

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 426

Taksonomia 26

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



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

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 stronach internetowych
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 2016

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

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
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
Jacek Batóg: Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis	13
Andrzej Bąk: Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
Grażyna Dehnel: <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
Andrzej Dudek: <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
Iwona Foryś: Analiza dyskryminacyjna w wyborze obiektów podobnych w procesie szacowania nieruchomości / The discriminant analysis in selection of similar objects in the real estate valuation process	51
Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz: Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
Iwona Konarzewska: Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria	69
Anna Król, Marta Targaszewska: Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
Marek Lubicz: Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
Aleksandra Łuczak: Wpływ różnych sposobów agregacji opinii ekspertów w FAHP na oceny priorytetowych czynników rozwoju / Influence of different methods of the expert judgments aggregation on assessment of priorities for evaluation of development factors in FAHP.....	99
Iwona Markowicz: Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity	108

Małgorzata Markowska, Danuta Strahl: Filary inteligentnego rozwoju a wrażliwość unijnych regionów szczebla NUTS 2 na kryzys ekonomiczny – analiza wielowymiarowa / Smart development pillars and NUTS 2 European regions vulnerability to economic crisis – a multidimensional analysis.....	118
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis	130
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis	139
Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta: Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce) / Problem of outliers in corporate bankruptcy prediction (case of manufacturing companies in Poland)	148
Wojciech Roszka: Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
Małgorzata Rószkiewicz: Czynniki różnicujące efektywność pracy ankietera w wywiadach <i>face-to-face</i> w środowisku polskich gospodarstw domowych / Factors affecting the efficiency of face-to-face interviews with Polish households.....	166
Adam Sagan, Marcin Pelka: Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data	174
Marcin Salamaga: Zastosowanie drzew dyskryminacyjnych w identyfikacji czynników wspomagających wybór kraju alokacji bezpośrednich inwestycji zagranicznych na przykładzie polskich firm / The use of classification trees in the identification of factors supporting the choice of FDI destination on the example of Polish companies.....	185
Agnieszka Stanimir: Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
Mirosława Sztemberg-Lewandowska: Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge	206
Tadeusz Trzaskalik: Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature	214

Joanna Trzęsiok: Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions	226
Hanna Wdowicka: Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
Artur Zaborski: Zastosowanie skalowania dynamicznego oraz metody wektorów dryfu do badania zmian w preferencjach / The use of dynamic scaling and the drift vector method for studying changes in the preferences.....	245

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego. Przewodniczącymi Komitetu Organizacyjnego konferencji byli prof. dr hab. Mirosław Szreder oraz dr hab. Krzysztof Najman, prof. nadzw. UG, sekretarzami naukowymi dr hab. Kamila Migdał-Najman, prof. nadzw. UG oraz dr hab. Anna Zamojska, prof. nadzw. UG, a sekretarzem organizacyjnym Anna Nowicka z Fundacji Rozwoju Uniwersytetu Gdańskiego.

Konferencja Naukowa została dofinansowana ze środków Narodowego Banku Polskiego.

Zakres tematyczny konferencji obejmował takie zagadnienia, jak:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczymi celami konferencji SKAD były prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 81 osób. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: AGH w Krakowie, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwer-

sytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, a także przedstawiciele NBP i PBS Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz trzynastu sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Józef Pocięcha, Eugeniusz Gatnar, Tadeusz Trzaskalik, Krzysztof Jajuga, Marek Walesiak, Barbara Pawełek, Feliks Wysocki, Ewa Roszkowska, Andrzej Sokołowski, Andrzej Bąk, Tadeusz Kufel, Mirosław Krzyśko, Krzysztof Najman, Małgorzata Rószkiewicz, Mirosław Szreder.

Teksty 25 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii „Taksonomia” nr 26. Pozostałe recenzowane artykuły znajdują się w „Taksonomii” nr 27.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pocięcha. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

- A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
- B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
- C. Organizacja konferencji SKAD PTS w latach 2016 i 2017.
- D. Wybór przedstawiciela Rady Sekcji SKAD PTS do IFCS.
- E. Dyskusja nad kierunkami rozwoju działalności Sekcji.

Prof. dr hab. Józef Pocięcha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajdują się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii „Taksonomia” nr 24 i 25 (PN UE we Wrocławiu nr 384 i 385). W „Przeglądzie Statystycznym” (zeszyt 4/2014) ukazało się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Międzyzdrojach, w dniach 8–10 września 2014 r. Prof. Barbara Pawełek przedstawiła także informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków i sympatyków SKAD.

W konferencji Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS – International Federation of Classification Societies) w dniach 6–8 lipca 2015 r. w Bolonii, zorganizowanej przez Università di Bologna, udział wzięło 19 osób z Polski (w tym 17 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 79,0%). Ponadto prof. Józef Pocięcha był członkiem Komitetu Naukowego Konferencji z ramienia SKAD, członkiem Międzynarodowego Komitetu Nagród IFCS oraz organizatorem i przewodniczącym sesji nt. „Classification models for forecasting of economic processes”.

W konferencji „European Conference on Data Analysis” (Colchester, 2–4 września 2015 r.) zorganizowanej przez The German Classification Society (GfKI) we współpracy z The British Classification Society (BCS) i Sekcją Klasyfikacji i Analizy Danych PTS (SKAD) udział wzięło 18 osób z Polski (w tym 14 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 66,0%). Ponadto profesorowie Krzysztof Jajuga oraz Józef Pociecha byli członkami Komitetu Naukowego konferencji, prof. Andrzej Dudek został poproszony przez organizatorów o przygotowanie referatu i wygłoszenie na Sesji Plenarnej „Cluster analysis in XXI century, new methods and tendencies”, prof. Krzysztof Jajuga był przewodniczącym sesji plenarnej, przewodniczącym sesji nt. „Finance and economics II” oraz organizatorem i przewodniczącym sesji nt. „Data analysis in finance”, prof. Józef Pociecha był organizatorem i przewodniczącym sesji nt. „Outliers in classification procedures – theory and practice”, prof. Andrzej Dudek był przewodniczącym sesji nt. „Machine learning and knowledge discovery II”.

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pociecha poinformował o dwóch wybranych konferencjach krajowych (były to XXXIV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015”, Łódź, 16–18 listopada 2015 r. i X Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 10–13 maja 2016 r.) oraz o trzech wybranych konferencjach zagranicznych. Konferencja „European Conference on Data Analysis” odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu w dniach 26–28 września 2017 r. W przeddzień tej konferencji, tj. 25.09.2017 r., odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2017”. Następna konferencja Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS) odbędzie się w 2017 r. w Tokio. W 2019 r. Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2019” organizuje prof. Andreas Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2016 zorganizuje Katedra Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

W kolejnej części zebrania dokonano wyboru przedstawiciela Rady Sekcji SKAD PTS do IFCS na kadencję 2016–2019. Powołano Komisję Skrutacyjną, której przewodniczącym został prof. Tadeusz Kufel, a członkami dr hab. Iwona Konarzewska i dr Dominik Rozkrut. Profesor Józef Pociecha poprosił zebranych o proponowanie kandydatur zgłaszając jednocześnie prof. Andrzeja Sokołowskiego. Wobec braku następnych kandydatur listę zamknięto. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 41 członków Sekcji. Profesor Andrzej Sokołowski został przedstawicielem Rady Sekcji SKAD PTS do

IFCS na kadencję 2016–2019, uzyskując następujący wynik: 39 głosów na „tak”, 1 głos na „nie”, 1 głos był nieważny.

W ostatnim punkcie zebrania dyskutowano nad kierunkami rozwoju działalności Sekcji obejmującymi następujące problemy: udział w międzynarodowym ruchu naukowym (wspólne granty, publikacje), umiędzynarodowienie konferencji SKAD (uczestnicy zagraniczni, dwujęzyczność konferencji), wydawanie własnego czasopisma.

Profesor Józef Pociecha zamknął posiedzenie Sekcji SKAD.

Krzysztof Jajuga, Marek Walesiak

Marcin Salamaga

Uniwersytet Ekonomiczny w Krakowie
e-mail: salamaga@uek.krakow.pl

**ZASTOSOWANIE DRZEW DYSKRYMINACYJNYCH
W IDENTYFIKACJI CZYNNIKÓW
WSPOMAGAJĄCYCH WYBÓR KRAJU ALOKACJI
BEZPOŚREDNICH INWESTYCJI ZAGRANICZNYCH
NA PRZYKŁADZIE POLSKICH FIRM**

**THE USE OF CLASSIFICATION TREES
IN THE IDENTIFICATION OF FACTORS
SUPPORTING THE CHOICE OF FDI DESTINATION
ON THE EXAMPLE OF POLISH COMPANIES**

DOI: 10.15611/pn.2016.426.19

Streszczenie: Celami artykułu są identyfikacja i hierarchizacja determinantów wyboru destynacji bezpośrednich inwestycji zagranicznych przez polskie firmy oraz segmentacja tych przedsiębiorstw. Przedmiotem badania są czynniki o charakterze kosztowym, rynkowym, efektywnościowym, prawnym, społecznym i politycznym, które mogą mieć znaczenie przy poszukiwaniu beneficjentów BIZ. Do klasyfikacji firm zastosowano drzewa decyzyjne tworzone zgodnie z procedurą CART. Metoda ta umożliwiła stworzenie profili przedsiębiorstw o zbliżonej polityce inwestycyjnej w zakresie priorytetów w wyborze destynacji BIZ. W artykule posłużono się danymi pochodzącymi z badania ankietowego przeprowadzonego wśród polskich przedsiębiorstw podejmujących BIZ lub planujących tę formę inwestowania.

Słowa kluczowe: bezpośrednie inwestycje zagraniczne, destynacja BIZ, badanie ankietowe, drzewa klasyfikacyjne, algorytm CART.

Summary: The purpose of this article is to identify and rank main factors determining the choice of Polish foreign direct investment (FDI) destination. The research includes the cost factors, market factors, efficiency factors, law factors, social factors, and political factors, which may be important in the search for the beneficiaries of FDI. In the paper the CART algorithm is used. This research method enables the creation of profiles of companies which have similar investment policy in terms of priorities of the choice of investment destination. In the research data from a survey conducted among Polish companies investing in FDI or planning this form of investing were used.

Keywords: FDI, FDI destination, survey, classification trees, CART algorithm.

1. Wstęp

Bezpośrednie inwestycje zagraniczne stały się ważnym elementem rozwoju gospodarczego w wielu krajach i odzwierciedlają stopień umiędzynarodowienia poszczególnych gospodarek. Zarówno różne teorie ekonomiczne, jak i praktyka gospodarcza pokazują, że dostawcami BIZ są kraje wysoko rozwinięte, o znacznym potencjale gospodarczym i z odpowiednią akumulacją kapitału. Do niedawna Polska była głównie beneficjentem BIZ, ale od kilku lat firmy z polskim kapitałem coraz częściej inwestują na rynkach zagranicznych, a eksportowane przez nie bezpośrednie inwestycje stanowią integralną część ich działalności międzynarodowej. Determinanty warunkujące podejmowanie decyzji o uruchomieniu bezpośrednich inwestycji zagranicznych są stosunkowo obszernie opisywane w krajowej i zagranicznej literaturze przedmiotu (zob. [Fry 1993]). Wśród wielu klasyfikacji determinantów BIZ często powtarzają się czynniki o charakterze kosztowym, rynkowym, handlowym oraz czynniki tworzące tzw. klimat inwestycyjny [Ozawa, Lukas 1993; Pilarska 2005; Rymarczyk 2004]. Część z nich ma charakter uniwersalny, a inne są bardziej typowe dla branży gospodarczej, w której działa firma. Czynniki lokalizacyjne, czyli decydujące o korzystniejszych warunkach produkcji w kraju goszczącym BIZ w porównaniu z krajem macierzystym są analizowane w tzw. teorii czynników lokalizacyjnych. W ramach tej teorii ekonomiści próbują wyjaśnić strukturę lokowania BIZ przy założeniu, że firmy dysponują przewagami konkurencyjnymi. To właśnie teoria czynników lokalizacyjnych stanowi inspirację dla autora do przeprowadzenia badania empirycznego determinantów BIZ. Zasadniczym celem tego artykułu jest zdiagnozowanie tych czynników, które mogą wpływać na wybór kierunku eksportu bezpośrednich inwestycji zagranicznych firm działających w Polsce. Analiza została wykonana na podstawie danych pochodzących z badania ankietowego, które przeprowadzono wśród polskich przedsiębiorstw będących dostawcami BIZ lub planującymi podjąć tę formę inwestycji w najbliższym czasie. Zależność pomiędzy motywacjami i kierunkami eksportu BIZ badano za pomocą drzew klasyfikacyjnych CART. Metoda ta umożliwiła również wyodrębnienie profili eksporterów BIZ, którzy kierują się podobnymi motywami, poszukując destynacji dla swoich inwestycji bezpośrednich. Warto podkreślić, że w literaturze przedmiotu trudno znaleźć badania z użyciem metod statystycznych mające na celu identyfikację motywów podejmowania BIZ w zależności od wyboru ich destynacji. Tymczasem, w przekonaniu autora, zastosowanie statystycznych metod eksploracji danych podnosi wiarygodność wyników końcowych i wzmacnia siłę prezentowanych wniosków. Zastosowanie metody CART należy uznać za oryginalne i jednocześnie skuteczne rozwiązanie w tego typu analizie. Wyniki przedstawione w artykule stanowią również przegląd aktualnych preferencji polskich przedsiębiorstw w zakresie kierunków eksportu BIZ w powiązaniu z motywami wyboru tych kierunków.

2. Metoda badania

Segmentację respondentów przeprowadzono, stosując drzewa klasyfikacyjne typu CART – *Classification and Regression Trees* [Breiman i in. 1984]. Wybór tej metody klasyfikacyjnej był podyktowany jakościowym charakterem danych pozyskanych w badaniu ankietowym oraz jej adekwatnością do problemu badawczego. Analizę CART cechuje wysoka efektywność predykcji jakościowej, w przypadku której stosowalność wielu metod statystycznych i ekonometrycznych jest ograniczona. Ponadto analiza CART jest odporna na występowanie jednostek nietypowych i dopuszcza wykorzystanie zmiennych charakteryzujących się brakami danych [Gatnar 2001]. Zaletami przedmiotowej metody są również: brak konieczności kodowania zmiennych jakościowych (np. na zmienne zero-jedynkowe), brak wymogów co do liniowości związków między zmiennymi oraz normalności ich rozkładów [Gatnar 2001].

W rezultacie stosowania tej metody otrzymuje się drzewa decyzyjne, czyli grafy o strukturze drzewiastej. Wynikowe drzewo składa się z węzła początkowego (korzenia), węzłów podlegających podziałowi na węzły potomne oraz węzłów końcowych (liści). 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 dyskryminacyjnych 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 przybierająca 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_t ($t = 1, 2, \dots, u$) oznacza klasę, do której należy obserwacja x_i .

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

$$\Delta Q(R) = Q(R) - \sum_{s=1}^S Q(R_s) p_s, \quad (3)$$

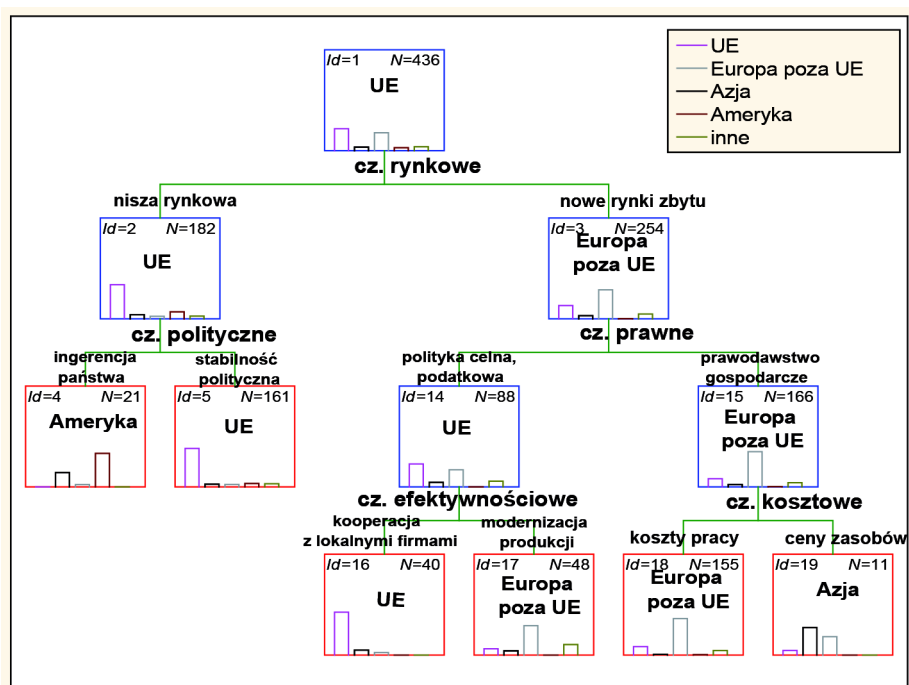
gdzie: $Q(R_s)$ – funkcja zróżnicowania obserwacji znajdujących się w podprzestrzeni R_s , p_s – frakcja obserwacji w podprzestrzeni R_s .

Maksymalizacja wyrażenia (3) zapewnia możliwie największą jednorodność otrzymanych podprzestrzeni, co generuje podział zmiennych uznawany za optymalny. W przypadku drzewa klasyfikacyjnego do oceny homogeniczności podprzestrzeni R_s stosuje się najczęściej błąd klasyfikacji, wskaźnik Giniego lub miarę entropii. W konstrukcji drzew klasyfikacyjnych niepożądanym zjawiskiem jest nadmierna złożoność modelu, która wiąże się ze wzrostem wartości błędu dla zbioru testowego. Przeciwdziałać temu efektowi można za pomocą przycinania krawędzi (*pruning*), co prowadzi do redukcji rozmiarów drzewa. Przycinanie krawędzi ma na celu eliminację tych gałęzi drzewa, których brak nie powoduje zbyt dużego pogorszenia jakości podziałów (monitorowanej za pomocą kryterium (3)). W efekcie powstają drzewa, z których wybierane jest to, dla którego błąd klasyfikacji w zbiorze testowym jest najmniejszy [Gatnar 2001]. Interpretacja drzewa dyskryminacyjnego CART odbywa się zgodnie z regułami klasyfikacyjnymi stworzonymi dla każdej ścieżki drzewa łączącej węzeł początkowy z węzłem końcowym. Jej celem jest identyfikacja kombinacji predyktorów warunkujących występowanie określonej wartości zmiennej zależnej. Źródłem danych wykorzystanych w procedurze CART było badanie ankietowe przeprowadzone w okresie od grudnia 2014 r. do kwietnia 2015 r. wśród polskich firm, które lokowały bezpośrednie inwestycje za granicą lub też na podstawie ich potencjału można przypuszczać, że będą zainteresowane tą formą lokowania kapitału. Kwestionariusz anonimowej ankiety rozesłano do ok. 800 przedsiębiorstw działających w różnych branżach gospodarki. Zwrot ankiet był na poziomie ok. 60%, a po ostatecznej selekcji nadesłanych odpowiedzi w badaniu uwzględniono 436 poprawnie wypełnionych kwestionariuszy. Spośród różnych czynników, o jakie pytano respondentów, zdecydowano się ostatecznie na uwzględnienie w niniejszej analizie sześciu grup czynników: kosztowych (np. *koszty pracy, ceny zasobów naturalnych*), rynkowych (np. *nowe rynki zbytu, poszukiwanie niszy rynkowej*), efektywnościowych (np. *kooperacja z lokalnymi firmami, modernizacja metod produkcji*), prawnych (np. *polityka celna i podatkowa, prawodawstwo gospodarcze i administracyjne*), politycznych (np. *stopień ingerencji państwa w gospodarkę, stabilność polityczna*), społecznych (np. *kwalifikacje kadry pracowniczej, poziom zamożności społeczeństwa*).

3. Wyniki segmentacji firm eksportujących BIZ

Rozwój gospodarczy Polski, rosnący potencjał i konkurencyjność polskich firm połączone z coraz większym stopniem światowej internacjonalizacji przedsiębiorstw spowodowały, że wiele krajowych firm szuka możliwości inwestowania poza granicami kraju. Rozwój nowych technologii, nowoczesne sposoby organizacji transportu sprawiają, że fizyczna odległość nie jest już barierą w eksporcie BIZ. Odbiorcami polskich inwestycji bezpośrednich mogą więc być (i są) nie tylko kraje europejskie, ale także azjatyckie czy afrykańskie. W związku z tym w pierwszym

etapie badania za pomocą metody CART dokonano segmentacji firm prowadzących BIZ (lub dopiero planujących tę formę inwestycji) ze względu na zmienną zależną *perspektywiczny kierunek eksportu BIZ*. Anketowani eksporterzy mogli wskazać na następujące aktualne lub potencjalne kierunki eksportu BIZ: kraje UE, kraje europejskie spoza UE (np. Norwegia, Rosja, Szwajcaria, Ukraina), Ameryka (zarówno Ameryka Północna, jak i Ameryka Południowa), Azja i inne (Afryka, Australia z Oceanią). Wyniki podziałów eksporterów BIZ w postaci drzewa klasyfikacyjnego przedstawiono na rys. 1.



Rys. 1. Drzewo klasyfikacyjne CART dla zmiennej zależnej *perspektywiczny kierunek eksportu BIZ*

Źródło: opracowanie własne.

Przedmiotowe drzewo ma 6 węzłów końcowych i 5 węzłów dzielonych. Utworzone w wyniku podziału metodą CART homogeniczne podzbiory mogą być traktowane jako segmenty dostawców BIZ, którzy preferują te same kierunki eksportu BIZ przy określonych kombinacjach kategorii cech reprezentujących czynniki kosztowe, rynkowe, efektywnościowe, prawne, polityczne i społeczne. Interpretację wyników analizy CART przeprowadzono dla wybranych segmentów eksporterów BIZ wyznaczonych przez węzły końcowe drzewa klasyfikacyjnego:

- jeżeli dostawca BIZ, wybierając rynek inwestycyjny, za priorytetowe uważa znalezienie niszy rynkowej oraz stabilność polityczną w kraju goszczącym BIZ

- (węzeł nr 5), to najprawdopodobniej wybierze rynek UE (z prawdopodobieństwem ok. 0,76);
- jeżeli przedsiębiorca kieruje się kryterium nowego rynku zbytu, prawodawstwem gospodarczym i administracyjnym oraz kosztami pracy w kraju przyjmującym BIZ (węzeł nr 18), to z prawdopodobieństwem ok. 0,72 będzie lokował swoje inwestycje bezpośrednio na rynkach tych krajów europejskich, które nie są członkami UE;
 - jeżeli dostawca BIZ, wybierając rynek inwestycyjny, za najważniejsze uważa znalezienie nowego rynku zbytu i duże znaczenie przypisuje prawodawstwu gospodarczemu oraz cenom zasobów naturalnych w kraju będącym beneficjentem BIZ (węzeł nr 19), to najprawdopodobniej wybierze rynek azjatycki (z prawdopodobieństwem 0,55).

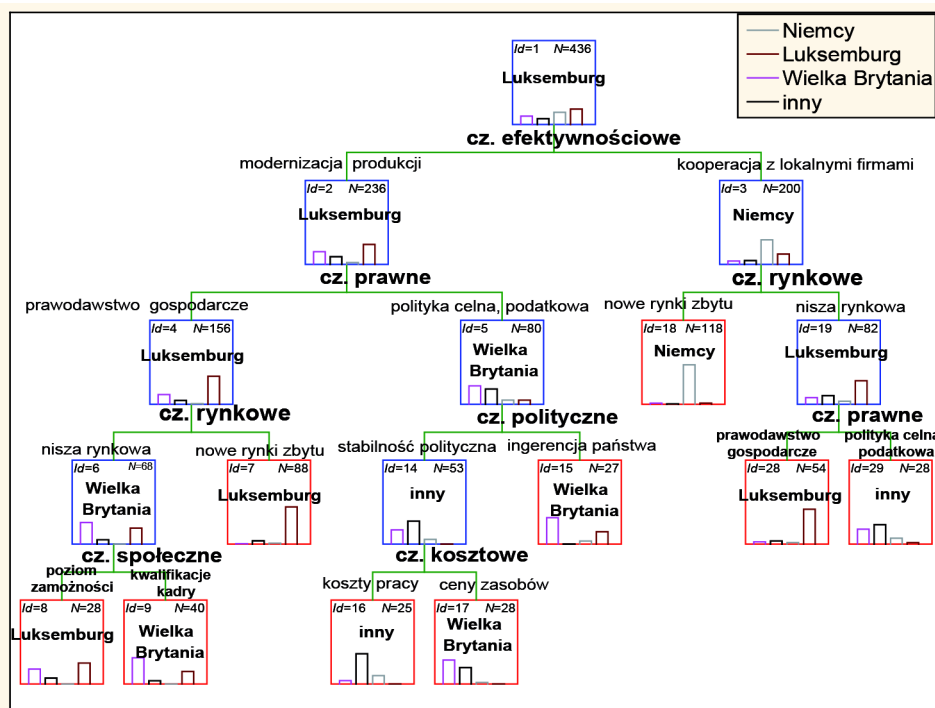
Najwyższy odsetek trafnie zakwalifikowanych eksporterów wystąpił wśród tych przedsiębiorstw, które za najbardziej perspektywiczny dla swoich inwestycji bezpośrednich uważają kierunek europejski poza UE (ok. 90,3% firm) oraz rynek UE (ok. 82,6% firm). Na podstawie badania siły poszczególnych predyktorów w różnicowaniu polskich firm dostarczających BIZ w przedmiotowej analizie CART stwierdzono, że największy wpływ na wybór perspektywicznego kierunku eksportu BIZ miały czynniki efektywnościowe i kosztowe, a najmniejsze znaczenie miały czynniki społeczne i prawne.

Przynależność Polski do Unii Europejskiej zapewniającej swobodny przepływ towarów, usług i osób, geograficzna bliskość wiodącej gospodarki europejskiej (Niemcy) sprawia, że kraje UE są ważnymi odbiorcami polskich BIZ. Unię Europejską jako docelową destynację dla BIZ w niniejszym badaniu wskazało najwięcej (ok. 44%) ankietowanych firm. W związku z tym kolejną analizę CART przeprowadzono dla zmiennej zależnej *perspektywiczny beneficjent BIZ z UE*. Respondentom przedstawiono wybrane kraje UE, które w ostatnich latach były głównymi beneficjentami polskich BIZ. Wzięto tutaj pod uwagę kraj reprezentujący najsilniejszą sąsiadującą z Polską gospodarkę UE (Niemcy), kraj stanowiący główny kierunek migracji ekonomicznej Polaków (Wielka Brytania), jak i kraj stanowiący ważne centrum światowych przepływów finansowych (Luksemburg). Pozostałe kraje UE zaliczono do kategorii *inne*. Wyniki segmentacji polskich eksporterów BIZ za pomocą analizy CART według zmiennej *perspektywiczny beneficjent BIZ z UE* przedstawiono na rys. 2.

Poniżej przedstawiono interpretację wybranych liści drzewa klasyfikacyjnego widocznego na rys. 2:

- jeżeli firma, szukając perspektywicznego odbiorcy BIZ wśród krajów UE, za istotne uważa modernizację metod produkcji, prawodawstwo gospodarcze i administracyjne, poszukiwanie niszy rynkowej oraz odpowiednie kwalifikacje kadry w kraju goszczącym BIZ (węzeł nr 9), to najprawdopodobniej wybierze Wielką Brytanię (z prawdopodobieństwem ok. 0,63);

- jeżeli firma, szukając perspektywicznego odbiorcy BIZ, kieruje się modernizacją metod produkcji, polityką celną i podatkową, a także stopniem ingerencji w gospodarkę rządu kraju przyjmującego BIZ (węzeł nr 15), to również najprawdopodobniej wybierze Wielką Brytanię (z prawdopodobieństwem ok. 0,63);
- jeżeli firma, poszukując odbiorcy BIZ wśród krajów UE, za priorytetowe uważa kooperację z miejscowymi firmami oraz poszukiwanie nowych rynków zbytu w kraju goszczącym BIZ (węzeł nr 18), to najprawdopodobniej wybierze Niemcy (z prawdopodobieństwem ok. 0,94);
- jeżeli eksporter BIZ kieruje się możliwością kooperacji z lokalnymi firmami, poszukiwaniem niszy rynkowej, a także prawodawstwem gospodarczym i administracyjnym w kraju będącym beneficjentem BIZ (węzeł nr 28), to z największym prawdopodobieństwem równym ok. 0,83 wybierze Luksemburg.



Rys. 2. Drzewo klasyfikacyjne CART dla zmiennej zależnej *perspektywiczny beneficjent BIZ z UE*

Źródło: opracowanie własne.

Najwięcej trafnie zakwalifikowanych eksporterów w aktualnej analizie CART znalazło się wśród tych, którzy za najbardziej perspektywiczny dla swoich bezpośrednich inwestycji zagranicznych uważają rynek niemiecki (ok. 87% eksporte-

rów), a w dalszej kolejności – Luksemburg (ok. 85%). Na podstawie badania siły poszczególnych predyktorów w różnicowaniu segmentów eksporterów BIZ stwierdzono, że największy wpływ na wybór perspektywicznego odbiorcy BIZ w UE miały czynniki rynkowe i kosztowe, a najmniej istotne okazały się czynniki społeczne i efektywnościowe. Analizując wyniki metody CART, warto zwrócić uwagę, że stosunkowo częste wskazywanie (przy różnych kombinacjach predyktorów) Luksemburga i Wielkiej Brytanii (krajów należących do światowych centrów finansowych) jako beneficjentów polskich BIZ może świadczyć o tym, że BIZ częściej mają postać przepływów finansowych niż inwestycji typu *greenfield*. Ułatwia to możliwości transferowania zysków polskich firm za granicę oraz sprzyja „optymalizacji podatkowej”. Istniejące przepisy prawne i podatkowe omawianych krajów stanowią również zachętę do rejestrowania w nich spółek córek, za pośrednictwem których BIZ prowadzone są również na rynkach krajów trzecich.

4. Zakończenie

Wyniki badań potwierdziły skuteczność analizy CART w segmentacji firm, które już lokują kapitał w formie BIZ lub planują to zrobić w niedalekiej przyszłości. W badaniu wyodrębniono profile eksporterów BIZ, którzy kierują się podobnymi motywami, poszukując krajów beneficjentów BIZ. Działalność gospodarcza o zasięgu międzynarodowym niesie za sobą określonego rodzaju ryzyka, np. ryzyko polityczne, rynkowe, walutowe i inne, które muszą być brane pod uwagę przez racjonalnie działające przedsiębiorstwo zamierzające lokować swój kapitał w formie BIZ. Tym bardziej więc staranny wybór kraju beneficjenta BIZ jest istotny, bo pozwala przynajmniej częściowo ograniczyć niektóre ryzyka towarzyszące międzynarodowej działalności gospodarczej firmy. W przedmiotowym badaniu najważniejsze w wyborze docelowego rynku eksportu BIZ okazywały się czynniki ekonomiczne (rynkowe, kosztowe) i to wydaje się uzasadnione z punktu widzenia przedsiębiorstwa maksymalizującego zysk. Czynniki polityczne w niewielkim stopniu determinowały wybory destynacji BIZ zapewne z powodu względnie wyższej stabilności politycznej szczególnie zachodnioeuropejskich krajów UE (w porównaniu z niektórymi krajami europejskimi leżącymi poza UE). Przedstawione wyniki badania mogą stanowić wsparcie w procesie decyzyjnym tych przedsiębiorstw, które planują uruchomienie inwestycji bezpośrednich i poszukują dla nich odpowiedniej destynacji. Ze względu na dynamiczne zmiany w sytuacji społeczno-gospodarczej i politycznej na świecie tego typu badania wymagają kontynuowania, co pozwoli śledzić aktualne trendy w obszarze motywów wyboru kierunku BIZ przez polskie firmy.

Literatura

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984, *Classification and Regression Trees*, CRC Press, London.
- Fry M., 1993, *Foreign Direct Investment in a Macroeconomic Framework: Finance Efficiency, Incentives and Distortions*, Working Paper WPS, no. 1141.
- Gatnar E., 2001, *Nieparametryczna metoda dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Gatnar E., 2009, *Drzewa klasyfikacyjne i regresyjne*, [w:] M. Walesiak, E. Gatnar (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa.
- Ozawa T., Lukas R., 1993, *On the determinants of Direct Foreign Investment: Evidence from East and Southeast Asia*, World Development, no. 21.
- Pilarska C., 2005, *Bezpośrednie inwestycje zagraniczne w teorii ekonomii*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Rymarczyk J., 2004, *Internacjonalizacja i globalizacja przedsiębiorstwa*, PWE, Warszawa.