór zmiennych niezależnych obejmował m.in.: charakterystyki absolwentów uczelni (wiek, kod geograficzny miejscowości zamieszkania, wysokość deklarowanego wynagrodzenia), informację o wysokości wpłat w poprzednich okresach oraz przynależność do trzech segmentów według metodologii PRIZM („Blue Blood Estates”<sup>1</sup>, „Urban Gold Coast”<sup>2</sup> i „Grey Power”<sup>3</sup>). W pierwszym kroku analizy zbudowano model CHAID, dzięki któremu zdyskretyzowano ciągłą zmienną niezależną „wysokość wpłat w poprzednich okresach”. Zmienna ta uczestniczyła w podziale drzewa i wydzieleniu siedmiu węzłów potomnych na pierwszym etapie podziału drzewa. W drugim kroku analizy zbudowano trzy dwumianowe modele logitowe w trzech podzbiórach absolwentów wyznaczonych za pomocą tej zmiennej. Redukcja liczby podziorów z siedmiu do trzech została dokonana subiektywnie przez badaczy. Schemat postępowania przedstawiono na rys. 4.

---

<sup>1</sup> W segmencie „Blue Blood Estates” znajdują się zamożni Amerykanie o bardzo wysokich – sięgających kilkuset tysięcy dolarów rocznie – dochodach, głównie małżeństwa z dziećmi, zamieszkujące przedmieścia, pracujący jako specjaliści i menedżerowie.

<sup>2</sup> W segmencie „Urban Gold Coast” znajdowały się głównie jednoosobowe gospodarstwa domowe zamieszkujące wysokie budynki wielorodzinne. Segment nie występuje w segmentacji PRIZM z 2011 r.

<sup>3</sup> W segmencie „Grey Power” znajdują się starsi Amerykanie, przedstawiciele klasy średniej, właściciele nieruchomości niekorzystający z domów spokojnej starości. Obecnie segment ten nazywa się „Gray Power”.



Rys. 4. Model hybrydowy CHAID-LOGIT

Źródło: opracowanie własne na podstawie [Lindahl, Winship 1994].

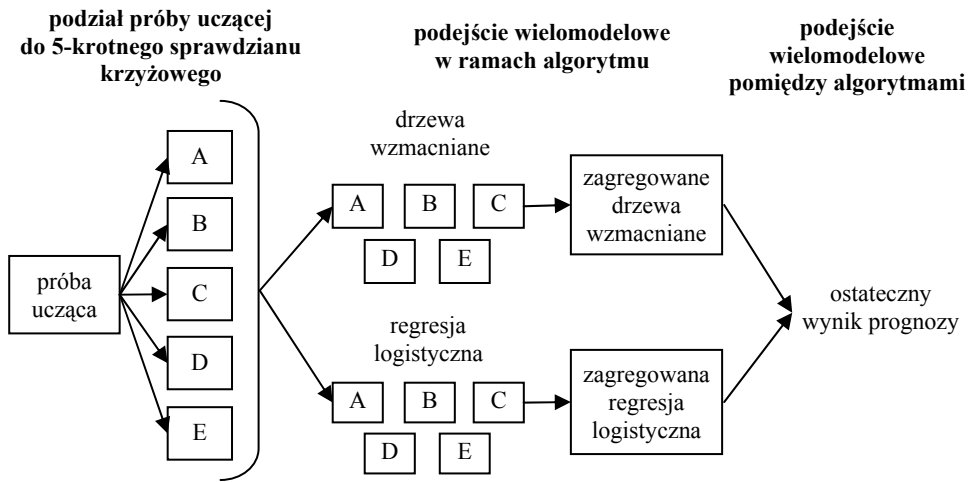
Bardziej zaawansowane hybrydy wykorzystujące algorytm CART polegały na włączaniu informacji o przynależności obiektu do liścia drzewa jako dodatkowych zmiennych niezależnych do modeli logitowych [Steinberg, Cardell 1998] lub na wykorzystaniu zmiennych konkurencyjnych w budowie hybryd CART-LOGIT [Łapczyński 2009].

Interesujący przykład modeli hybrydowych znaleźć można również w pracy M. Wei i in. [2008]. Celem badań było zbudowanie modelu prognostycznego na potrzeby sprzedaży krzyżowej. Autorzy wykorzystali tzw. podejście wielomodelowe w ramach jednego algorytmu (*within-algorithm ensemble*) oraz podejście wielomodelowe pomiędzy algorytmami (*cross-algorithm ensemble*). Próbę uczącą podzielono na 5 rozłącznych podzbiorów i zastosowano 5-krotny sprawdzian krzyżowy (walidację krzyżową). W pierwszym podejściu wykorzystano drzewa wzmacniane i regresję logistyczną, natomiast w drugim uśredniono wynik prognozy, opierając się na obu modelach. Trafność predykcji drzew wzmacnianych była nieco lepsza od trafności predykcji uzyskanej za pomocą modelu logitowego. Uśredniony współczynnik korelacji pomiędzy wynikami prognozy obu modeli dla każdego podzbioru był równy 0,73 i został zinterpretowany przez autorów jako niezbyt wysoki. W kolejnej fazie uśredniono wynik prognozy dla obu narzędzi z wszystkich podzbiorów, otrzymując trafność predykcji po 5-krotnej walidacji krzyżowej. Współczynnik korelacji pomiędzy tymi wartościami wzrósł do 0,75. Procedurę budowy modeli przedstawiono na rys. 5.

Podejście wielomodelowe może być tutaj traktowane jako hybrydowe z punktu widzenia łączenia narzędzi klasycznych (regresja logistyczna) z narzędziami *data mining* (drzewa wzmacniane).

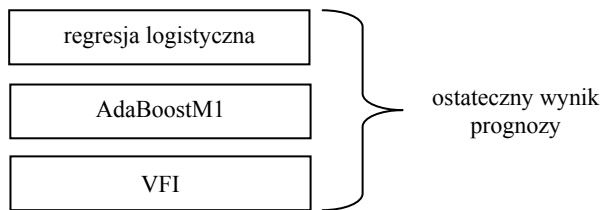
W tym samym roku zaproponowano nieco inne podejście do prognozowania sprzedaży krzyżowej [Qiu, Wang, Bi 2008]. Autorzy wykorzystali regresję logistyczną, algorytm AdaBoostM1 oraz podejście nazywane głosowaniem przedziałów klasowych zmiennych niezależnych<sup>4</sup> (*voting feature intervals*, VFI). VFI jest nieprzyrostowym (*non-incremental*) algorytmem klasyfikacyjnym, który tworzy prze-

<sup>4</sup> Algorytm VFI został po raz pierwszy opisany w 1997 r. [Demiroz, Guvenir 1997]. Jego implementację można znaleźć w programie Weka.



Rys. 5. Model hybrydowy: regresja logistyczna i wzmacniane drzewa klasyfikacyjne

Źródło: opracowanie własne na podstawie [Wei i in. 2008].



Rys. 6. Model hybrydowy: regresja logistyczna – AdaBoostM1 – VFI

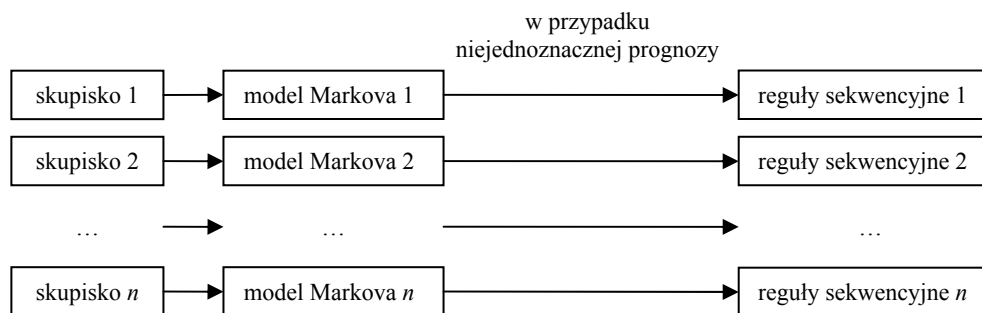
Źródło: opracowanie własne na podstawie [Qiu, Wang, Bi 2008].

działy zmiennych na podstawie zbioru obserwacji z predefiniowanymi klasami. Metodę tę uznaje się za tolerancyjną z punktu widzenia braków danych. Zmienne ciągłe są dyskretyzowane, natomiast dla zmiennych kategoryalnych ich warianty są traktowane jako odrębne „przedziały klasowe”<sup>5</sup>. Dla wszystkich przedziałów klasowych każdej zmiennej niezależnej zestawia się liczbę obiektów należących do poszczególnych kategorii zmiennej zależnej. Budowa przedziałów klasowych odbywa się poprzez minimalizację funkcji entropii, a predykcja poprzez głosowanie na podstawie próby testowej. Autorzy traktują *de facto* podejście hybrydowe jako podejście wielomodelowe, z tą różnicą, że ostatecznym wynikiem analizy jest uśrednione prawdopodobieństwo przynależności obiektu do klasy obliczone za pomocą trzech całkiem różnych narzędzi analitycznych. Schemat tego modelu hybrydowego przedstawiono na rys. 6.

<sup>5</sup> Tworzy się szereg strukturalny dla cechy jakościowej.

### 3. Łączenie modeli bezwzorcowych

Interesującą propozycję bezwzorcowego modelu hybrydowego przedstawiono w 2008 r. [Khalil, Li, Wang 2008]. W celu skutecznego prognozowania odwiedzin witryny internetowej połączono reguły sekwencyjne, analizę skupisk i model Markowa. Hybryda została nazwana zintegrowanym modelem predykcyjnym (*integrated prediction model*). Na pierwszym etapie pogrupowano sesje (przypadki odpowiadające odwiedzinom witryny) za pomocą metody  $k$ -średnich. Zbiór zmiennych odnosił się do liczby odwiedzanych stron w obrębie witryny. W drugim kroku zbudowano modele Markowa oddzielnie dla każdego skupiska. Ich celem było przewidywanie odwiedzin kolejnej strony na podstawie sekwencji stron odwiedzanych wcześniej. Szansa na bardziej trafną prognozę odwiedzin konkretnej strony jest tym większa, im większy jest zbiór elementów (stron) tę wizytę poprzedzających. Zbyt długie i tym samym zbyt szczegółowe sekwencje powodują z kolei nadmierną złożoność modelu i niewyczerpującą analizę wszystkich możliwych sekwencji, jakie mogą wystąpić w przyszłości. Autorzy zdecydowali się zatem ograniczyć liczbę elementów poprzednika w sekwencji stron do dwóch. W sytuacji gdy model Markowa nie był w stanie trafnie prognozować odwiedzin następnej strony, do analizy włączano reguły sekwencyjne, których zadaniem było znalezienie elementów poprzedzających te dwa zidentyfikowane wcześniej przez model Markowa. Przykładowy problem z interpretacją modelu Markowa mógłby przedstawiać się następująco: „jeżeli otworzył  $A$  i  $B$ , to otworzył  $C$  (prawdopodobieństwo = 0,47)” oraz „jeżeli otworzył  $A$  i  $B$  to otworzył  $D$  (prawdopodobieństwo = 0,53)”. Zbyt mała różnica między wynikami sprawia, że za pomocą reguł sekwencyjnych poszukuje się elementów poprzedzających zdarzenia  $A$  i  $B$ . Mogłoby się wówczas okazać, że nowe hybrydowe reguły sekwencyjne przybiorą następującą postać: „jeżeli otworzy  $K$  i  $A$  i  $B$ , to otworzy  $C$  (prawdopodobieństwo = 1,00)” oraz „jeżeli otworzy  $W$  i  $A$  i  $B$ , to otworzy  $D$  (prawdopodobieństwo = 1,00)”. Schemat hybrydyzacji przedstawiono na rys. 7.



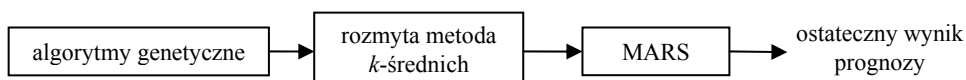
Rys. 7. Model hybrydowy  $k$ -średnich – modele Markowa – reguły sekwencyjne

Źródło: opracowanie własne na podstawie [Khalil, Li, Wang 2008].

W literaturze znaleźć można też inne sposoby łączenia reguł asocjacyjnych i sekwencyjnych z technikami grupowania (SOM,  $k$ -średnich, uogólniona analiza skupisk (EM)) [Łapczyński 2011, w druku]. Przykłady te dotyczą prognozowania sprzedaży krzyżowej na rynku usług.

#### 4. Łączenie modeli wzorcowych i bezwzorcowych

Przykładem mieszanego modelu złożonego<sup>6</sup> jest model hybrydowy zastosowany do prognozowania bankructw przedsiębiorstw na rynku finansowym [Martin i in. 2011]. Łączenie narzędzi polegało na sekwencyjnym użyciu algorytmów genetycznych, rozmytej metody  $k$ -średnich i metody MARS. Algorytmy genetyczne ułatwiły połączenie wskaźników finansowych banków w nowe zmienne komponenty. Rozmyta metoda  $k$ -średnich (*fuzzy c-means algorithm*)<sup>7</sup> pozwoliła na określenie prawdopodobieństwa przynależności obiektu do więcej niż jednego skupienia. Oznacza to, że skupiska nie są rozłączne, a niektóre obiekty mogą należeć do więcej niż jednego z nich. W tym wypadku utworzono 3 skupiska, gdzie w każdym z nich znajdowały się odpowiednio: przedsiębiorstwa, które zbankrutowały, przedsiębiorstwa, które nie zbankrutowały, oraz przedsiębiorstwa wciąż działające, jednak znajdujące się na skraju bankructwa. W każdym skupieniu z osobna wykorzystano następnie metodę MARS dla celów dyskryminacyjnych. Schemat modelu hybrydowego przedstawiono na rys. 8.



**Rys. 8.** Model hybrydowy: algorytmy genetyczne – rozmyta metoda  $k$ -średnich – MARS

Źródło: opracowanie własne na podstawie [Martin i in. 2011].

Profilowanie skupisk przy użyciu tego narzędzia pozwoliło otrzymać lepsze rozwiązanie niż w przypadku pojedynczych modeli, które zostały jednocześnie zbudowane z wykorzystaniem tego samego zbioru zmiennych niezależnych. Autorzy porównali wyniki otrzymane za pomocą hybrydy z wynikami analizy dyskryminacyjnej, regresji logistycznej, modelu probitowego i modelu hazardu (*hazard model*).

<sup>6</sup> Przez pojęcie „mieszany” należy w tym wypadku rozumieć łączenie modeli wzorcowych i bezwzorcowych, a nie tzw. *mixture models*.

<sup>7</sup> Rozmyta metoda  $k$ -średnich została po raz pierwszy opisana w 1973 r. [Dunn 1973]. Metodę zmodyfikowano osiem lat później [Bezdek 1981].

## 5. Podsumowanie

Budowa złożonych modeli predykcyjnych wydaje się obecnie standardem w prognozowaniu zjawisk rynkowych. Bez względu na to, czy predykcja dotyczy zmiennych ilościowych (modele regresyjne), czy zmiennych jakościowych (modele dyskryminacyjne), podejmuje się wiele prób łączenia modeli i narzędzi analitycznych z różnych dziedzin. Ograniczona objętość artykułu nie pozwoliła na pełną prezentację wszystkich pomysłów badaczy z ostatniego dziesięciolecia. Warto tutaj jednak nadmienić, że o modelach hybrydowych pisano już w kontekście analizy *conjoint* [Green, Krieger 1996] czy w kontekście modeli wyborów konsumenckich [Bodapati, Drolet 2005]. Do tworzenia modeli hybrydowych włączane są bardzo często algorytmy genetyczne [Brabazon, Keenan 2004; Pan 2010], używa się sekwencyjnie sieci neuronowych [Flores, Loaleza 2009] lub wprowadza kilka zmodyfikowanych na różne sposoby zbiorów zmiennych niezależnych [Bandyopadhyay, 2007]. Z obszaru analitycznego CRM na uwagę zasługują z kolei prace dotyczące sprzedaży krzyżowej czy analizy migracji klientów [Bose, Chen 2009; Kim 2006; Zou, Wang, Chang 2008].

## Literatura

- Bandyopadhyay A., *Mapping corporate drift toward default Part 2: A hybrid credit-scoring model*, „The Journal of Risk Finance” 2007, vol. 8, no. 1, s. 46–55.
- Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press, New York 1981.
- Bodapati A.V., Drolet A., *A hybrid choice model that uses actual and ordered attribute value information*, „Journal of Marketing Research” 2005, vol. 42, August, s. 256–265.
- Bose I., Chen X., *Hybrid models using unsupervised clustering for prediction of customer churn*, [w:] *Proceedings of IMECS 2009, March 18–20/2009*, Hong Kong 2009, s. 1–6.
- Brabazon A., Keenan P.B., *A hybrid genetic model for the prediction of corporate failure*, „Computational Management Science” 2004, Springer-Verlag, s. 293–310.
- Breiman L., *Random forests*, „Machine Learning” 2001, vol. 45, Kluwer Academic Publishers, s. 5–32.
- Demiroz G., Guvenir H.A., *Classification by voting feature intervals*, [w:] *Proceedings of the Ninth European Conference on Machine Learning*, Springer-Verlag, 1997, s. 85–92.
- Dunn J.C., *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*, „Journal of Cybernetics” 1973, vol. 3, s. 32–57).
- Flores J.J., Loaleza R., *Financial time series forecasting using a hybrid neural-evolutionary approach*, [w:] *Proceedings 15th SIGEF International Conference*, Lugo, Spain, 2009, s. 547–555.
- Gatnar E., *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa 2008.
- Green P.E., Krieger A.M., *Individualized hybrid models for conjoint analysis*, „Management Science” 1996, vol. 42, no. 6, s. 850–867.
- Khalil F., Li J., Wang H., *Integrating recommendation models for improved web page prediction accuracy*, [w:] *Proceedings of Thirty-First Australasian Computer Science Conference (ACSC2008)*,

- Wollongong, Australia "Conferences in Research and Practice in Information Technology (CRPIT)", 2008, vol. 74, s. 1–10.
- Kim Y., *Toward a successful CRM: Variable selection, sampling, and ensemble*, „Decision Support Systems” 2006, vol. 41, no. 2, s. 542–553.
- Lindahl W.E., Winship C., *A logit model with interactions for predicting major gift donors*, „Research in Higher Education” 1994, vol. 35, no. 6, s. 729–743.
- Łapczyński M., *Hybrydowe modele asocjacyjne w sprzedaży krzyżowej i uzupełniającej na rynku usług*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Poznaniu, Poznań 2011 (artykuł złożony do druku).
- Łapczyński M., *Modele hybrydowe CART-LOGIT w analizie danych rynkowych*, [w:] J. Dziechciarz (red.), *Projektowanie, ocena i wykorzystanie danych rynkowych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 51, UE, Wrocław 2009, s. 85–95.
- Martin A. i in., *A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means and MARS*, „International Journal on Soft Computing (IJSC )” 2011, vol. 2, no. 1/, s. 12–24.
- Pan W.-T., *Performing stock price prediction use of hybrid model*, „Chinese Management Studies” 2010, vol. 4, no. 1, s. 77–86.
- Qiu D., Wang Y., Bi B., *Identify cross-selling opportunities via hybrid classifier*, „International Journal of Data Warehousing and Mining” 2008, vol. 4, no. 2, s. 55–62.
- Steinberg D., Cardell N.S., *The hybrid CART-logit model in classification and data mining*, <http://www.salford-systems.com/resources/whitepapers/index.html>.
- Wei M., Chai L., Wei R., Huo W., *A solution to the cross-selling problem of PAKDD-2007: Ensemble model of treenet and logistic regression*, „International Journal of Data Warehousing and Mining” 2008, vol. 4, no. 2, s. 9–14.
- Zou N., Wang J., Chang G.-L., *A reliable hybrid prediction model for real-time travel time prediction with widely spaced detectors*, [w:] *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15/2008*, s. 91–96.

## COMBINING METHODS AND TOOLS IN BUILDING PREDICTIVE MODELS

**Summary:** The purpose of this article is to characterize two main strategies used in combining analytical tools during the construction of predictive models. The first is the multi-model approach (ensemble models), in which the final outcome of the analysis is achieved by the aggregation of partial solutions. The second strategy utilizes a hybrid model, in which the final result of the analysis is obtained by combining at least two different analytical tools or models. This procedure is usually two-stage, and the results of one analysis are the basis for the next step for the construction of the second model. A review of literature included in the article is related to complex predictive models used mainly in the field of marketing research.

**Keywords:** hybrid models, ensemble models, predictive models.