

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 385

Taksonomia 25

**Klasyfikacja i analiza danych –
teoria i zastosowania**

Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2015

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp.....	9
Tomasz Bartłomowicz: Segmentacja konsumentów na podstawie preferencji wyrażonych uzyskanych metodą Maximum Difference Scaling	11
Barbara Batóg, Jacek Batóg, Andrzej Niemiec, Wanda Skoczylas, Piotr Waśniewski: Zastosowanie metod klasyfikacyjnych w identyfikacji kluczowych indyktorów osiągnięć w zarządzaniu wynikami przedsiębiorstw	20
Iwona Bąk: Wykorzystanie statystycznej analizy danych w badaniach turystyki transgranicznej na obszarach chronionych.....	28
Beata Bieszk-Stolorz: Ocena stopnia deprecjacji kapitału ludzkiego z wykorzystaniem nieliniowych modeli regresji.....	37
Mariola Chrzanowska, Nina Drejerska: Małe i średnie przedsiębiorstwa w strefie podmiejskiej Warszawy – określenie znaczenia lokalizacji z wykorzystaniem drzew klasyfikacyjnych.....	45
Adam Depta: Próba modelowania strukturalnego jakości życia osób jękaających się jako konstrukt ukrytego na podstawie kwestionariusza SF-36v2	53
Katarzyna Dębkowska: Wielowymiarowa analiza kondycji finansowej przedsiębiorstw sektora e-usług	63
Krzysztof Dmytrów, Mariusz Doszyń: Taksonomiczna procedura wspomagania kompletacji produktów w magazynie	71
Mariusz Doszyń, Sebastian Gnat: Propozycja procedury taksonomiczno-ekonometrycznej w indywidualnej wycenie nieruchomości.....	81
Marta Dziechciarz-Duda, Anna Król: Zastosowanie analizy <i>unfolding</i> i regresji hedonicznej do oceny preferencji konsumentów	90
Katarzyna Frodyma: Współzależność między poziomem rozwoju gospodarczego a udziałem energii ze źródeł odnawialnych w końcowym zużyciu w krajach Unii Europejskiej.....	99
Hanna Gruchociak: Porównanie struktury lokalnych rynków pracy wyznaczonych przy wykorzystaniu różnych metod w Polsce w latach 2006 i 2011 .	111
Alicja Grześkowiak, Agnieszka Stanimir: Postrzeganie środowiska pracy przez starszą i młodszą generację pracowników	120
Marta Hozer-Koćmiel, Christian Lis: Klasyfikacja krajów nadbałtyckich ze względu na czas prac wykonywanych w gospodarstwie domowym	129
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel: Zegar cyklu koniunkturalnego państw UE i USA w latach 1995-2013 w świetle badań synchronizacji.....	138
Aleksandra Łuczak: Wykorzystanie rozszerzonej interwałowej metody TOPSIS do porządkowania liniowego obiektów	147

Aleksandra Łuczak, Feliks Wysocki: Zintegrowane podejście do ustalania współczynników wagowych dla cech w zagadnieniach porządkowania linowego obiektów	156
Małgorzata Markowska, Danuta Strahl: Wykorzystanie klasyfikacji dynamicznej do identyfikacji wrażliwości na kryzys ekonomiczny unijnych regionów szczebla NUTS 2.....	166
Aleksandra Matuszewska-Janica, Marta Hozer-Koćmiel: Struktura zatrudnienia oraz wynagrodzenia kobiet i mężczyzn a przedmiotowa struktura gospodarcza w państwach UE.....	178
Anna M. Olszewska: Zastosowanie analizy korespondencji do badania związku pomiędzy zarządzaniem jakością a innowacyjnością przedsiębiorstw	187
Małgorzata Podogrodzka: Metoda aglomeracyjna w ocenie przestrzennego zróżnicowania starości demograficznej w Polsce	195
Ewa Roszkowska, Tomasz Wachowicz: Ocena ofert negocjacyjnych spoza dopuszczalnej przestrzeni negocjacyjnej.....	201
Ewa Roszkowska, Tomasz Wachowicz: Zastosowanie metody <i>unfolding</i> do wspomagania procesu negocjacji	210
Małgorzata Rószkiewicz: Próba diagnozy uwarunkowań poziomu wskaźnika braku odpowiedzi w środowisku polskich gospodarstw domowych.....	219
Marcin Salamaga: Próba identyfikacji muzycznych profili melomanów z wykorzystaniem drzew klasyfikacyjnych i regresyjnych	229
Agnieszka Sompolska-Rzechuła: Określenie czynników wpływających na prawdopodobieństwo poprawy poziomu rozwoju społecznego z wykorzystaniem modelu logitowego	239
Iwona Staniec: Wykorzystanie analizy czynnikowej w identyfikacji konstruktywów ukrytych determinujących ryzyko współpracy.....	248
Agnieszka Stanimir: Skłonność do zagranicznej mobilności młodszych i starszych osób	257
Mirosława Sztemberg-Lewandowska: Problemy decyzyjne w funkcjonalnej analizie głównych składowych.....	267
Tomasz Szubert: Demograficzno-społeczne determinanty określające subiektywny status jednostki w polskim społeczeństwie	276
Piotr Tarka: Własności 5- i 7-stopniowej skali Likerta w kontekście normalizacji zmiennych metodą Kaufmana i Rousseeuwa	286
Joanna Trzęsiok: Nielklasyczne metody regresji a problem odporności	296
Katarzyna Wawrzyniak: Ocena podobieństwa wyników uporządkowania województw uzyskanych różnymi metodami porządkowania	305
Katarzyna Wójcik, Janusz Tuchowski: Wykorzystanie metody opartej na wzorcach w automatycznej analizie opinii konsumenckich.....	314
Anna Zamojska: Zastosowanie analizy falkowej w ocenie efektywności funduszy inwestycyjnych	325

Summaries

Tomasz Bartłomowicz: Segmentation of consumers based on revealed preferences obtained with the Maximum Difference Scaling method	19
Barbara Batóg, Jacek Batóg, Andrzej Niemiec, Wanda Skoczylas, Piotr Waśniewski: Application of classification methods to identify the key performance indicators of performance management	27
Iwona Bąk: The application of statistical data analysis in the studies of cross-border tourism in protected areas.....	36
Beata Bieszk-Stolorz: Evaluating human capital depreciation by means of non-linear regression models.....	44
Mariola Chrzanowska, Nina Drejerska: Small and medium enterprises in the Warsaw suburban zone – determination of a localization’s role using classification trees	52
Adam Depta: An attempt of structural modelling of the quality of life of stuttering people as a latent construct, based on SF-36v2 questionnaire ...	62
Katarzyna Dębkowska: Multidimensional analysis of financial condition of e-business services	70
Krzysztof Dmytrów, Mariusz Doszyń: Taxonomic procedure of supporting order-picking of products in a warehouse	80
Mariusz Doszyń, Sebastian Gnat: Taxonomic and econometric methods in individual real estate evaluation.....	89
Marta Dziechciarz-Duda, Anna Król: The application of unfolding analysis and hedonic regression in the investigation of consumers’ preferences	98
Katarzyna Frodyma: Interdependence between the level of economic development and the share of renewable energy in gross final energy consumption in the European Union.....	110
Hanna Gruchociak: Comparison of local labour markets structure designated using different methods in Poland in 2006 and 2011 years.....	119
Alicja Grześkowiak, Agnieszka Stanimir: Perception of working environment by older and younger generation of workers.....	128
Marta Hozer-Koćmiel, Christian Lis: Classification of the Baltic Sea Region countries due to the time of household work.....	137
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel: Business cycle clock for the EU and the USA in 1995-2013 in the light of synchronization research.....	146
Aleksandra Łuczak: The use of the extended interval TOPSIS methods for linear ordering of objects.....	155
Aleksandra Łuczak, Feliks Wysocki: Integrated approach for determining the weighting coefficients for features in issues of linear ordering of objects.....	165

Małgorzata Markowska, Danuta Strahl: The application of dynamic classification for the identification of vulnerability to economic crisis in the EU NUTS 2 regions	177
Aleksandra Matuszewska-Janica, Marta Hozer-Koćmiel: The structure of male and female employment and remuneration vs. the basic economy structure in the EU countries	186
Anna M. Olszewska: The application of the correspondence analysis for the study of the relations between quality management and innovation in the enterprises.....	194
Małgorzata Podogrodzka: Agglomeration method in the age and ageing in Poland by voivodships.....	200
Ewa Roszkowska, Tomasz Wachowicz: Scoring the negotiation offers from the outside of the feasible negotiation space	209
Ewa Roszkowska, Tomasz Wachowicz: Application of the unfolding analysis to negotiation support.....	218
Małgorzata Rószkiewicz: An attempt to diagnose the determinants of non-response rate in Polish households surveys	228
Marcin Salamaga: Attempt to identify music lovers profiles using classification and regression trees	238
Agnieszka Sompolska-Rzechuła: The definition of factors influencing the probability of improving the level of human development using the logit model.....	247
Iwona Staniec: The use of factor analysis to identify hidden constructs – determinants of the cooperation risk	256
Agnieszka Stanimir: Willingness to mobility abroad among younger and older persons	266
Mirosława Sztemberg-Lewandowska: Decision problems in functional principal components analysis.....	275
Tomasz Szubert: Socio-demographic factors determining subjective social status of an individual in Polish society	285
Piotr Tarka: Normalization methods of variables and measurement on 5 and 7 point Likert scale	295
Joanna Trzęsiok: Non-classical regression methods vs. robustness	304
Katarzyna Wawrzyniak: The evaluation of the similarity of the voivodships' orderings obtained by means of different methods.....	313
Katarzyna Wójcik, Janusz Tuchowski: Using pattern-based opinion mining.....	324
Anna Zamojska: Mutual funds performance measurement – wavelets analysis approach.....	333

Marcin Salamaga

Uniwersytet Ekonomiczny w Krakowie

e-mail: salamaga@uek.krakow.pl

PRÓBA IDENTYFIKACJI MUZYCZNYCH PROFILI MELOMANÓW Z WYKORZYSTANIEM DRZEW KLASYFIKACYJNYCH I REGRESYJNYCH

Streszczenie: Celem artykułu jest segmentacja miłośników muzyki poważnej pod względem wysokości dochodów, które są oni skłonni przeznaczyć na kulturę muzyczną, oraz pod względem preferowanych gatunków muzycznych czy twórców muzyki. Do klasyfikacji melomanów zastosowano drzewa regresyjne i klasyfikacyjne tworzone zgodnie z procedurami CART i CHAID, opierając się na danych pochodzących z badania ankietowego przeprowadzonego wśród odbiorców muzyki poważnej. Narzędzia te umożliwiły stworzenie muzycznych profili melomanów, które pomogą w lepszym poznaniu zapotrzebowania na produkty muzyczne w różnych grupach odbiorców muzyki.

Słowa kluczowe: drzewa klasyfikacyjne, drzewa regresyjne, algorytm CART, algorytm CHAID, muzyka.

DOI: 10.15611/pn.2015.385.25

1. Wstęp

Muzyka poważna uchodzi za sztukę dostępną raczej dla wąskiego grona odbiorców. Jej elitarność można wyjaśnić niedostateczną promocją tego rodzaju muzyki w środkach masowego przekazu, jak również koniecznością poniesienia pewnego wysiłku intelektualnego koniecznego do zrozumienia wartości artystycznej dzieła muzycznego. Wśród osób, które podejmą ten wysiłek, muzyka staje się źródłem głębokich doznań estetycznych. Ponadto, jak mawiał znany publicysta muzyczny Jerzy Waldorff: „muzyka łagodzi obyczaje” [Waldorff 1982]. W dobie powszechnej komercjalizacji również muzyka poważna do pewnego stopnia stała się „produktem” podlegającym ekonomicznym prawom rynku. Jego uczestnikami są m.in. wydawnictwa muzyczne, firmy fonograficzne, studia nagrań, środki masowego przekazu, agencje muzyczne, impresariaty artystyczne i oczywiście słuchacze, melomani reprezentujący popytową stronę rynku. Aby przygotować optymalny

„produkt muzyczny”, którego nośnikiem może być płyta kompaktowa, program muzyczny, plik internetowy, jak również koncert na żywo, konieczne jest zbadanie rynku muzycznego. Celem takiej analizy jest m.in. poznanie gustów i preferencji muzycznych melomanów czy szerzej – poznanie ich kultury muzycznej. Empiryczne badania upodobań muzycznych są od dawna podejmowane w literaturze psychologicznej, socjologicznej, rzadziej marketingowej (por. [Hargreaves, North 1997; Kamińska 2002; Ginocchio 2006; Białkowski, Grusiewicz 2009]). Niniejsze opracowanie również wpisuje się w nurt badań nad preferencjami i gustami muzycznymi osób słuchających muzyki. Celem artykułu jest wyodrębnienie profili respondentów o zbliżonych gustach muzycznych ze względu na preferowany sposób odbierania muzyki oraz wysokość dochodów, które są oni skłonni przeznaczyć na kulturę muzyczną¹. Do klasyfikacji melomanów zastosowano drzewa klasyfikacyjne i regresyjne tworzone zgodnie z procedurami CART i CHAID, opierając się nadanych pochodzących z internetowego badania ankietowego przeprowadzonego wśród odbiorców muzyki poważnej (użytkowników internetowych serwisów muzycznych). Za zastosowaniem tych właśnie narzędzi przemawia m.in. swoboda w doborze cech ilościowych i jakościowych bez konieczności kodowania kategorii tych ostatnich, uchylenie założeń co normalności rozkładu zmiennych, a także przejrzystość interpretacji wyników i możliwość ich graficznej wizualizacji. Przedmiotowe procedury od dawna są stosowane m.in. w naukach ekonomicznych [Gatnar 2001; Bai, Yen, Yang 2008; Łapczyński 2010], rolniczych [Ferraro, Rive-ro, Ghersa 2009; Ślaska-Grzywna 2010], technicznych [Ma i in. 2013] czy medycznych [Berney i in. 2011; Zahir i in. 2013; Berney i in. 2011]. Również w muzykologii można znaleźć nieliczne przykłady wykorzystania drzew dyskryminacyjnych w klasyfikacji stylów, gatunków muzycznych itp. (por. [Abundo 2009; Abeßer, Lukashevich, Bräuer 2012]). Deficytowe są natomiast publikacje naukowe, w których wykorzystuje się drzewa klasyfikacyjne i regresyjne w segmentacjach rynku muzycznego. Zatem proponowane podejście badawcze poszerza dotychczasowy wachlarz metod analizy rynku muzycznego i w tym sensie jest oryginalne.

Dzięki utworzeniu muzycznych profili melomanów możliwe jest lepsze zdiagnozowanie zapotrzebowania na produkty muzyczne w różnych grupach odbiorców muzyki. Wyniki takiego badania mogą więc być wykorzystane w optymalizowaniu oferty muzycznej sal koncertowych, teatrów operowych, wydawnictw fonograficznych czy muzycznych programów w radiu i telewizji.

¹ Przez wydatki na kulturę muzyczną rozumiano wydatki na zakup biletów wstępu do filharmonii lub teatru operowego, płyt muzycznych CD, DVD, literatury muzycznej, jak również wydatki na kształcenie muzyczne.

2. Metoda badania

Segmentację respondentów przeprowadzono, stosując drzewa klasyfikacyjne i regresyjne (CART– *Classification and Regression Trees*) [Breiman i in. 1984] oraz automatyczną detekcję interakcji metodą chi-kwadrat (CHAID – *Chi-square Automatic Interaction Detection*) [Kass 1983]. Obie metody generują drzewa decyzyjne, czyli grafy o strukturze drzewiastej. W konstrukcji takich drzew można wydzielić węzeł początkowy (korzeń), węzły podlegające podziałowi (macierzyste) na węzły potomne oraz węzły końcowe (liście). W przypadku drzew klasyfikacyjnych zmienna zależna jest mierzona na słabych skalach (nominalnej lub porządkowej), a drzewa regresyjne mają zmienną zależną mierzona na mocnej skali (co najmniej przedziałowej). Za pomocą drzew klasyfikacyjnych oraz regresyjnych dokonywany jest podział wielowymiarowej przestrzeni zmiennych na S rozłącznych segmentów zgodnie z funkcją klasyfikacyjną [Gatnar 2009]:

$$f(x_i) = \sum_{s=1}^S \alpha_s I(x_i \in R_s), \quad (1)$$

gdzie: R_s ($s = 1, 2, \dots, S$) – podprzestrzeń przestrzeni X^m , α_s – parametry modelu, I – funkcja wskaźnikowa przyjmująca wartość 1 (gdy $x_i \in R_s$) lub 0 (gdy $x_i \notin R_s$).

Jeśli Y jest nominalną zmienną, to:

$$\alpha_s = \arg \max_t p(P_t | x_i \in R_s), \quad (2)$$

gdzie P_s ($s = 1, 2, \dots, u$) oznacza klasę, do której należy obserwacja x_i .

Jeśli natomiast zmienna Y jest mierzona na skali co najmniej przedziałowej, to parametr α_s jest przeciętną wartością zmiennej Y dla obserwacji znajdujących się w podprzestrzeni R_s :

$$\alpha_s = \frac{1}{n_s} \sum_{x_i \in R_s} y_i, \quad (3)$$

gdzie n_s – liczba obserwacji w podprzestrzeni R_s .

W celu oceny jakości podziału przestrzeni R na rozłączne segmenty R_1, R_2, \dots, R_S stosuje się funkcję postaci [Gatnar 2009]:

$$\Delta Q(R) = Q(R) - \sum_{s=1}^S Q(R_s) p_s, \quad (4)$$

gdzie: $Q(R_s)$ – funkcja zróżnicowania obserwacji znajdujących się w podprzestrzeni

R_s ,
 p_s – frakcja obserwacji w podprzestrzeni R_s .

Za najlepszy podział przestrzeni zmiennych uznawany jest taki, który zapewni największą jednorodność otrzymanych podprzestrzeni, co jest możliwe przy maksymalizacji wyrażenia (4). Postać funkcji $Q(R_s)$ zależy od rodzaju tworzonego drzewa dyskryminacyjnego. W przypadku, gdy budowane jest drzewo regresyjne, najczęściej stosowana jest funkcja kwadratowa [Gatnar 2009]:

$$Q(R_s) = \frac{1}{n_s} \sum_{x_i \in R_s} (y_i - \alpha_s)^2, \quad (5)$$

gdzie α_s jest parametrem obliczanym zgodnie ze wzorem (3).

Jeśli budowane jest drzewo klasyfikacyjne, to do oceny homogeniczności podprzestrzeni R_s stosuje się przeważnie błąd klasyfikacji, wskaźnik Giniego lub miarę entropii. W procesie budowy drzew dyskryminacyjnych każda ze zmiennych merytorycznych jest poddawana dyskretyzacji [Gatnar 2009], w wyniku czego jej wartości dzielone są na dwa rozłączne przedziały, po czym wybierana jest ta zmienna, która maksymalizuje kryterium (4). W efekcie z każdego węzła drzewa dyskryminacyjnego wychodzą dwa węzły potomne. W sytuacji, gdy zmienna X jest mierzona na skali słabej, to zbiór jej kategorii jest dzielony na takie dwa podzbiory, aby wartość wyrażenia (4) była maksymalna². W konstrukcji drzew klasyfikacyjnych i regresyjnych niepożądanym zjawiskiem jest nadmierna złożoność modelu, któremu towarzyszy wzrost wartości błędu dla zbioru testowego. Jednym ze sposobów przeciwdziałania temu efektowi jest przycinanie krawędzi (*pruning*), co prowadzi do redukcji rozmiarów drzewa. Przycinanie krawędzi jest procesem sekwencyjnym i polega na eliminacji tych gałęzi drzewa, których brak nie powoduje zbyt dużego pogorszenia jakości podziałów (kontrolowanej za pomocą kryterium (4)). W efekcie powstaje zbiór drzew, z których wybierane jest to, dla którego błąd klasyfikacji dla zbioru testowego³ jest najmniejszy [Gatnar 2001].

W metodzie CART o ostatecznym kształcie drzewa, jego głębokości decyduje przebieg krzywych współczynnika błędu obliczonego na zbiorze uczącym i współczynnika błędu obliczonego na zbiorze testowym [Łapczyński 2010]. Interpretacja drzewa dyskryminacyjnego CART odbywa się zgodnie z regułami klasyfikacyjnymi stworzonymi dla każdej ścieżki drzewa łączącej korzeń z liściem. Wybór analizy CART do klasyfikacji melomanów wynikał przede wszystkim z jej efektywności w predykcji jakościowej, w przypadku której stosowalność wielu metod statystycznych i ekonometrycznych jest ograniczona. Dodatkowo analiza CART jest odporna na występowanie jednostek nietypowych i dopuszcza wykorzystanie zmiennych charakteryzujących się brakami danych [Gatnar 2001].

² W metodzie CART wydzielane są dwa przedziały wartości (bądź podzbiory kategorii) dla zmiennych X , w wyniku czego powstaje drzewo binarne. Podział na większą liczbę przedziałów (podzbiorów kategorii) skutkuje drzewami niebinarnymi.

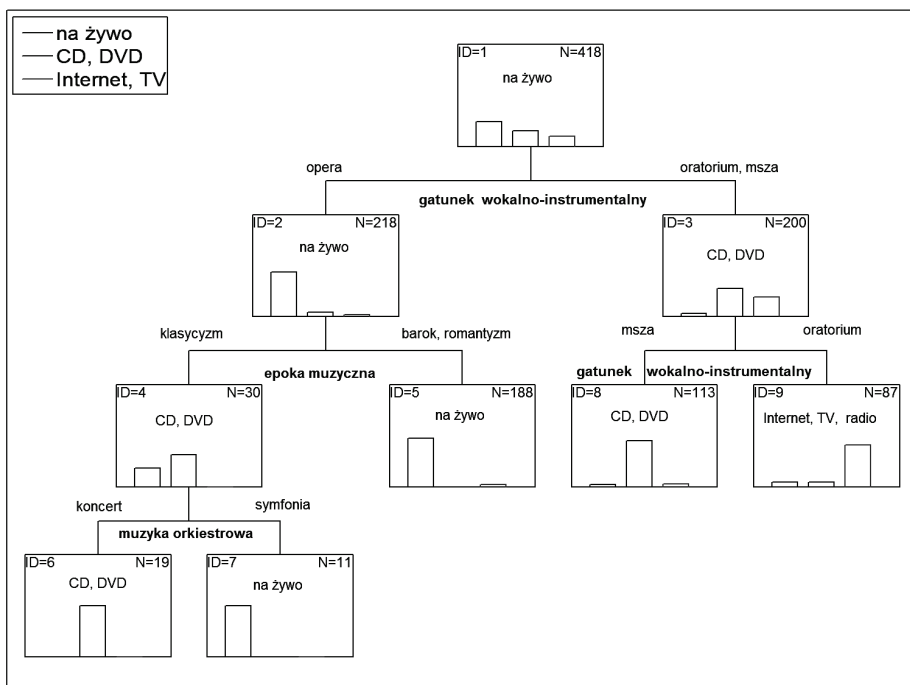
³ Ewentualnie błąd klasyfikacji wyznaczany za pomocą walidacji krzyżowej.

Do konstrukcji drzewa regresyjnego zastosowano algorytm CHAID, który w przeciwieństwie do procedury CART może tworzyć drzewa z większą liczbą węzłów potomków niż 2 oraz nie dokonuje przycinania drzewa. W analizie CHAID następuje stopniowe łączenie podzbiorów kategorii zmiennej X , przy czym procesem tym steruje test χ^2 (dla zmiennych nominalnych) lub test F (dla zmiennych metrycznych). Algorytm sprawdza wszystkie pary kategorii zmiennej X (predyktora) pod kątem różnicowania zmiennej zależnej Y . Pary kategorii nieistotnie różniące tę zmienną łączone są w jedną kategorię, po czym algorytm poszukuje kolejnej pary itd. Proces kończy się, kiedy kolejne podziały nie są możliwe (np. ze względu na powstanie zbyt małych grup w stosunku do zadanej wstępnie liczebności). Analiza kolejnych podziałów pozwala na znalezienie zmiennych najsilniej różnicujących zmienną zależną i umożliwia utworzenie ich rankingu. W węzłach regresyjnego drzewa CHAID podawana jest wartość średnia i wariancja (odchylenie standardowe) zmiennej zależnej. W interpretacji liści drzewa identyfikuje się kombinacje predyktorów warunkujących występowanie określonej wartości zmiennej zależnej. Do zalet przedmiotowej metody należą: możliwość uwzględnienia predyktorów jakościowych oraz brak konieczności ich kodowania (np. na zmienne zero-jedynkowe), brak wymogów co do liniowości związków między zmiennymi oraz normalności ich rozkładów [Gatnar 2001].

W analizach CART i CHAID wykorzystano dane pochodzące z ankiet, które w okresie od stycznia do czerwca 2014 r. rozsyłano drogą elektroniczną do ok. 780 respondentów – użytkowników muzycznych serwisów internetowych, uczestników forów dyskusyjnych poświęconych muzyce, bywalców sal koncertowych i teatrów operowych. Zwrot ankiet był na poziomie 55%, a po ostatecznej selekcji nadesłanych odpowiedzi w analizie uwzględniono 418 poprawnie wypełnionych ankiet. Pytania zawarte w elektronicznym kwestionariuszu umożliwiły m.in. poznanie preferencji muzycznych melomanów, ich estetyki muzycznej i kultury muzycznej. Respondentów pytano o: preferowaną formę odbioru muzyki (warianty odpowiedzi: *na żywo, płyty audio CD i DVD, Internet, TV*) ulubioną epokę muzyczną (*barok, klasycyzm, romantyzm*), preferowany gatunek muzyki wokalnoinstrumentalnej (*msza, opera, oratorium*), preferowany gatunek muzyki orkiestrowej (*koncert, symfonia*), ulubionego kompozytora epoki baroku (*Bach, Haendel, Vivaldi*), klasycyzmu (*Beethoven, Haydn, Mozart*), romantyzmu (*Brahms, Chopin, Schumann*), XX w. (*Bartok, Debussy, Strawiński*), ulubionego twórcę oper (*Puccini, Wagner, Verdi*), preferowany nurt muzyki XX w. (*impresjonizm, neoklasycyzm*), preferowanie muzyki programowej (*tak, nie*), wysokość miesięcznych wydatków na kulturę muzyczną. Wszystkie z wymienionych cech mierzone były na słabej skali pomiaru (skala nominalna) z wyjątkiem cechy *wydatki na kulturę muzyczną*, którą zmierzono na skali metrycznej (skala ilorazowa). Ze względu na cel badania i charakter zmiennych zastosowanie procedury CART i CHAID w profilowaniu melomanów wydaje się uzasadnione. Przedmiotowe analizy przeprowadzono w pakiecie komputerowym Statistica.

3. Wyniki badań

Poziom oraz charakter kultury muzycznej społeczeństwa jest wypadkową m.in. jego stosunku do muzyki, wrażliwości muzycznej, sposobu postrzegania muzyki i preferencji muzycznych. Efekt syntezy tych elementów znajduje odzwierciedlenie m.in. w sposobie odbioru muzyki przez melomanów. W związku z tym w pierwszej kolejności przeprowadzono segmentację melomanów metodą CART ze względu na zmienną zależną *preferowana forma odbioru muzyki*. Wyniki w postaci drzewa klasyfikacyjnego przedstawiono na rys. 1. Przedmiotowe drzewo ma 5 węzłów końcowych i 4 węzły dzielone. Powstałe w wyniku podziału metodą CART homogeniczne podzbiory mogą być traktowane jako segmenty melomanów, którzy preferują podobny sposób odbioru muzyki (na żywo, z płyt audio CD, DVD lub za pośrednictwem Internetu, telewizji, radia) przy określonych kombinacjach kategorii cech reprezentujących gust muzyczny melomanów, ich estetykę muzyczną, wrażliwość muzyczną itp. W interpretacji wyników analizy CART skupiono się na charakterystyce wybranych segmentów melomanów wyznaczonych przez węzły końcowe drzewa klasyfikacyjnego.



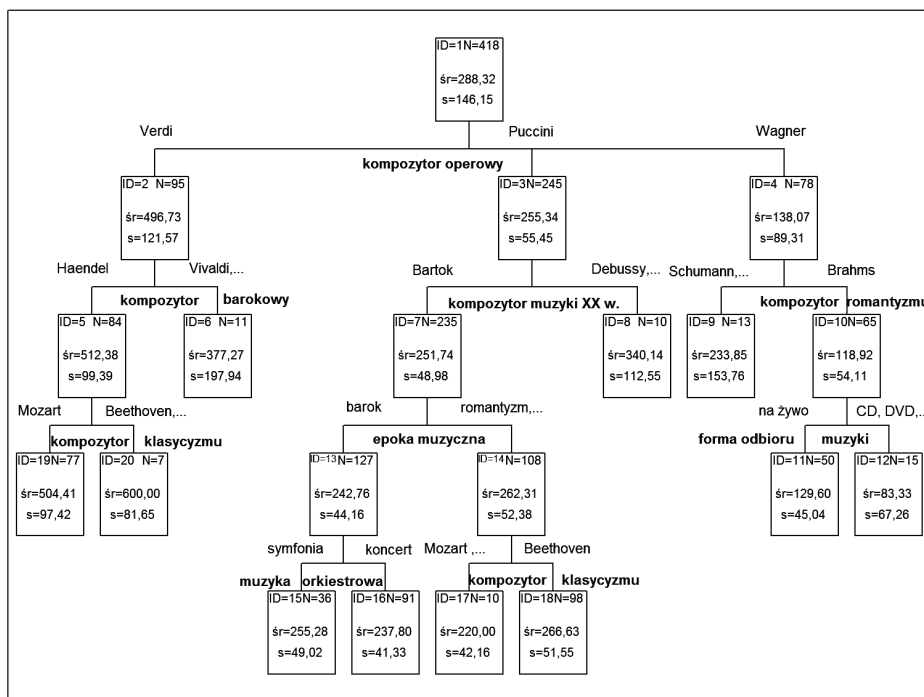
Rys. 1. Drzewo klasyfikacyjne CART dla zmiennej zależnej *preferowana forma odbioru muzyki*

Źródło: opracowanie własne na podstawie wyników ankiety internetowej przeprowadzonej w okresie 2.01.2014-30.06 2014.

Analizując kolejne poziomy drzewa CART, począwszy od korzenia po liście końcowe, można sformułować następujące reguły:

- jeżeli respondent spośród gatunków muzyki wokально-instrumentalnej preferuje operę i jednocześnie lubi muzykę epoki barokowej lub epoki romantycznej (węzeł nr 5), to najprawdopodobniej słucha muzyki poważnej na żywo (z prawdopodobieństwem 0,952),
- jeżeli respondent spośród gatunków muzyki wokально-instrumentalnej preferuje operę i jednocześnie gustuje w muzyce epoki klasycyzmu, a spośród gatunków muzyki orkiestrowej najbardziej ceni koncert (węzeł nr 6), to najprawdopodobniej słucha muzyki poważnej z płyt audio CD i DVD (z prawdopodobieństwem bliskim 1,000),
- jeżeli respondent spośród gatunków muzyki wokально-instrumentalnej najbardziej ceni mszę (węzeł nr 8), to najprawdopodobniej słucha muzyki poważnej z płyt audio CD i DVD (z prawdopodobieństwem 0,903).

Odsetki trafnie zakwalifikowanych melomanów do grup preferujących słuchanie muzyki na żywo, z płyt audio CD, DVD oraz za pośrednictwem Internetu, telewizji



Rys. 2. Drzewo regresyjne CHAID dla zmiennej zależnej wysokość miesięcznych wydatków na kulturę muzyczną

Źródło: opracowanie własne na podstawie wyników ankiety internetowej przeprowadzonej w okresie 2.01.2014-30.06 2014.

i radia wyniosły odpowiednio 93,60%, 93,80% i 82,56%. Na podstawie badania siły poszczególnych predyktorów w różnicowaniu segmentów melomanów w ramach przedmiotowej analizy CART stwierdzono, że największy wpływ na formę odbioru muzyki ma preferowany gatunek muzyki orkiestrowej, a w dalszej kolejności preferowany gatunek wokalnoinstrumentalny i ulubiona epoka muzyczna.

Do klasyfikacji melomanów pod względem wysokości dochodów, które są oni skłonni przeznaczyć na kulturę muzyczną (cecha mierzona na silnej skali metrycznej), zastosowano drzewo regresyjne tworzone zgodnie z procedurą CHAID (rys. 2).

Drzewo przedstawione na rys. 2 ma 11 węzłów końcowych i 9 węzłów dzielonych. Ankietowana grupa respondentów została podzielona na jak najbardziej jednorodną grupę pod względem zmiennej zależnej i zarazem możliwie najbardziej różniące się między sobą. Zmienną najsilniej różnicującą respondentów pod względem wysokości miesięcznych wydatków na kulturę muzyczną jest *ulubiony kompozytor operowy*. Interpretując wyniki analizy CHAID, skupiono się na wybranych liściach drzewa regresyjnego (z większą liczbą respondentów). Oto niektóre reguły wynikające z przedmiotowej analizy:

- jeżeli respondenci gustują w operach Pucciniego, spośród kompozytorów XX w. preferują muzykę Bartoka, najbardziej cenią muzykę barokową, a w muzyce orkiestrowej przedkładają koncert ponad symfonię, to ich wydatki na kulturę muzyczną mają rozkład ze średnią 237,80 zł i odchyleniem standardowym 41,33 zł,
- jeżeli melomani preferują opery Pucciniego, chętniej słuchają muzyki Bartoka niż Debussy'ego czy Strawińskiego, najbardziej cenią muzykę epoki romantyzmu, a spośród kompozytorów klasycyzmu wybierają Beethovena, to ich wydatki na kulturę muzyczną mają rozkład ze średnią 266,63 zł i odchyleniem standardowym 51,55 zł,
- jeżeli respondenci chętniej słuchają oper Wagnera niż Verdiego i Pucciniego, spośród kompozytorów epoki romantycznej najbardziej preferują Brahmsa, a muzyki najchętniej słuchają żywo, to ich wydatki na kulturę muzyczną mają rozkład ze średnią 129,60 zł i odchyleniem standardowym 45,04 zł. Do oceny stopnia homogeniczności powstałych węzłów i tym samym trafności predykcji wydatków na kulturę muzyczną można wykorzystać współczynnik zmienności. Najniższą wartość przedmiotowy współczynnik osiągnął w końcowym węźle nr 16 (ok. 17,38%), a najwyższą w węźle nr 9 (ok. 65,75%). Zatem najwyższą precyzję predykcji wydatków osiągnięto w grupie melomanów, którzy jednocześnie preferują muzykę Pucciniego, Bartoka, muzykę epoki baroku i chętniej słuchają koncertu niż symfonii. Z kolei najsłabszą dokładność predykcji wydatków na kulturę muzyczną można zaobserwować w grupie respondentów, którzy gustują w operach Wagnera oraz preferują muzykę Schumanna i Chopina.

4. Zakończenie

Wyniki badań w niniejszym opracowaniu potwierdziły przydatność drzew regresyjnych i klasyfikacyjnych w segmentacji miłośników muzyki poważnej. Dowiedziono, że melomani są w znacznym stopniu zróżnicowani pod względem sposobu odbierania muzyki, upodobań muzycznych i wrażliwości muzycznej. W szczególności w przedmiotowej analizie udało się połączyć aspekt ekonomiczny (wydatki na kulturę muzyczną) z trudno mierzalnymi kategoriami kształtującymi kulturę muzyczną. Dzięki temu wyodrębniono „ekonomiczno-muzyczne” profile respondentów i wskazano najbardziej istotne czynniki segmentacji. Pozwoliło to na wyrażenie kombinacji gustów i preferencji muzycznych w kwotach pieniężnych, które melomani są skłonni przeznaczyć na zaspokojenie swoich potrzeb muzycznych. Przedstawione wyniki mogą więc stanowić wsparcie w identyfikacji zapotrzebowania melomanów na produkty muzyczne w postaci koncertu symfonicznego, przedstawienia operowego, płyty audio CD, DVD czy audycji muzycznej w Internecie, telewizji lub radiu. Predykcja kwoty dochodów przeznaczanych na ich zakup stanowi cenną informację o sile ekonomicznej wyróżnionych profili i pomoże w dostosowaniu wachlarza produktów muzycznych do oczekiwań konkretnych grup melomanów. W szczególności wyniki badania mogą pomóc w optymalizowaniu repertuarów filharmonii, teatrów operowych, oferty wydawnictw muzycznych, fonograficznych, muzycznej oferty telewizyjnej i radiowej. Wydaje się, że przedmiotowe badanie może być interesujące zarówno dla analityków rynku muzycznego, menedżerów firm wydawnictw muzycznych, fonograficznych, dyrektorów filharmonii, teatrów operowych, jak również psychologów, socjologów oraz osób związanych ze środowiskiem muzycznym. Ponieważ badanie ankietowe było przeprowadzone na próbie osób, które korzystają z Internetu (ankieta internetowa), więc w uogólnianiu wyników należy zachować ostrożność. Z tego też względu badanie w przyszłości powinno być powtarzane być może na liczniejszej próbie melomanów (również tych, którzy nie korzystają z Internetu), co pozwoli zaktualizować wyniki w zakresie kształtowania się ich gustów, preferencji muzycznych i szeroko pojętej kultury muzycznej.

Literatura

- Abeßer J., Lukashevich H., Bräuer P., 2012, *Classification of music genres based on repetitive basslines*, Journal of New Music Research, vol. 41, no. 3, s. 239-257.
- Abundo C., Monterola C., Tugaff J., Valenzuela J.F., 2009, *Prediction of potential hit song and musical genre using artificial neural networks*, International Journal of Modern Physics C, vol. 20, no. 11, s. 1697-1718.
- Bai B., Yen J., Yang X., 2008, *False financial statements: characteristics of China's listed companies and CART detecting approach*, International Journal of Information Technology & Decision Making, vol. 7, no. 2, s. 339-359.

- Berney S.C., Gordon I.R., Opdam H.I., Denehy L., 2011, *A classification and regression tree to assist clinical decision making in airway management for patients with cervical spinal cord injury*, Spinal Cord, vol. 49, no. 2, s. 244-250.
- Białkowski A., Grusiewicz M., 2009, *Zainteresowania muzyczne i główne formy ich realizowania przez uczniów*, Wychowanie Muzyczne w Szkole, nr 3, s. 19-29.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984, *Classification and Regression Trees*, CRC Press, London.
- Ferraro D.O., Rivero D.E., Ghersa C.M., 2009, *An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees*, Field Crops Research, vol. 112, no. 2/3, s. 149-157.
- Gatnar E., 2009, *Drzewa klasyfikacyjne i regresyjne*, [w:] Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa.
- Gatnar E., 2001, *Nieparametryczna metoda dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- GINOCCHIO J.F., 2006, *Music Style Preference: A Ranking of Musical Styles and Comparison by Age, Gender, Ethnicity, Music Training, and Rural, Suburban, Or Urban Upbringing*, Ball State University.
- HARGREAVES D.J., NORTH A.C., 1997, *The Social Psychology of Music*, Oxford: Oxford University Press.
- KAMIŃSKA B., 2002, *Upodobania muzyczne-problemy i wyniki badań*, [w:] Białkowski B., Smoleńska-Zielińska B. (red.), *Blżej muzyki, blżej człowieka*, Wydawnictwo UMCS, Lublin.
- KASS G.V., 1983, *An exploratory technique for investigating large quantities of categorical data*, Applied Statistics, vol. 29, no. 2, s. 119-127.
- ŁAPCZYŃSKI M., 2010, *Drzewa klasyfikacyjne i regresyjne w badaniach marketingowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- MA J., WANG T., GAO X., WANG S., WANG Z., 2013, *Classification and regression tree-based adaptive damping control of inter-area oscillations using wide-area signals*, Transmission & Distribution, vol. 8, no. 6, s. 1177-1186.
- ŚLASKA-GRZYWNA B., 2010, *Wykorzystanie drzew klasyfikacyjnych w analizie cech sensorycznych brokułu włoskiego po obróbce cieplnej*, Inżynieria Rolnicza, vol. 3, nr 121, s. 209-2015.
- WALDORFF J., 1982, *Muzyka łagodzi obyczaje*, Polskie Wydawnictwo Muzyczne, Kraków.
- ZAHIR S.T., BINESH F., MIROULIAEI M., KHAJEH E., NOSHAD S., 2013, *Malignancy risk assessment in patients with thyroid nodules using classification and regression trees*, Journal of Thyroid Research, vol. 2013, s. 1-9.

ATTEMPT TO IDENTIFY MUSIC LOVERS PROFILES USING CLASSIFICATION AND REGRESSION TREES

Summary: The purpose of this article is the segmentation of classical music lovers by spending size on musical culture and by preferred music genres, or composers. For the classification of music lovers there are used regression and classification trees created in accordance with the CART and CHAID algorithms based on data from a survey conducted among classical music lovers. These research methods have enabled the creation of music lovers profiles, which help in better knowledge of the demand for music products in different groups of respondents.

Keywords: classification trees, regression trees, CART algorithm, CHAID algorithm, music.