

Mariusz Trojanowski

Uniwersytet Warszawski
mtrojanowski@wz.uw.edu.pl

PRZEWIDYWANIE ZACHOWAŃ KONSUMENTÓW ZA POMOCĄ ANALIZY SCORINGOWEJ

PREDICTION OF CONSUMER BEHAVIOR BY SCORING ANALYSIS

DOI: 10.15611/pn.2018.525.07

JEL Classification: M31, C8

Streszczenie: Przewidywanie zachowań nabywców, z jednej strony, jest coraz trudniejsze, z drugiej zaś rosną możliwości techniczne gromadzenia i analizowania wielu danych na temat nabywców (*Big Data*) oraz odkrywania niewidocznych na pierwszy rzut oka zależności w zachowaniach konsumentów (*Data Mining*). Celem artykułu jest przedstawienie sposobu przeprowadzenia analizy scoringowej, należącej do grupy analiz predykcyjnych, służących określeniu prawdopodobieństwa wystąpienia pewnych zdarzeń, w tym przypadku pozytywnej reakcji adresatów działań marketingowych. W tekście wykorzystano analizę przypadku organizacji społecznej SOS Wioski Dziecięce, której komunikaty marketingowe skierowane były do potencjalnych donatorów, wybranych na podstawie wyników analizy scoringowej. Analiza ta pokazuje, jak cennym uzupełnieniem tradycyjnych badań marketingowych może być analiza scoringowa. Artykuł jest jednocześnie postulatem szerszego zainteresowania się tym kierunkiem prowadzenia badań i analiz nabywców.

Słowa kluczowe: analiza nabywców, analiza scoringowa, predykcja, przewidywanie zachowań, *Data Mining*, SOS Wioski Dziecięce.

Summary: Predicting buyer behavior is more and more difficult today. On the other hand, the technical possibilities of gathering and analyzing a lot of data about buyers (*Big Data*) and discovering the invisible dependencies in consumer behavior (*Data Mining*) are growing. The aim of the article is to present the method of conducting the scoring analysis, belonging to the group of predictive analyzes, used to determine the probability of occurrence of certain events, in this case the positive reaction of the addressees of marketing activities. The text uses the case study of the social organization “SOS Children’s Villages”, whose marketing messages were addressed to potential donors, selected based on the results of the scoring analysis. The case study shows how a valuable complement to traditional marketing research can be scoring analysis. The article is also a postulate of a wider interest in this direction of research and analysis of buyers.

Keywords: customers’ analysis, scoring analysis, prediction of consumer behavior, *Data Mining*, SOS Children’s Villages.

1. Wstęp

Badanie zachowań nabywców jest jednym z podstawowych kanonów marketingu i jednym z jego najważniejszych obszarów analitycznych [Sojkin 2009, s. 57]. Szczególnie ważne jest przewidywanie reakcji konsumentów na różnego rodzaju stymulacje marketingowe, co ostatecznie sprowadza się do tego, czy adresat działań dokona wyboru sprzedawanej marki, czy z niej zrezygnuje. Co ważne, przewidywanie zamierzeń konsumentów jest jednym z najtrudniejszych zadań stojących przed badaczami marketingowymi [Mazurek-Łopacińska (red.) 2016, s. 147]. W rzeczywistości częściej polega na badaniu nie samych zamiarów czy planów zakupowych, ale na rozpoznawaniu uwarunkowań decyzyjnych konsumentów, a przez to prognozowaniu ich zamiarów zakupowych [Kaczmarczyk 2007, s. 150-151]. Inaczej mówiąc, badania zachowań nabywców najczęściej odpowiadają na pytania, jak i dlaczego zachowują się oni obecnie w taki, a nie inny sposób, zaś o wiele trudniej jest odpowiedzieć na pytanie, jak zachowają się w przyszłości [Woś, Rachocka, Kasperek-Hoppe 2011, s. 103]. Ważne jest zbieranie różnorodnych informacji o konsumentach oraz posługiwanie się w tym celu wieloma zróżnicowanymi metodami i źródłami [Mróz 2013, s. 231-236].

W obszarze badań marketingowych mamy do dyspozycji wiele różnych metod i technik, które służą przewidywaniu zachowań nabywców, jak chociażby: metody prognozowania o charakterze subiektywnym [Kędzior 2005, s. 176-177], badanie postaw z założeniem, że ich znajomość daje możliwość przewidywania późniejszych zachowań [Kieźel, Smyczek 2015, s. 177], badanie zamiarów zakupowych [Światowy 2016, s. 206-207] i inne. Niestety, zdolność przewidywania przyszłych zachowań nabywców na podstawie wymienionych metod jest dość ograniczona ze względu na różne czynniki, w tym przede wszystkim z uwagi na deklaratorywność uzyskanych odpowiedzi respondentów oraz zmieniające się czynniki sytuacyjne. Uzyskane w danym momencie wnioski z badań oraz określane na tej podstawie zamiary zakupowe mogą w skrajnym wypadku zupełnie różnić się z rzeczywistymi decyzjami zakupowymi. Rozminięcie jest tym bardziej prawdopodobne, gdyż coraz trudniej możemy przewidywać rozwój zmian w otoczeniu rynkowym. One zawsze utrudniały badanie przyszłych zachowań nabywców. Jednakże współcześnie dynamika zmian ma niespotykaną jak dotąd skalę [Mazurek-Łopacińska (red.) 2016, s. 23]. Wiele zjawisk, takich jak zmiany w otoczeniu ekonomicznym, kulturowym, a zwłaszcza technologicznym, ma charakter nieciągły. Przez to ich przebieg jest trudniejszy do przewidzenia [Lewandowska 2010]. W tej sytuacji bardzo trudno jest antycypować zachowania konsumentów.

Z drugiej strony nowoczesne technologie informatyczne, jak również postęp w zakresie metod badań oraz analityki marketingowej tworzą nowe, niespotykane dotąd, a na pewno rzadziej używane, możliwości opisywania konsumentów i przewidywania ich przyszłych zachowań [Mróz 2013, s. 241-245]. Postęp technologiczny i rozwój Internetu napędzają zjawisko *Big Data* – tworzą wielkie możliwości

agregowania ogromnych ilości informacji o nabywcach, w tym informacji o ich zachowaniach, łączenia różnych baz danych i wyciągania wniosków [Maślankowski 2015; Miciuła, Miciuła 2015]. Analizy marketingowe prowadzone w oparciu o agregowane wielkie zbiory danych (*Data Mining*) wydają się już nie tylko uzupełnieniem tradycyjnych metod badawczych, ale też mogą stać się podstawą w badaniu przyszłych zachowań nabywców, a na pewno dane zapisane w obszernych bazach danych mogą bardzo pomagać w działaniach marketingowych [Sojkin 2009, s. 119]. Na razie obserwuje się przewagę deklaratywnego zainteresowania przedsiębiorstw koncepcją *Big Data* niż faktycznego jej zastosowania w praktyce rynkowej [Kryśkiewicz 2017].

2. Zgłębianie wiedzy o klientach – *Data Mining*

Podstawą analiz marketingowych prowadzonych przez przedsiębiorstwa z zakresu *Big Data* jest elektroniczna baza danych, co współcześnie oznacza różne zbiory danych wymagające łącznego analizowania [Saboo Kumar, Park 2016, s. 911-939; Weinert 2017]. Wiedza o nabywcach zapisana w bazie danych jest sumą wiedzy o pojedynczych nabywcach. W bazie danych zapisuje się dane imiennie-adresowe lub tylko dane na temat IP komputera czy smartfona użytkownika, dane o jego aktywności w sieci, dane transakcyjne – co kupował, jak często, gdzie, jaką metodą dokonywał płatności itd. [Bird 2011, s. 253-255; Trojanowski 2010, s. 62-65]. Od jakości bazy danych, aktualności i kompletności zawartych w niej informacji zależy jakość dokonywanych analiz.

Wśród prowadzonych na użytek marketingu analiz opartych na bazie danych można wymienić takie analizy, jak: analizę zachowań zakupowych klientów, w tym analizy typu RFM (*Recency-Frequency-Money*), pokazujące, jak często i jakiej wartości produkty kupuje dany klient [Chang, Hsiao-Ping 2011; Kozielski (red.) 2006, s. 130-131]; analizę rentowności każdego klienta oraz wartości życiowej LTV (*Live-Time Value*) pozwalającą ustalić wartość dochodu, jaki firma może osiągnąć, obsługując danego klienta w określonym przedziale czasu [Grzegorzczak 2007, s. 119; Hosseini, Mohammadzadeh 2016, s. 355-367; Venkatesan, Kumar 2004, s. 108]; analizę segmentacyjną opartą na wartości każdego klienta, ale także na wzorcach zachowań zakupowych [Migut 2005, s. 15].

Szczególne znaczenie mają w tej grupie analizy charakteryzujące profile klientów i ich zachowania, pozwalające budować modele zachowań poprzez zgłębianie danych lub tłumacząc wprost, tzw. kopanie w danych zgromadzonych w bazie (*Data Mining*). W tym zakresie warto przede wszystkim wymienić analizę prawdopodobieństwa pozytywnej reakcji (tzw. analiza scoringowa).

Metody *Data Mining* polegają na zestawianiu ze sobą różnych informacji pochodzących z różnych zbiorów danych, a przez to pozwalają na wyszukiwanie w nich interesujących związków oraz nieznanych wcześniej zależności [Hand, Mannila, Smyth 2001]. *Data Mining* wykorzystuje zaawansowane narzędzia statystyczne,

narzędzia analiz z wbudowanymi modułami sztucznej inteligencji oraz komputery o ogromnych możliwościach obliczeniowych. Dzięki tym technologiom możliwe jest szybkie przeszukiwanie wielkich ilości informacji i tworzenie reguł asocjacyjnych, tj. reguł, jakie wiążą ze sobą poszczególne dane (tabela 1). Reguły asocjacyjne mogą być oparte na zmiennych jakościowych lub ilościowych, które dotyczą jednego lub więcej wymiarów, cech czy atrybutów [Paszyła 2005]. W przypadku transakcji zmienne jakościowe określają atrybuty poddane obserwacji. Na przykład, czy produkt A został kupiony („1”), czy nie („0”), w jakiej wersji został kupiony (tutaj jest więcej wartości, z których każda odpowiada np.: konkretnej wersji produktu, jaki jest wiek klienta itd.). Dane te mogą być zapisane w postaci tabelarycznej. Natomiast zmienne ilościowe określają ilość lub wartość kupionego produktu i pozwalają bardziej precyzyjnie zdefiniować zależności.

Tabela 1. Reguła asocjacyjna

„JEŻELI [poprzednik] TO [następnik]	
Reguła zapisywana jest za pomocą warunków:	
[warunki poprzednika] => [warunki następnika]	
Przykłady:	[zakup kawy] => [zakup śmietanki do kawy]
	[wiek klienta] => [intensywność spożycia określonego produktu]

Źródło: opracowanie własne.

Reguły asocjacyjne pozwalają wyodrębnić pewne zależności między zmiennymi oraz określać prawdopodobieństwo ich powtarzania. Najważniejsze prawidłowości, które są istotne z punktu widzenia marketingu, to zależności pomiędzy zachowaniami nabywców oraz sekwencje zdarzeń. W tym celu używa się m.in. analizy koszykowej, która pierwotnie była wykorzystywana w supermarketach do odkrycia tego, które produkty były zwykle kupowane przez klientów łącznie. Wiedza ta służyła detalistom do podejmowania decyzji odnośnie do ułożenia asortymentu w jednostce handlowej oraz planowania promocji. Inną metodą jest analiza sekwencji. Służy do tworzenia reguł zachowań konsumenckich, a przy tym dodatkowo uwzględnia kwestię czasu, co pozwala określić nie tylko, CO klient prawdopodobnie kupi, ale także KIEDY to zrobi.

Badanie zależności w zachowaniach konsumentów może także służyć do przewidywania przyszłych zachowań klientów poprzez budowanie tzw. modeli predykcyjnych.

3. Modele predykcyjne (scoringowe)

W wyniku analiz *Data Mining* można zbudować model predykcyjny, czyli przewidujący, jak zachowają się bądź jak nie zachowują się klienci o określonej charakterystyce w określonej sytuacji [Lee, Lee, Sohn 2013].

Modele predykcyjne wykorzystywane w badaniach marketingowych mają konkretne zastosowanie. Ich rolą jest przewidywanie rzeczywistości. Jednocześnie mogą nie sprawdzać się w jej wyjaśnieniu, co pokazuje problem najlepszej relacji między poziomem błędu a wariancją uzyskanych estymatorów – tzw. *bias-variance trade off* [Sagan 2014, s. 5-7]. Modele predykcyjne obciążone są błędem specyfikacji, czyli słabo wyjaśniają rzeczywistość, ale mają dużą moc przewidywania jej stanu przyszłego. Rolę wyjaśniania rzeczywistości pełnią modele eksplanacyjne, które mają pokazywać zależności przyczynowo-skutkowe między różnymi zmiennymi. Te modele powinny cechować się zarówno dużą mocą wyjaśniającą (niskim błędem specyfikacji), jak i dużą mocą predykcyjną (niskim błędem przewidywania). Tendencje w badaniach społecznych pokazują, że przechodzi się od modeli eksplanacyjnych do predykcyjnych [Sagan 2016, s. 276-284].

Do modeli predykcyjnych należą modele scoringowe (w skrócie scoringi). Są to modele punktowe, a ściślej mówiąc, modele przyznające punkty. Punkty przyznawane są wybranym do analizy cechom badanego zdarzenia z przeszłości, a następnie otrzymane wyniki zamieniane są na prawdopodobieństwo wystąpienia tego zdarzenia w przyszłości [Hsieh 2004, s. 623-633]. Każdy model scoringowy jest tworzony do badania ściśle określonego zdarzenia, które da się opisać zero-jedynkowo, czyli albo zdarzenie nastąpi, albo nie nastąpi.

Modele scoringowe są wykorzystywane do różnych celów. Najczęściej stosuje się je w bankowości do określenia prawdopodobieństwa, czy klient spłaci kredyt, czy nie spłaci [Hryniewicz 2016], ale także w ubezpieczeniach, marketingu, medycynie i innych obszarach. Marketing jest obszarem coraz częstszego stosowania scoringów [Malthouse 2010]. Co ciekawe, modele scoringowe po raz pierwszy zastosowano w marketingu już w latach 30. w USA w sprzedaży wysyłkowej, gdy stosowano te metody do wybierania adresatów kampanii, rokujących największe prawdopodobieństwo odpowiedzi na ofertę [Korol 2013, s. 51].

Niezbędnym elementem do rozpoczęcia procesu budowy modelu jest posiadanie zbioru danych historycznych, opisujących podobne zdarzenie w niedawnej przeszłości, oraz skonfigurowanie (zestawienie) tych cech z faktem, czy zdarzenie nastąpiło, czy nie nastąpiło. Jest to tzw. próba ucząca. W przypadku prognozowania reakcji nabywców na określone działanie marketingowe dane historyczne muszą pokazywać zachowania odbiorców wobec podobnej akcji, jaka miała miejsce w przeszłości. Ich zachowanie jest zmienną zależną, a zmiennymi niezależnymi są różne cechy nabywców, takie jak wiek, płeć, miejsce zamieszkania itd., zachowania zakupowe, jak: zakup innego produktu w przeszłości, częstotliwość kupowania różnych produktów, wielkość wydatków, fakt niedawnego zakupu innego produktu, na podstawie wcześniej odkrytych sekwencji itd., posiadanie określonych produktów (wyposażenie domu, konkretna marka auta czy telefonu komórkowego itd. W oparciu o te dane tworzy się algorytmy drzew decyzyjnych CHAID, obrazujące zależności zaobserwowane w zbiorze danych historycznych, czyli w zbiorze uczącym [Jaśko 2012, s. 155-169]. Następnie w wyniku przekształceń matematycznych powstaje karta sco-

ringowa, w której każdemu atrybutowi wybranej cechy przypisuje się wartość punktową. Suma punktów uzyskanych przez zespół cech opisujących badane zdarzenie pozwala odczytać rezultat modelu scoringowego. Każdej wartości punktowej jest przypisane prawdopodobieństwo zaistnienia badanego zdarzenia [Wysiński 2013, s. 253-268]. Wracając do przypadku akcji marketingowej, na podstawie danych wejściowych wyznacza się wzór na prawdopodobieństwo, że dany klient odpowie pozytywnie na kolejną podobną akcję.

Poniższy przykład przedstawia budowę modelu predykcyjnego w odniesieniu do akcji wysyłkowej. Dotyczy on optymalnej selekcji klientów z bazy danych do przedstawienia im oferty określonego produktu [Goldian 2002]. Dane wejściowe do modelu uzyskano z akcji testowej skierowanej do wybranych odbiorców o określonej z góry charakterystyce. Następnie dokonano wyboru zmiennej zależnej, którą było zamówienie oferowanego produktu. Zmienna miała dwie wartości – sukces (1, np. zamówienie produktu) i porażka (0, np. brak zamówienia produktu). Dokonano wyboru także zmiennych opisujących adresata, czyli zmiennych niezależnych determinujących zmienną zależną. Były to takie zmienne, jak: wiek klienta (A), od ilu miesięcy konsument był zapisany w bazie (B), ile miesięcy upłynęło od ostatniej pozytywnej reakcji (C), na ile akcji zareagował pozytywnie (D). Zmienną zależną (złożenie zamówienia) określono jako (E). Dane zawarte są w tabeli 2.

Tabela 2. Przykład modelu predykcyjnego – dane wejściowe do budowy modelu oraz policzone na podstawie modelu prawdopodobieństwo reakcji odbiorcy

Zmienne Klienci	A	B	C	D	E	PS
Klient nr 1	25	37	11	11	0	0,67
Klient nr 2	24	29	23	3	0	0,17
Klient nr 3	28	46	20	4	0	0,22
Klient nr 4	26	46	2	9	1	0,65
Klient nr 5	22	49	2	13	1	0,83*
Klient nr 6	24	49	13	6	0	0,39
Klient nr 7	35	32	8	3	1	0,23
Klient nr 8	26	49	1	15	1	0,88*
Klient nr 9	30	30	15	3	1	0,20

Źródło: opracowanie na podstawie: [Goldian 2002].

Posiadając dane wejściowe (A-E), można sformułować model. Do tego celu służy jedna z metod statystycznych, tzw. regresja logistyczna. Wagi poszczególnych zmiennych można ustalić przy pomocy jednego z pakietów statystycznych. Wynikiem jest wzór uzależniający prawdopodobieństwo sukcesu (PS) od wybranych i wyliczonych zmiennych zależnych. Prawdopodobieństwo sukcesu określa równanie:

$$P(S) = \frac{e^{a_1 * X_1 + a_2 * X_2 + a_3 * X_3 + a_4 * X_4 + \dots}}{1 + e^{a_1 * X_1 + a_2 * X_2 + a_3 * X_3 + a_4 * X_4 + \dots}},$$

gdzie: $P(S)$ – prawdopodobieństwo sukcesu; X_1, X_2, X_3, X_4 – zmienne niezależne; a_1, a_2, a_3 – wagi zmiennych niezależnych.

Otrzymane wagi i wzór na sukces: A:-0,01790; B:0,0026; C:0,03850; D:0,2295.

Policzone wartości prawdopodobieństwa dla każdego klienta według otrzymanego wzoru zawarte są w tabeli 2 (oznaczenie w tabeli symbolem $P(S)$). Wyniki wskazują, że największe prawdopodobieństwo pozytywnej reakcji na daną akcję występuje w przypadku klientów nr 8 i 5. W efekcie zastosowania powyższego modelu scoringowego do właściwej akcji marketingowej zostali wybrani adresaci o charakterystyce podobnej do wyróżnionych powyżej klientów. Decyzją kierownictwa marketingowego może być ewentualne obniżenie oczekiwań co do prawdopodobieństwa sukcesu i objęcie akcją również odbiorców nr 1 i 4, o ile prawdopodobieństwo będzie satysfakcjonujące.

Scoringi powstają na podstawie zachowań innych klientów, którzy w przeszłości dokonali zakupu danej oferty. I to jest w całej analizie bardzo ważne założenie, a zarazem podstawa ryzyka, jakie jest wpisane w budowę samego modelu. Chodzi o to, że klienci, którzy mają kupić dany produkt teraz, niekoniecznie muszą zachować się w taki sam sposób, jak inne osoby zachowały się wcześniej. Mogły się zmienić warunki rynkowe oraz rodzaj prowadzonych działań marketingowych przez przedsiębiorstwo i jego konkurentów. Z drugiej strony istnieje ogólna prawidłowość mówiąca, że osoby o podobnej charakterystyce podobnie się zachowują. Modele scoringowe korzystają z tej zasady, choć należy pamiętać o jej względności.

4. Analiza przypadku zastosowania analizy scoringowej przez SOS Wioski Dziecięce ¹

4.1. O SOS Wioski Dziecięce – zarys sytuacji

Stowarzyszenie SOS Wioski Dziecięce jest organizacją pożytku publicznego, pomagającą dzieciom opuszczonym i osieroconym. W Polsce funkcjonują cztery wioski: w Biłgoraju, Kraśniku, Siedlcach i Karlinie, z których każda obejmuje kilkanaście domów jednorodzinnych. Pierwsza wioska SOS powstała w Biłgoraju w 1984 roku. Oprócz wiosek SOS Stowarzyszenie prowadzi także Dom Młodzieży SOS w Lublinie, Wspólnotę Mieszkaniową SOS dla młodzieży w Kraśniku oraz rodzinne domy czasowego pobytu dla dzieci. Stowarzyszenie jest członkiem światowego stowarzyszenia SOS – Kinderdorf International.

Stowarzyszenie musi pozyskiwać fundusze na swoją działalność. Ma tutaj konkurencję w postaci wielu organizacji typu non profit działających w Polsce. Kilka

¹ Za zgodą Acxiom Polska. Zob. [Trojanowski 2010, s. 122].

z nich ma ogólnopolski, a nawet światowy zasięg, znaną powszechnie markę i największe gromadzone środki. Są to takie organizacje, jak: UNICEF, międzynarodowa organizacja ekologiczna – WWF, Fundacja Dr Clown, pomagająca chorym dzieciom przebywającym w szpitalach i placówkach specjalnych, Wielka Orkiestra Świątecznej Pomocy, Fundacja Polsat, Fundacja TVN i inne.

W celu pozyskiwania donatorów Stowarzyszenie stosuje różnego rodzaju narzędzia marketingowe. Są to przede wszystkim: akcje marketingu bezpośredniego (wysyłki poczty elektronicznej i tradycyjnej), ulotki pozostawiane pod drzwiami mieszkań i domów (beadresowy *door drop*), dołączanie oferty do ofert innych firm (akcje *co-brandingowe*), medialna reklama w prasie, radiu i telewizji oraz akcje specjalne (*eventy*).

Wszystkie akcje muszą być planowane tak, aby ryzyko ponoszonych inwestycji było minimalne. Szczególny nacisk kładziony jest na te kampanie, których efekty można precyzyjnie mierzyć i określać, ile środków pozyskano w wyniku każdej z nich. Dlatego kluczowe w marketingu Stowarzyszenia jest możliwie precyzyjne analizowanie reakcji odbiorców akcji – potencjalnych donatorów – i na tej podstawie określanie, do kogo akcja powinna być skierowana.

4.2. Opis przypadku – zastosowanie analizy scoringowej

Celem analizowanej akcji marketingowej było podniesienie jej odzewu, tj. zwiększenie odsetka osób pozytywnie reagujących na prośbę o wpłaty funduszy na cele Stowarzyszenia. Dodatkowym celem było podniesienie średniej wartości jednorazowej wpłaty. Analizowany przypadek wykorzystania modelu scoringowego zaczął się od testu baz danych oferowanych przez partnera Stowarzyszenia, jakim był broker baz danych, firma Acxiom. Ze względu na fakt, że do testowej akcji nie było możliwości dobrania bazy według kryteriów określających skłonność do wpłacania pieniędzy na cele charytatywne, pierwszym krokiem w kierunku doprecyzowania grupy docelowej było umieszczenie przez SOS Wioski Dziecięce pytań w Ogólnopolskim Kwestionariuszu Acxiom, ściśle weryfikujących grupę celową (tabela 3).

Tabela 3. Pytania umieszczone w kwestionariuszu Acxiom w celu określenia ogólnych charakterystyk donatorów

Pytanie A14. Czy kiedykolwiek wspomagał(a) lub zamierza wspomagać Pan(i) akcje charytatywne? Wspomagałem(am): Tak <input type="checkbox"/> Nie <input type="checkbox"/> Zamierzam wspomagać: Tak <input type="checkbox"/> Nie <input type="checkbox"/>	
Pytanie A15. Jakie cele zamierza Pan(i) wspierać finansowo:	
01 <input type="checkbox"/> głodujące dzieci na świecie	03 <input type="checkbox"/> schroniska dla zwierząt
02 <input type="checkbox"/> osierocone dzieci	04 <input type="checkbox"/> inne

Źródło: Ogólnopolski Kwestionariusz Produktów i Usług, Acxiom Polska.

Zawarte w badaniu Acxiom pytania pozwoliły pozyskać bazę danych ponad 100 000 osób, które deklarowały chęć wspierania akcji charytatywnych (w pyta-

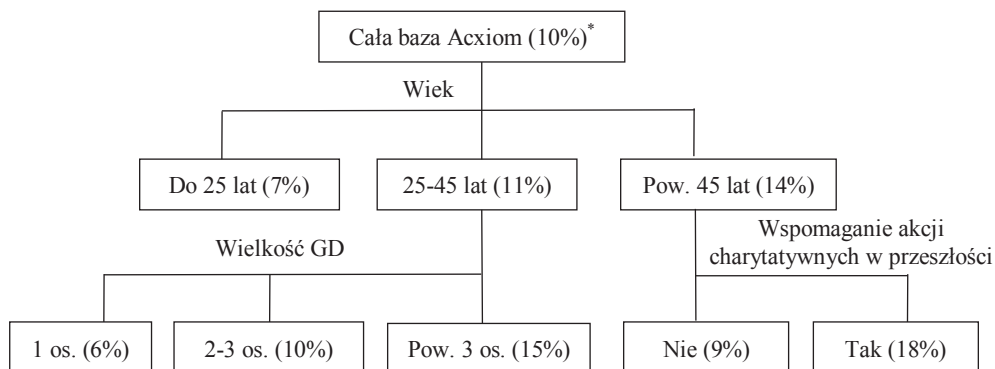
niu A14 – tabela 3 – zaznaczyły opcję „Tak” Zamierzam wspomagać). SOS Wioski Dziecięce zdecydowały się na wysyłkę pocztową swojej oferty do tych właśnie osób z propozycją wpłacenia jakiegokolwiek kwoty na konto Stowarzyszenia. Niespełna jedna piąta adresatów (18%) pozytywnie zareagowała na tę ofertę. Ta grupa 1800 osób stała się tzw. próbą uczącą do przeprowadzonej analizy scoringowej.

Ważne jest to, że wszyscy respondenci wypełniający kwestionariusz Acxiomu, w tym także osoby z próby uczącej, podali bardzo dużo informacji na swój temat, odpowiadając na szereg pytań, nie tylko dotyczących ich charakterystyki demograficznej (m.in. wiek, płeć, miejsce zamieszkania, struktura gospodarstwa domowego), ale także ich preferencji, dokonywanych wyborów itd. Dzięki temu próba ucząca była opisana przez bardzo wiele cech, a każda cecha miała przyporządkowany konkretny atrybut. Te dane pozwoliły opisać profil każdego z 1800 donatorów.

Dzięki powyższym danym można było przeprowadzić pierwszy etap analizy scoringowej, tzn. analizę profilu grupy uczącej – czyli grupy obecnych donatorów. Przeprowadzenie tej analizy pozwoliło zidentyfikować główne prawidłowości w danych i zobrazować cechy charakterystyczne dla analizowanej grupy. Wyniki analizy wskazywały, że na wsparcie działań SOS częściej decydują się osoby powyżej 45. roku życia, które już kiedyś brały udział w innych akcjach dobroczynnych, osoby z gospodarstw domowych powyżej trzech osób, kobiety, osoby z wyższym wykształceniem, czytające częściej gazety opiniotwórcze, takie jak „Gazeta Wyborcza” czy „Rzeczpospolita”. Z tym profilem wiązało się również szereg innych cech, które w istotny sposób wyróżniały grupę donatorów z ogółu osób znajdujących się w bazie Acxiom, takie jak pozycja zawodowa, hobby, zainteresowania, wyposażenie gospodarstwa domowego w sprzęt AGD i RTV i szereg innych cech spośród ponad 1000 zmiennych opisujących respondentów Ogólnopolskiego Kwestionariusza Produktów i Usług. Im więcej ww. cech posiadała dana osoba, tym większe było prawdopodobieństwo jej pozytywnej reakcji na ofertę donacji.

Drugim etapem analizy było wykorzystanie technik *Data Mining* i zeksplorowanie całej bazy Acxiom oraz innych dostępnych baz, w których posiadaniu było Stowarzyszenie, pod kątem wyszukiwania w nich osób o profilu najmocniej zbliżonym do profilu obecnych donatorów z próby uczącej. Aby wykorzystać w pełni potencjał informacji zawarty w uzyskanych danych, do opracowania scoringu określającego skłonność do udzielenia wsparcia finansowego dla SOS Wioski Dziecięce wykorzystano algorytmy drzew decyzyjnych CHAID i regresji logistycznej. Rysunek 1 pokazuje algorytm drzewa decyzyjnego, jaki wykorzystano w analizie. Algorytm, wychodząc od analizy całej bazy Acxiom, szuka parametru najsilniej determinującego fakt udzielenia donacji. W tym przypadku takim parametrem okazał się wiek, ponieważ w grupie wiekowej powyżej 45 lat odsetek donatorów wyniósł 14% w porównaniu z niższymi grupami wiekowymi (7% w grupie do 25. r.ż. i 11% w grupie 25-45 lat). Kolejnymi najsilniejszymi predyktorami okazały się wielkość gospodarstwa domowego oraz fakt wspomagania akcji charytatywnych w przeszłości. W ostatecznym kształcie drzewo podzieliło bazę na kilkadziesiąt segmentów, z których każdy miał

określone prawdopodobieństwo udzielenia donacji. Jednak ze względu na fakt, że część segmentów wygenerowanych przez drzewo była stosunkowo duża, w celu wygenerowania scoringu o bardziej ciągłej charakterystyce została przeprowadzona dodatkowa analiza z wykorzystaniem regresji logistycznej.



* W nawiasach podane są odsetki oznaczające procent penetracji grupy docelowej.

Rys. 1. Algorytm budowy drzewa decyzyjnego obrazujący cechy donatorów

Źródło: dane wewnętrzne SOS Wioski Dziecięce i Acxiom Polska.

Tabela 4. Segmenty wydzielone z pełnej bazy danych Acxiom posiadające najwyższe podobieństwo do profilu donatorów (największe prawdopodobieństwo pozytywnej reakcji)

Nr segmentu	Liczba osób w segmencie	Proc. osób o profilu donatora w segmencie	Liczebność skumulowana
1	1344	57,7%	1344
2	1373	36,9%	2717
3	1380	23,5%	4097
4	1372	22,6%	5469
5	1374	13,4%	6843
6	1378	12,6%	8221
7	1388	10,9%	9609
8	1388	7,6%	10 997
9	1387	6,6%	12 384
n

* Ramką zaznaczone rekomendowane segmenty.

Źródło: dane wewnętrzne SOS Wioski Dziecięce i Acxiom Polska.

Finalnie w wyniku analizy uzyskano bardzo szczegółowy podział bazy na wiele mikrosegmentów. Raport z analizy w postaci tabeli ze szczegółowymi informacjami o wielkości segmentu, penetracji grupy docelowej donatorów stanowił wygodne i łatwe w wykorzystaniu narzędzie, pozwalające decydom podjąć optymalne decyzje odnośnie do wyboru do właściwej akcji najlepszych adresatów (czyli segmentów mających największe prawdopodobieństwo pozytywnej reakcji na ofertę donacji) i określenia jej zasięgu (tabela 4).

Po teście bazy dobranej scoringowo do właściwej akcji marketingowej (wysyłki oferty donacji) wybrano już tylko te segmenty, których indeks penetracji osób z grupy uczącej był najwyższy. Przy czym były to osoby, do których wcześniej nie był kierowany komunikat testowy. Z bazy wyeliminowano także osoby, których dochody netto były poniżej 1000 zł na gospodarstwo domowe. Powstał pakiet *Data Pool*, tj. baza wyselekcjonowanych scoringiem osób, do których SOS WD miało roczną licencję z prawem wielokrotnej komunikacji marketingowej oraz dostępem do usług analitycznych.

4.3. Wyniki akcji

Wysyłka pocztowa z zachętą do przekazania na cele stowarzyszenia SOS Wioski Dziecięce przyniosła następujące wyniki:

- odzew na akcję na bazie scoringowej, liczony jako procent osób, które pozytywnie zareagowały na zachętę zawartą w wysyłce (*response rate*), w porównaniu z poprzednimi akcjami nieopartymi na analizie scoringowej, zwiększył się o 43%,
- zwiększyła się średnia wartość wpłaty poszczególnych donatorów o 83%,
- koszt dotarcia do potencjalnych donatorów zmniejszył się o 25%.

Wyżej wymienione wyniki w porównaniu z wcześniejszymi akcjami marketingowymi były niespodziewanie pozytywne i bardzo zadowalające dla decydentów marketingowych. Tak dobre wyniki akcji wynikały przede wszystkim z właściwie przeprowadzonej analizy klientów. Obecnie baza aktywnych donatorów SOS Wioski Dziecięce liczy kilkaset tysięcy osób. Każda akcja marketingowa jest poprzedzona wnikliwą analizą wyników akcji poprzednich, wyciągane są wnioski, robione analizy profilu i odzewu.

5. Zakończenie

Złożoność i dynamika zmian, jakie zachodzą w otoczeniu rynkowym, oraz związana z nimi zmienność w zachowaniach konsumentów coraz bardziej utrudniają badanie zachowań nabywców, a w szczególności przewidywanie ich reakcji na różne działania marketingowe. W tej sytuacji warto wykorzystywać pojawiające się i stale poszerzane możliwości analityczne oparte na nowoczesnych technologiach informatycznych (*Data Mining*). Pozwalają one na zestawianie ze sobą wielkich i zróżnicowanych zbiorów danych, a przez to odkrywanie niewidocznych na pierwszy rzut oka zależności i prawidłowości w zachowaniach klientów.

Analizy oparte na *Data Miningu* nie są czymś odkrywczym nowym, bo przecież wykorzystują znane od lat metody statystyki opisowej i wnioskowania statystycznego. To, co jest relatywnie nowe i czym warto się interesować, to poszerzane w szybkim tempie możliwości agregowania danych z różnych rozbudowanych zbiorów danych (*Big Data*), w tym baz własnych oraz kupowanych baz geodemograficznych i lifestylowych. Stosowanie analiz predykcyjnych wymaga jednak zwiększania świadomości oraz specjalistycznej wiedzy i kompetencji analitycznych wśród marketerów. Warto bowiem zaznaczyć, że analizy oparte na *Data Miningu* nie są w zupełności zautomatyzowane. Nadal definiowanie problemu, celu analizy i nadzór nad całością należą do analityka, co nie zmienia się na przestrzeni ostatnich kilku lat [Sokołowski 2005, s. 5-11]. Mają też, jak wiele innych podejść analitycznych, swoje ograniczenia. Niemniej jednak, jak pokazuje opisany tutaj przypadek SOS Wioski Dziecięcej, kierunek ten jest tym bardziej warty uwagi, gdyż modele predykcyjne, w tym scoringowe, dzięki badaniu prawdopodobieństwa reakcji nabywców, służą realnemu zgłębieniu wiedzy o nabywcach. Pomagają w odpowiedzi na tak ważne pytania, jak to, jakie produkty, kiedy i, przede wszystkim, komu oferować, jak planować różne akcje marketingowe. Analizy scoringowe wydają się ważnym uzupełnieniem innych metod i technik analizy zachowań nabywców.

Literatura

- Bird D., 2011, *Zdrowy rozsądek w marketingu bezpośrednim i interaktywnym*, MT Biznes, Warszawa.
- Chang H.-Ch., Hsiao-Ping T., 2011, *Group RFM analysis as a novel framework to discover better customer consumption behavior*, Expert Systems with Applications, vol. 38, issue 12, s. 14499-14513.
- Goldian G., 2002, *Baza danych jako narzędzie do budowania indywidualnego komunikatu do konsumenta*, prezentacja na konferencji Boomerang, SMB, Warszawa.
- Grzegorzczak M., 2007, *Problem pomiaru wartości klienta*, Problemy Zarządzania, nr 16, s. 119.
- Hand D., Mannila H., Smyth P., 2001, *Principles of Data Mining*, MIT Press, Cambridge.
- Hosseini Z.Z., Mohammadzadeh M., 2016, *Knowledge discovery from patients' behavior via clustering-classification algorithms based on weighted eRFM and CLV model: An empirical study in public health care services*, Iranian Journal of Pharmaceutical Research, 15 (1), s. 355-367.
- Hryniewicz K., 2016, *Modele scoringowe. Czym jest scoring i jakie ma zastosowanie?*, Metodolog.pl, marzec, 13.
- Hsieh N.-Ch., 2004, *An integrated data mining and behavioral scoring model for analyzing bank customers*, Expert Systems with Applications, vol. 27, issue 4, November, s. 623-633.
- Jaśko P., 2012, *Modele scoringu kredytowego z wykorzystaniem narzędzi data mining – analiza porównawcza*, Zastosowania metod statystycznych w badaniach naukowych IV, StatSoft Polska, Kraków.
- Kaczmarczyk S., 2007, *Zastosowanie badań marketingowych. Zarządzanie marketingowe i otoczenie przedsiębiorstwa*, PWE, Warszawa.
- Kędzior Z., 2005, *Badania rynku. Metody. Zastosowania*, PWE, Warszawa.
- Kieźel E., Smyczek S., 2015, *Zachowania konsumentów. Procesy unowocześnień konsumpcji*, Wolters Kluwer, Warszawa.
- Korol T., 2013, *Nowe podejście do analizy wskaźnikowej w przedsiębiorstwie*, Wolters Kluwer, Warszawa.

- Kozielski R. (red.), 2006, *Wskaźniki marketingowe*, wyd. II, Oficyna Ekonomiczna, Kraków.
- Kryśkiewicz Ł., 2017, *Analityka Big Data w Polsce. Jak polskie firmy wykorzystują #BigData, aby efektywnie konkurować na rynku*, Dziennik Internautów Biznes i Prawo, 17.08.
- Lee M.Y., Lee A.S., Sohn S.Y., 2013, *Behavior scoring model for coalition loyalty programs by using summary variables of transaction data*, Expert Systems with Applications, vol. 40, iss. 5, April, s. 1564-1570.
- Lewandowska M., 2010, *Tendencje zmian w makroocenieniu – perspektywa polskich przedsiębiorstw*, Zeszyty Naukowe. Szkoła Główna Handlowa. Kolegium Gospodarki Światowej, nr 27, s. 155-179.
- Malthouse E.C., 2010, *Accounting for the long-term effects of a marketing contact*, Expert Systems with Applications, vol. 37, iss. 7, July, s. 4935-4940.
- Maślankowski J., 2015, *Analiza jakości danych pozyskiwanych ze stron internetowych z wykorzystaniem rozwiązań Big Data*, Roczniki Kolegium Analiz Ekonomicznych SGH, nr 38, s. 167-177.
- Mazurek-Łopacińska K. (red.), 2016, *Badania marketingowe. Metody, techniki i obszary aplikacji na współczesnym rynku*, PWN, Warszawa.
- Miciuła I., Miciuła K., 2015, *Kluczowe trendy dla budowania biznesu w branży Big Data*, Zeszyty Naukowe Uniwersytetu Szczecińskiego, nr 863, Studia Informatica, nr 36, s. 51-63.
- Migut G., 2005, *Segmentacja użytkowników serwisu WWW z użyciem metod statystycznych i sieci neuronowych*, [w:] *Data Mining: poznaj siebie i swoich klientów*, Zeszyt z serii „Zastosowania Statystyki i Data Mining”, StatSoft Polska, Warszawa-Kraków.
- Mról B., 2013, *Konsument w globalnej gospodarce. Trzy perspektywy*, Oficyna Wydawnicza SGH, Warszawa.
- Paszyła A., 2005, *Przykład badania wzorców zachowań klientów za pomocą analizy koszykowej*, [w:] *Data Mining: poznaj siebie i swoich klientów*, Zeszyt z serii „Zastosowania Statystyki i Data Mining”, StatSoft Polska, Warszawa-Kraków.
- Saboo A.R., Kumar V., Park I., 2016, *Using Big Data to model time-varying effects for marketing resource (re)allocation*, Big Data & Analytics in Networked Business, MIS Quarterly, vol. 40, no. 4, s. 911-939.
- Sagan A., 2014, *Wprowadzenie do modelowania zjawisk społecznych i przykłady zastosowań w statystyce*, StatSoft Polska, DaneWiedzaSukces, Kraków.
- Sagan A., 2016, *Modelowanie marketingowe a paradygmat marketingu*, Handel Wewnętrzny, nr 5.
- Sojkin B., 2009, *Informacyjne podstawy decyzji marketingowych*, PWE, Warszawa.
- Sokołowski A., 2005, *Data Mining – automat czy metoda naukowa?*, [w:] *Data Mining: poznaj siebie i swoich klientów*, Zeszyt z serii „Zastosowania Statystyki i Data Mining”, StatSoft Polska, Warszawa-Kraków.
- Światowy G., 2016, *Pomiar zmiennych jakościowych*, [w:] Mazurek-Łopacińska K. (red.), *Badania marketingowe. Metody, techniki i obszary aplikacji na współczesnym rynku*, PWN, Warszawa.
- Trojanowski M., 2010, *Marketing bezpośredni. Koncepcja – zarządzanie – instrumenty*, PWE, Warszawa.
- Venkatesan R., Kumar V., 2004, *A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy*, Journal of Marketing, vol. 68.
- Weinert A., 2017, *Wykorzystanie systemów BIG DATA w procesie wyborów strategicznych*, Ekonomika i Organizacja Przedsiębiorstwa, nr 2 (805), s. 48-58.
- Woś J., Rachocka J., Kasperek-Hoppe M., 2011, *Zachowania konsumentów. Teoria i praktyka*, wyd. II, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.
- Wysiński P., 2013, *Zastosowanie scoringu kredytowego w zarządzaniu ryzykiem kredytowym*, International Business and Global Economy, Instytut Handlu Zagranicznego Uniwersytetu Gdańskiego, nr 32, s. 253-268.