

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 336

Badania marketingowe – nowe podejścia oraz metody na współczesnym rynku

Redaktorzy naukowci

Krystyna Mazurek-Łopacińska

Magdalena Sobocińska



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Joanna Szynal
Redaktor techniczny: Barbara Łopusiewicz
Korektor: K. Halina Kocur
Łamanie: Małgorzata Czupryńska
Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2014

ISSN 1899-3192
ISBN 978-83-7695-476-9

Wersja pierwotna: publikacja drukowana

Druk i oprawa:
EXPOL, P. Rybiński, J. Dąbek, sp.j.
ul. Brzeska 4, 87-800 Włocławek

Spis treści

Wstęp	9
Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Badania kodów kulturowych w tworzeniu marketingowych modeli biznesu	11
Anna Olejniczuk-Merta: Rozwój innowacji społecznych a badania marketingowe	22
Monika Hajdas: Techniki pomiaru kompatybilności marek i idei kulturowych.....	31
Paweł Chlipała: Triangulacja podejść metodologicznych w badaniach naukowych z dziedziny marketingu	39
Adam Sagan: Analiza rzetelności skal w wielopoziomowych modelach pomiaru	49
Adam Sagan, Mariusz Łapczyński: Modele hybrydowe CART-logit w analizie procesu podejmowania decyzji w gospodarstwie domowym.....	60
Anna Myrda: Segmentacja łańcuchów środków-celów: miary podobieństwa sekwencji i ilościowe wskaźniki jakości grupowania a wyniki grupowania	70
Kamila Pilch: Asymetryczne skalowanie wielowymiarowe w wizerunkowych badaniach jednostek terytorialnych.....	79
Alicja Kusińska: Analizy wielowymiarowe jako źródło wiedzy o zachowaniach konsumentów na rynku	89
Krzysztof Błoński: Wykorzystanie metod wielowymiarowych do analizowania związku między emocjami a satysfakcją klienta	99
Sylwester Białowas, Iwona Olejnik: Poziom opiekuńczości państwa a zachowania oszczędnościowe – analiza wielowymiarowa	110
Jadwiga Stobiecka: Interpretacyjne konsekwencje oceny stabilności opinii respondentów w badaniach konsumpcji, oszczędzania i inwestowania.....	118
Piotr Tarka: HOMALS – wielowymiarowa analiza korespondencji jako metoda konstrukcji skali pomiarowej w badaniach marketingowych.....	129
Lukasz Skowron: Zastosowanie modelowania ścieżkowego do wyznaczenia przebiegu procesu lojalnościowego wśród klientów lubelskich centrów handlowych.....	140
Ireneusz P. Rutkowski: Metody CMMI i SGMM oceny dojrzałości procesu innowacji i wprowadzania produktu na rynek.....	152
Hanna Hall: Nowy konsument a zmiany w metodach jego badania.....	163
Tomasz Olejniczak: Techniki badawcze wykorzystywane w badaniu cyklu życia gospodarstwa domowego	174

Anna Dąbrowska, Arkadiusz Wódkowski: Kompetencje konsumentów w świetle badań ilościowych	185
Sylwia Makomaska: Wpływ muzyki tła na reakcje konsumentów w miejscu sprzedaży – problematyka interdyscyplinarności badań	195
Lucyna Witek: Metodyczne aspekty badania postaw konsumentów (na przykładzie rynku produktów ekologicznych)	205
Magdalena Olejniczak: Zróżnicowanie technik badawczych w badaniu motywacji zakupowych konsumentów żywności funkcjonalnej.....	215
Agata Dziakowicz: Metody badań marketingowych na rynku dóbr luksusowych.....	224
Wanda Patrzałek, Aleksandra Perchla-Włosik: Zastosowanie analizy semiologicznej w badaniach wpływu mody na zachowania młodych konsumentów	233
Agata Stolecka-Makowska: Zastosowanie podejścia interpretacyjnego w badaniu zmian zachowań nabywczych konsumentów podlegających akulturacji	244
Arkadiusz Wódkowski: Zmiana paradygmatu w marketingowych badaniach jakościowych?.....	257
Grzegorz Maciejewski: Zogniskowane wywiady grupowe w badaniach zachowań młodych dorosłych na rynku usług finansowych	266
Iga Rudawska: Zastosowanie zogniskowanego wywiadu grupowego do oceny jakości obsługi pacjentów przewlekle chorych	275
Ewa Nowakowska, Adam Sagan: Kontryfaktyczno-porównawcze studium przypadku w marketingu usług zdrowotnych.....	284
Krzysztof Kapera, Mariusz Kuziak: Skuteczność wybranych metod komunikacji z respondentami w badaniach internetowych	296
Iwona Escher: Niejednoznaczność statusu metodologicznego internetowego wywiadu grupowego i jego poszczególnych odmian	310
Magdalena Daszkiewicz, Sylwia Wrona: Zogniskowane wywiady grupowe online jako alternatywa dla tradycyjnych metod gromadzenia danych – szanse rozwoju i wyzwania dla badaczy	321
Olgierd Witczak: Potencjał wykorzystania serwisów społecznościowych w badaniach jakościowych	331
Agnieszka Dejnaka: Facebook jako obszar prowadzenia badań marketingowych.....	339
Robert Wolny: Możliwości wykorzystania obserwacji w Internecie w badaniach rynku e-usług.....	348

Summaries

Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Research of cultural codes in creating marketing models of business	21
---	----

Anna Olejniczuk-Merta: The development of social innovation and marketing research.....	30
Monika Hajdas: Techniques for measuring the compatibility of brands and cultural ideas.....	38
Paweł Chlipała: Triangulation of methodological approaches in scientific research of marketing field	48
Adam Sagan: Reliability analysis in multilevel measurement models	59
Adam Sagan, Mariusz Łapczyński: CART-logit hybrid models in the analysis of decision-making process in the households	69
Anna Myrda: Segmentation of Means-End Chains: sequence dissimilarity measures and quantitative cluster validity indexes vs. clustering results...	78
Kamila Pilch: Asymmetric multidimensional scaling in the research of territorial units image	88
Alicja Kusińska: Multidimensional analysis as a source of knowledge about consumer behaviour.....	98
Krzysztof Błoński: The use of multidimensional methods to analyze the relationship of emotions and customer satisfaction.....	109
Sylwester Białowas, Iwona Olejnik: The level of the state's social security and its influence on saving behaviour – multidimensional analysis.....	117
Jadwiga Stobiecka: Interpretative consequences of the assessment of respondents' opinions stability in the studies of consumption, saving and investing.....	128
Piotr Tarka: HOMALS – multiple correspondence analysis as the method for measurement scale construction in marketing research.....	139
Lukasz Skowron: The usage of the Structural Equation Modeling for determining the loyalty building process among the customers of the shopping centers located in Lublin.....	151
Ireneusz P. Rutkowski: CMMI and SGMM methods of maturity evaluation of the product innovation process and introduction of a product on the market	162
Hanna Hall: New consumer and changes in the methods of their research....	173
Tomasz Olejniczak: Research techniques used in the study of the household life cycle.....	184
Anna Dąbrowska, Arkadiusz Wódkowski: Consumer competences in the light of quantitative research	194
Sylwia Makomaska: The effects of background music on consumers response in the place of commerce – the problem of interdisciplinary research.....	204
Lucyna Witek: Methodical aspects of research of consumers' attitudes (on the example of market of green products).....	214
Magdalena Olejniczak: The diversity of marketing research techniques in the study of purchase motivation of functional food consumers.....	223

Agata Dziakowicz: Methods of marketing research for the luxury goods market	232
Wanda Patrzalek, Aleksandra Perchla-Wlosik: Studies of the impact of fashion on the behavior of young consumers using semiological analysis	243
Agata Stolecka-Makowska: The use of an interpretative approach in a study of purchasing behaviour changes of consumers acculturation	256
Arkadiusz Wódkowski: Change of paradigm in qualitative market research?	265
Grzegorz Maciejewski: The focus group interview in the research of the young-adults behaviors on the financial services market	274
Iga Rudawska: The use of focus group interview to evaluate the service quality of chronically ill patients	283
Ewa Nowakowska, Adam Sagan: Comparative-counterfactual case research in health service marketing	295
Krzysztof Kapera, Mariusz Kuziak: Effectiveness of selected methods of communication with respondents in online surveys	309
Iwona Escher: The ambiguity of the methodological status of <i>online group interview</i> and its particular types	320
Magdalena Daszkiewicz, Sylwia Wrona: Online focus group interviews as an alternative for traditional methods of data collection – opportunities for development and challenges to researchers	330
Olgierd Witczak: The potential of using social network sites in qualitative research	338
Agnieszka Dejnaka: Facebook as a marketing research area	347
Robert Wolny: The use of Internet observations in research of e-services market	357

Adam Sagan, Mariusz Łapczyński

Uniwersytet Ekonomiczny w Krakowie

MODELE HYBRYDOWE CART-LOGIT W ANALIZIE PROCESU PODEJMOWANIA DECYZJI W GOSPODARSTWIE DOMOWYM*

Streszczenie: Celem artykułu jest budowa modelu hybrydowego CART-logit, łączącego drzewa klasyfikacyjne (algorytm CART) z dwumianowym modelem logitowym. Zmienna zależna odnosi się do procesu podejmowania decyzji w polskich gospodarstwach domowych i dotyczy dóbr trwałego użytku. Badania wykazały, że hybrydyzacja tych dwóch narzędzi analitycznych poprawiła trafność predykcji modelu i pogłębiła merytoryczną interpretację zależności między zmiennymi.

Słowa kluczowe: modele hybrydowe CART-logit, proces podejmowania decyzji w gospodarstwie domowym, dobra trwałego użytku.

DOI: 10.15611/pn.2014.336.06

1. Wstęp

W analizie procesu podejmowania decyzji można wyróżnić dwa podstawowe podejścia. W pierwszym, reprezentowanym najsilniej przez reprezentantów teorii przetwarzania informacji (*information processing theory*), dominują modele eksplanacyjne, wyjaśniające strukturę procesu postępowania konsumenta, relacje zachodzące między wartościami osobowymi, postawami a preferencjami itp. W nurcie tym popularnymi metodami analizy są modele ścieżkowe i strukturalne (SEM), pozwalające na ocenę siły efektów mediacyjnych i błędów pomiaru dla mierzonych zmiennych ukrytych.

Drugie podejście w obszarze analiz postępowania konsumenta związane jest z podejściem behawioralnym (*behavioral perspective models*), kładącym nacisk na trafność dokonywanych predykcji na podstawie odpowiednio dobranych predyktorów.

* Artykuł powstał na podstawie badań prowadzonych w ramach grantu NCN nr UMO-2011\01\B\HS4\04812, realizowanych przez Centrum Badań i Ekspertyz Uniwersytetu Ekonomicznego w Katowicach.

Celem artykułu jest ocena zastosowania modeli hybrydowych w badaniach procesu podejmowania decyzji, pozwalających na poprawę trafności predykcyjnej, oraz połączenie zalet modeli predykcyjnych i eksplanacyjnych w zintegrowanym układzie analitycznym.

2. Idea modeli hybrydowych CART-logit

Model hybrydowy CART-logit wykorzystany w niniejszych badaniach łączy algorytm drzew klasyfikacyjnych i regresyjnych CART z dwumianowymi modelami logitowymi. CART [Breiman i in. 1993] jest uznawany za najbardziej zaawansowany algorytm do budowy modeli drzew¹. Zmienna zależna i zmienne niezależne mogą znajdować się na dowolnym poziomie pomiaru, a sama analiza nie wymaga spełnienia właściwie jakichkolwiek założeń dotyczących rozkładu zmiennych². Dwumianowe modele logitowe są popularnym narzędziem stosowanym w sytuacji, gdy badacz dysponuje dwuwariantową zmienną zależną i zbiorem dowolnych zmiennych niezależnych (modele te charakteryzują się również słabszymi wymogami w porównaniu z modelami regresji liniowej, dotyczącymi rozkładu zmiennych).

Cechy algorytmu CART odróżniające go od modeli logitowych to:

- automatyczny wybór najlepszych predyktorów (budowa rankingu ważności zmiennych niezależnych),
- brak konieczności transformacji zmiennych (np. logarytmowania, pierwiastkowania),
- automatyczne odkrywanie efektów interakcji,
- odporność na obserwacje nietypowe,
- nie wymagane zastępowanie braków danych,
- w trakcie budowy modelu wymagany jedynie umiarkowany nadzór ze strony badacza.

Algorytm CART rozpoznaje strukturę danych, jednak przy bardzo rozbudowanym modelu drzewa nie zapewnia czytelnej prezentacji modelu. Zdarza się ponadto, że drzewo o dużej liczbie liści przedstawia bardzo prostą zależność między zmiennymi.

Budowa modeli logitowych wymaga z kolei dużej wiedzy badacza, związanej z poprawną specyfikacją modelu, i najczęściej trwa znacznie dłużej niż budowa drzewa klasyfikacyjnego. Modele logitowe są wrażliwe na obserwacje nietypowe i wymagają imputacji braków danych (przypadki z brakami danych są usuwane z analizy). Dużą zaletą tego podejścia jest możliwość obliczenia unikatowego praw-

¹ O drzewach klasyfikacyjnych i regresyjnych w języku polskim pisali m.in.: Gatnar [2001] i Łapczyński [2010].

² Czasami pojawiają się problemy natury technicznej. Zdarza się, że do analizy nie można włączyć jakościowych zmiennych niezależnych z dużą liczbą kategorii albo zbyt dużej liczby zmiennych niezależnych. Dotyczy to jednak zwykle analizowania dużych zbiorów obserwacji, gdzie liczba predyktorów przekracza 1000, a liczba kategorii zmiennych jakościowych jest wyższa od 100.

dopodobieństwa przynależności do klasy (kategorii zmiennej zależnej) dla każdego przypadku.

Za jedną z pierwszych prób łączenia drzew klasyfikacyjnych z modelami logitowymi można uznać hybrydowe podejście CHAID-logit [Lindahl, Winship 1994, s. 729-743]. Hybrydyzacja polegała na sekwencyjnym użyciu tych narzędzi analitycznych. Po wstępnej eksploracji zbioru obserwacji za pomocą algorytmu CHAID podzielono przypadki na rozłączne podzbiory, wykorzystując do tego celu węzły końcowe drzewa. W drugim etapie procedury dla każdego podzbioru budowano niezależne dwumianowe modele logitowe.

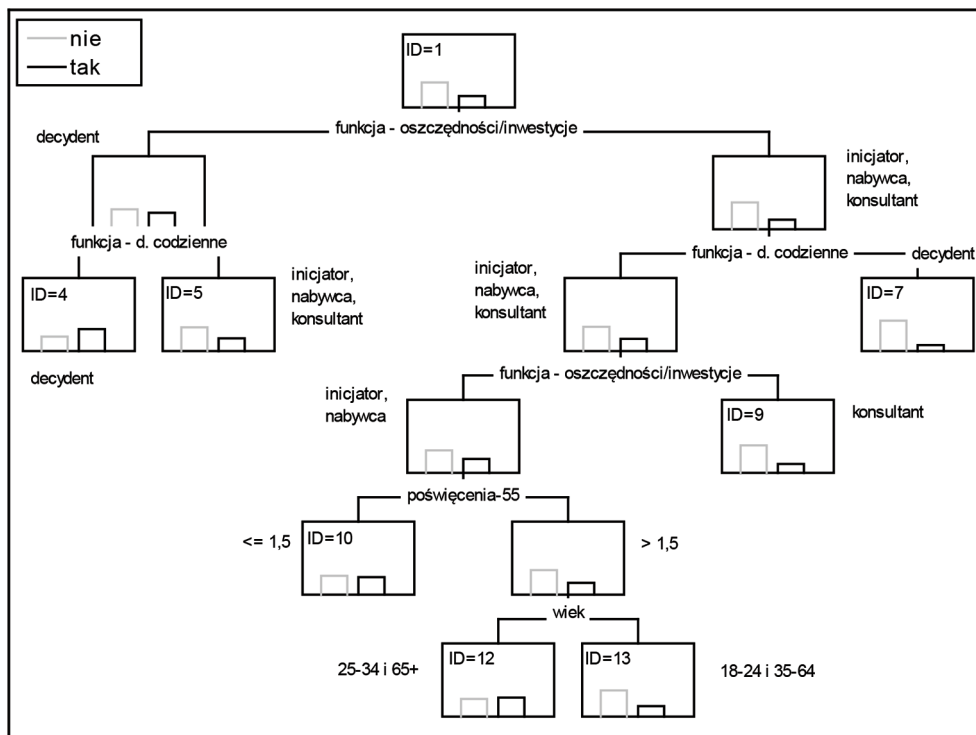
Inną koncepcję hybrydyzacji zaproponowano kilka lat później [Steinberg, Cardell 1998b], łącząc algorytm CART z modelami logitowymi. Tym razem procedura również była dwuetapowa, jednak do zbioru zmiennych niezależnych w modelu logitowym wprowadzono dodatkową zmienną sztuczną, której kategorie informowały o przynależności przypadku do węzła końcowego drzewa klasyfikacyjnego CART. Drzewo powstało w oparciu o ten sam zbiór zmiennych objaśniających, a każdy liść uwzględniał interakcję między predyktorami. Autorzy uznali, że taki sposób hybrydyzacji jest bardziej skuteczny, gdyż podział zbioru obserwacji na podzbiory według pierwszej koncepcji wiąże się z nadmierną redukcją liczebności (przypadki są rozdzielane na liście) oraz utratą informacji (zdarza się, że zbiory zmiennych niezależnych są różne w różnych podzbiorach). Do zalet podejścia CART-logit zaliczyli ponadto wyższą trafność predykcji modelu hybrydowego, szybsze odkrywanie interakcji przez algorytm CART oraz – najczęściej – brak konieczności zastępowania braków danych³.

3. Model drzew klasyfikacyjnych CART dla decydentów w odniesieniu do dóbr trwałych

Przykład modelu hybrydowego CART-logit dotyczy procesu podejmowania decyzji w gospodarstwach domowych⁴. Zmienna zależna odnosi się do roli, jaką odgrywa badany w tym procesie, i ze względu na ograniczoną objętość artykułu dotyczy jedynie decydentów (bez uwzględnienia np. informatorów czy inspiratorów) oraz dóbr trwałego użytku.

³ Wyróżniono kilka sposobów postępowania z brakami danych w modelach hybrydowych CART-logit: 1) ignorowanie braków danych – uzasadniane małą wrażliwością algorytmu na ten problem; 2) przypisanie przypadkom z brakami danych prawdopodobieństw otrzymanych za pomocą algorytmu CART, a pozostałym przypadkom – prawdopodobieństw uzyskanych za pomocą modelu hybrydowego; 3) imputacja braków danych wszelkimi znanymi badaczowi sposobami; 4) wprowadzenie zmiennych sztucznych, informujących o braku danych.

⁴ Badania przeprowadzono w 2012 roku na ogólnopolskiej reprezentatywnej kwotowej, zagnieżdżonej próbie gospodarstw domowych. Badaniami objęto 1100 respondentów z 440 rodzin z 11 województw.



Rys. 1. Model drzewa klasyfikacyjnego CART dla funkcji decydenta w procesie podejmowania decyzji w odniesieniu do dóbr trwałego użytku

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

W pierwszym etapie analizy zbudowano drzewo klasyfikacyjne CART⁵ (rys. 1), z którego wynika, że decydenci w zakresie dóbr trwałego użytku to osoby, które:

- są decydentami w zakresie oszczędzania i inwestowania i jednocześnie są decydentami w zakresie dóbr codziennych (węzeł nr 4);
- pełnią inne role w procesie podejmowania decyzji o nabywaniu dóbr codziennych oraz oszczędzaniu i inwestowaniu, przedkładają dobro rodziny nad własne korzyści i znajdują się w przedziale wiekowym 25-34 albo 65+ (węzeł nr 12).

Ogólna trafność predykcji dla modelu wynosi 66%, trafność predykcji dla kategorii „decydent” – 59%, a trafność predykcji dla kategorii „nie-decydent” – 70%. Z rankingu ważności predyktorów wynika, że zmiennymi niezależnymi mającymi największy wpływ na podejmowanie decyzji w zakresie zakupu dóbr trwałych są

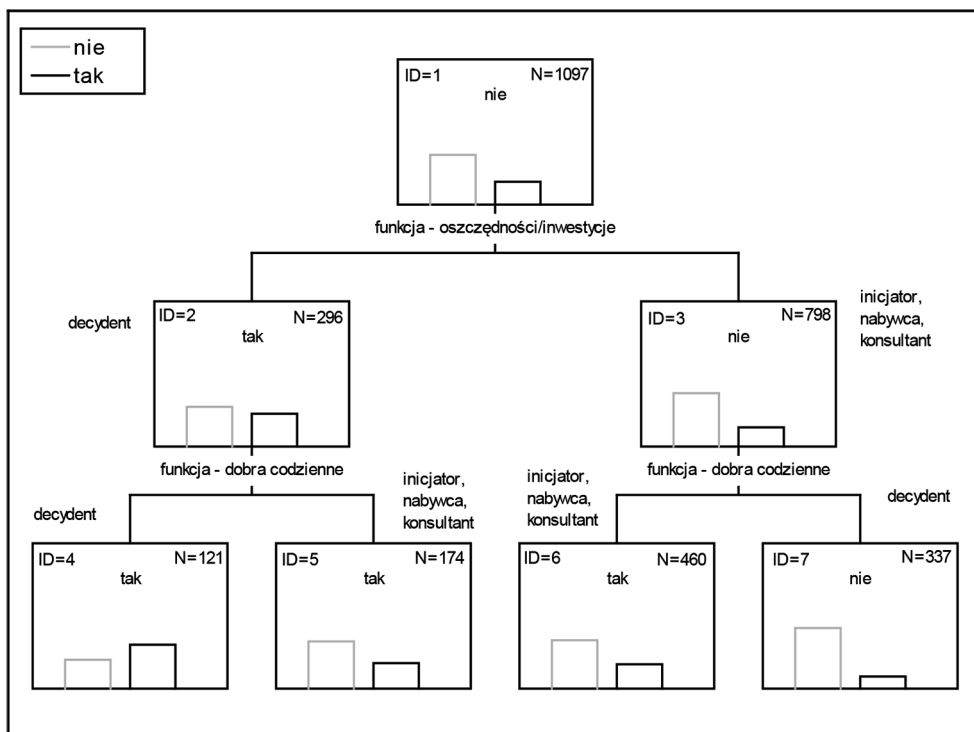
⁵ Ustawienia algorytmu były następujące: prawdopodobieństwa *a priori* równe, koszty błędnych klasyfikacji równe, minimalna liczebność węzła macierzystego 27, minimalna liczebność liścia 27, brak sprawdzianu krzyżowego (walidacji krzyżowej).

funkcje, jakie pełni badany w zakresie dóbr codziennych (ocena = 100) oraz w zakresie oszczędzania i inwestowania (ocena = 63).

4. Model hybrydowy CART-logit dla decydentów w odniesieniu do dóbr trwałych

Przed przystąpieniem do hybrydyzacji zmniejszono głębokość drzewa do dwóch poziomów. Liczba liści równa 4 sprawia, że do modelu logitowego trafiła dodatkowa 4-wariantowa zmienna niezależna. Zredukowany model drzewa klasyfikacyjnego CART (rys. 2) można zinterpretować następująco:

- jeżeli respondent jest decydentem w odniesieniu do oszczędności i inwestycji i jednocześnie jest decydentem w odniesieniu do dóbr kupowanych codziennie, to jest decydentem w odniesieniu do dóbr trwałego użytku z prawdopodobieństwem 0,60 (liść nr 4);



Rys. 2. Zredukowany model drzewa klasyfikacyjnego CART dla funkcji decydenckiego w procesie podejmowania decyzji w odniesieniu do dóbr trwałego użytku

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

- jeżeli respondent jest decydentem w zakresie oszczędności i inwestycji i jednocześnie nie pełni roli decydenta w zakresie dóbr kupowanych codziennie, to jest decydentem w odniesieniu do dóbr trwałego użytku z prawdopodobieństwem 0,35 (liść nr 5);
- jeżeli respondent nie jest decydentem w odniesieniu do oszczędności i inwestycji i jednocześnie nie jest decydentem w odniesieniu do dóbr codziennych, to jest decydentem w odniesieniu do dóbr trwałego użytku z prawdopodobieństwem 0,34 (liść nr 6);
- jeżeli respondent nie pełni roli decydenta w zakresie oszczędności i inwestycji i jednocześnie jest decydentem w zakresie dóbr codziennych, to jest decydentem w odniesieniu do dóbr trwałego użytku z prawdopodobieństwem 0,16 (liść nr 7).

Do modelu logitowego (wyniki przedstawiono w tab. 1) trafiło 10 zmiennych niezależnych, w tym jedna odnosząca się do liścia nr 4 w modelu CART. Ogólnie w grupie predyktorów zwiększających prawdopodobieństwo przynależności respondenta do kategorii decydentów w zakresie dóbr trwałego użytku znajdują się:

- pozycja ze skali LOV – radość życia,
- podział środków zgodnie z zasadą „radość życia czerpię w większym stopniu z dóbr i usług, które służą wszystkim członkom rodziny, niż tych kupowanych na moje osobiste potrzeby” (zmienna „podział środków-26”),
- płeć męska,

Tabela 1. Wyniki hybrydowego modelu CART-logit dla funkcji decydenta w procesie podejmowania decyzji w odniesieniu do dóbr trwałego użytku

Zmienna	Ocena	Błąd standardowy	Wartość p	Iloraz szans
Stała	-1,66	0,54	0,00	0,19
LOV – radość życia	0,10	0,04	0,01	1,11
LOV – samorealizacja	-0,11	0,04	0,01	0,90
Podział środków-26	0,20	0,07	0,01	1,22
Poświęcenia-55	-0,17	0,08	0,03	0,85
Mężczyzna	0,42	0,15	0,00	1,53
Wykształcenie średnie	0,49	0,22	0,03	1,63
Wykształcenie wyższe	0,50	0,22	0,03	1,64
Dobra codzienne – decydent	-0,92	0,18	0,00	0,40
Oszczędności i inwestycje – inicjator	0,37	0,19	0,05	1,44
Węzeł nr 4	2,11	0,26	0,00	8,27

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

- wykształcenie średnie i wyższe,
- pełnienie roli inicjatora w zakresie oszczędności i inwestycji,
- przynależność do liścia nr 4 (tożsame z pełnieniem roli decydenta w zakresie oszczędności i inwestowania i jednocześnie pełnieniem roli decydenta w zakresie dóbr codziennych).

W grupie zmiennych niezależnych zmniejszających prawdopodobieństwo pełnienia funkcji decydenta w zakresie dóbr trwałego użytku znajdują się:

- pozycja ze skali LOV – samorealizacja,
- kierowanie się zasadą, że „uszczerbek na własnym wizerunku nie ma znaczenia, jeżeli w zamian można uzyskać poczucie większego bezpieczeństwa finansowego” (zmienna „poświęcenia-55”),
- pełnienie funkcji decydenta w odniesieniu do dóbr codziennych.

Efekt interakcji z węzła nr 4 wzbogaca merytoryczną interpretację modelu. O ile pełnienie funkcji decydenta jedynie w odniesieniu do dóbr codziennych jest stymulantą, o tyle w połączeniu z funkcją decydenta w odniesieniu do oszczędzania i inwestycji staje się stymulantą w modelu. Iloraz szans równy 8,27 informuje, że prawdopodobieństwo pełnienia roli decydenta w zakresie dóbr trwałego użytku jest ponad 8-krotnie wyższe wśród osób, które są decydentami w zakresie dóbr codziennych i decydentami w zakresie oszczędzania i inwestowania, niż w grupie respondentów pełniących inne role w procesie podejmowania decyzji (*ceteris paribus*)⁶.

Podsumowując, można stwierdzić, że osobami decydującymi o zakupie dóbr trwałych są przede wszystkim mężczyźni, posiadający co najmniej średnie wykształcenie, dbający bardziej o członków rodziny niż o własne korzyści, nie rezygnując przy tym z przyjemności i radości życia. Pełnią oni jednocześnie funkcje decydenta w przypadku pozostałych produktów i usług nabywanych przez gospodarstwo domowe.

Tabela 2. Macierz błędnych klasyfikacji dla modelu hybrydowego CART-logit

Obserwowane	Przewidywane		Razem
	nie-decydent	decydent	
Nie-decydent	446	246	692
Decydent	107	209	316
Razem	553	455	1008

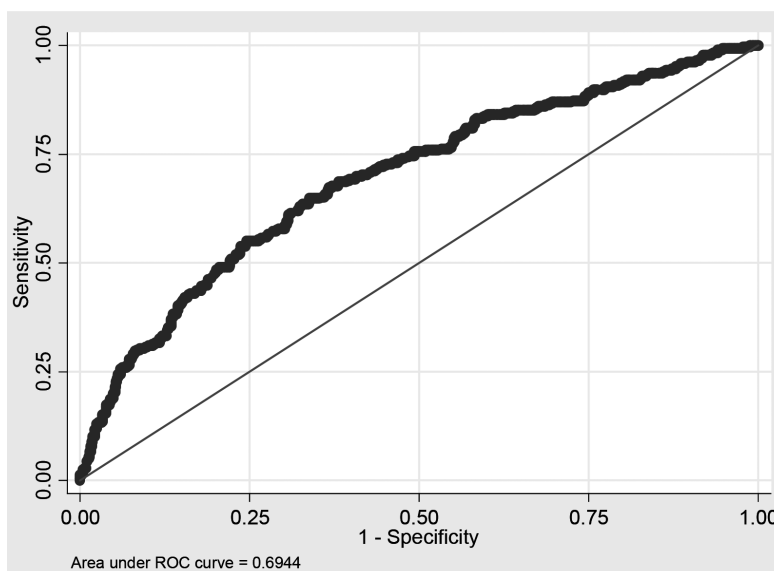
Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

W tabeli 2 znajduje się macierz błędnych klasyfikacji dla modelu hybrydowego CART-logit. Zmodyfikowaną wartość graniczną ustalono na poziomie udziału de-

⁶ Inne wyróżnione w badaniach role pełnione w procesie podejmowania decyzji to: nabywca (ten, który dokonuje fizycznie zakupu), inicjator (ten, który sugeruje podjęcie decyzji), informator (ten, który poszukuje informacji o możliwości zakupu) oraz konsultant/ doradca (ten, który sugeruje, doradza lub odradza zakup).

cydentów w zbiorze obserwacji (0,3135). Ponad 66% przypadków należących do kategorii „decydent w zakresie dóbr trwałego użytku” zostało sklasyfikowanych poprawnie przy ogólnej trafności predykcji równej 65%. Wartość współczynnika R^2 McFaddena wynosi 0,07, zaś wartość współczynnika R^2 Cragga i Uhlera 0,14. Na podstawie danych z macierzy błędnych klasyfikacji można obliczyć kilka wskaźników jakości rozwiązania:

- czułość (*sensitivity, accuracy+, recall, truepositive rate, hit rate*), czyli procent poprawnie sklasyfikowanych przypadków z kategorii „1” (tu: decydenci)⁷: czułość = $\frac{209}{316} = 0,661$;
- specyficzność (*specificity, accuracy-, true negative rate*), czyli procent poprawnie sklasyfikowanych przypadków z kategorii „0” (tu: nie-decydenci)⁸: specyficzność = $\frac{446}{692} = 0,645$;
- precyzja (*precision, positive predictive value, response rate*), czyli procent przypadków z przewidywanej kategorii „1”, będących rzeczywiście przypadkami z kategorii „1” (tu: decydenci): precyzja = $\frac{209}{455} = 0,459$.



Rys. 3. Krzywa ROC dla modelu hybrydowego CART-logit

Źródło: opracowanie własne z wykorzystaniem programu STATA 13.

⁷ Procent przypadków prawdziwie dodatnich.

⁸ Procent przypadków prawdziwie ujemnych.

Czułość i specyficzność bywają wykorzystywane do budowy wykresu krzywej operacyjno-charakterystycznej (ROC) [Sagan 2011], którą przedstawiono na rys. 3. Obszar pod krzywą ROC (*area under ROC curve*, AUC), który przyjmuje wartości od 0 do 1, jest tutaj równy 0,6944, co wskazuje na dosyć dobrą klasyfikację przypadków. Wartości na osi X (*1-specificity*) są interpretowane jako odsetek przypadków fałszywie dodatnich, czyli innymi słowy tych nie-decydentów, którzy zostali sklasyfikowani jako decydenci. Wartości na osi Y (*sensitivity*) to z kolei odsetek przypadków prawdziwie dodatnich, czyli decydentów, którzy zostali sklasyfikowani jako decydenci. Punkt na krzywej ROC wysunięty najwyżej w lewą stronę (najdalej na „północny zachód”) wskazuje na sytuację, gdy występuje najwyższy procent przypadków prawdziwie dodatnich i jednocześnie najniższy procent przypadków fałszywie dodatnich. Dla „idealnego” rozwiązania punkt ten miałby współrzędne (0;1).

5. Podsumowanie

Łączenie różnych narzędzi analitycznych w budowie jednego modelu predykcyjnego (prognostycznego) lub opisowego jest z całą pewnością czasochłonne i wymaga od badacza dużej wiedzy i doświadczenia. Wysiłek poniesiony w trakcie analizy bywa jednak kompensowany lepszym rozwiązaniem. Określenie „lepszy” odnosi się w tym wypadku do wyższej trafności predykcji kategorii zmiennej zależnej, którą zainteresowany jest badacz, oraz pogłębionej interpretacji badanego zjawiska. Podejścia hybrydowe wyznaczają nowe kierunki w badaniach zachowań konsumentów, pozwalające na łączenie podejść eksplanacyjnych i predykcyjnych.

Dzięki hybrydyzacji algorytmu CART i modelu logitowego uzyskano wyższą wartość współczynnika czułości (*recall*), a także bardziej dokładną interpretację zależności pomiędzy zmiennymi (ilorazy szans w modelu logitowym) oraz szybko odkryto interesujące interakcje pomiędzy zmiennymi (dzięki algorytmowi CART). Wydaje się, że ten sposób postępowania podczas budowy hybrydowych modeli predykcyjnych przynosi wiele korzyści, co potwierdzają również inne – nieliczne jak dotąd – badania w tym obszarze [Ware 2003; Gansky, Cheng, Pollick 2005; Łapczyński 2009].

Literatura

- Breiman L. i in., *Classification and Regression Trees*, Chapman and Hall, New York 1993.
- Gansky S.A., Cheng N., Pollick H.F., *Predicting Early Childhood Caries with Individual, Family and Neighborhood Factors*, Proceedings of IADR, Baltimore 2005.
- Gatnar E., *Nieparametryczna metoda dyskryminacji i regresji*, PWN, Warszawa 2001.
- 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., *Drzewa klasyfikacyjne i regresyjne w badaniach marketingowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków 2010.

- Łapczyński M., *Modele hybrydowe CART-LOGIT w analizie danych rynkowych*, [w:] J. Dziechciarz (red.), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu 2009, nr 51, s. 85-95.
- Sagan A., *Krzywe operacyjno-charakterystyczne w ewaluacyjnych badaniach marketingowych*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie nr 864, Kraków 2011, s. 15-17.
- Steinberg D., Cardell N.S., *Improving Data Mining with New Hybrid Methods*, <http://www.salford-systems.com>, 27 maja 1998a (data pobrania pliku: październik 2010).
- Steinberg D., Cardell N.S., *The hybrid CART-logit model in classification and data mining*, www.salford-systems.com, 1998b (data pobrania pliku 10.01.2008).
- Ware B.C., *Reader Involvement: Who? What? Where?*, Proceedings of Worldwide Readership Research Symposium, Munich 2003, s. 371-394.

CART-LOGIT HYBRID MODELS IN THE ANALYSIS OF DECISION-MAKING PROCESS IN THE HOUSEHOLDS

Summary: The aim of this article is to develop a hybrid CART-logit model that combines decision trees (CART algorithm) with binomial logit model. The dependent variable refers to the decision-making process in Polish households in the area of durable goods. The results have shown that hybridization of these two analytical tools can improve the accuracy of prediction and enrich the interpretation of relationships among variables.

Keywords: CART-logit hybrid models, decision-making process, durable goods.