

*Grażyna Trzpiot**
*Alicja Ganczarek-Gamrot***

DRZEWA DECYZYJNE W STATYSTYCZNEJ ANALIZIE DECYZJI NA PRZYKŁADZIE WIRTUALNYCH ŁAŃCUCHÓW DOSTAW

Streszczenie. Celem artykułu jest określenie stopnia oddziaływania stosowanej przez przedsiębiorstwa: technologii informacyjno-komunikacyjnej, kompetencji logistycznych, oraz ich branży transportowo-spedycyjno-logistycznej na rozwój wirtualnych łańcuchów dostaw. Ponadto zaproponowanie działań, umożliwiających efektywne kształtowanie się rozwoju wirtualnych łańcuchów dostaw w danych przedsiębiorstwach. Dysponując informacjami badania ankietowego przedsiębiorstw Województwa Śląskiego, których oddziaływanie na rozwój regionalny i konkurencyjność gospodarki jest zauważalny, za pomocą drzew decyzyjnych przeprowadzono statystyczną analizę decyzji wyboru najkorzystniejszych działań z punktu widzenia poziomu rozwoju technologii informacyjno-komunikacyjnej oraz rozwiązań logistycznych. Wyniki analizy zostaną wykorzystane do pracy z zakresu logistyki na temat „Czynniki rozwoju wirtualnych łańcuchów dostaw.”

Słowa kluczowe: wirtualny łańcuch dostaw, drzewa klasyfikacyjne, metody podziału, reguły stopu.

I. WPROWADZENIE

Rozwój rynków finansowych i towarowych prowadzi do powstawania nowych organizacji a tym samym nowych zależności w otoczeniach tych rynków. Złożoność kształtujących się procesów wymaga szerszego spojrzenia na analizowane problemy. Metody statystyki jednowymiarowej stają się niewystarczające do pełnej analizy tych procesów.

Dzięki osiągnięciom matematyków w dziedzinach algebry liniowej oraz geometrii wielowymiarowej na przełomie XIX i XX wieku pojawiły się początki Statystycznej Analizy Wielowymiarowej (SAW). Korzystając z możliwości uogólnienia klasycznych metod statystyki jednej zmiennej na przypadki wielowymiarowe, jak również postępu technik obliczeniowych, metody statystyki wielowymiarowej rozwijają się zarówno pod względem teorii, jak również zastosowań (Jajuga (1993)).

* Prof. AE, Dr hab., Katedra Statystyki, Akademia Ekonomiczna w Katowicach.

** Dr, Katedra Statystyki, Akademia Ekonomiczna w Katowicach.

Łańcuch dostaw rozumiany jako przepływ surowców, materiałów, podzespołów i wyrobów gotowych od momentu pozyskania tych pierwszych, do momentu konsumpcji wyrobu finalnego przez użytkownika końcowego, można postrzegać jako organizację łączącą przedsiębiorstwa uczestniczące w procesie dostarczania danego produktu na rynek. Wirtualny łańcuch dostaw jest organizacją tymczasową powołaną do konkretnego celu, koncentrującą się na kliencie, wykorzystującą technologie informatyczne (Adamczyk (2005)).

Celem jaki postawili sobie autorzy artykułu jest wykorzystanie technik SAW do określenia stopnia oddziaływania stosowanej przez przedsiębiorstwa: technologii informacyjno-komunikacyjnej, kompetencji logistycznych, oraz ich branży transportowo-spedycyjno-logistycznej na rozwój wirtualnych łańcuchów dostaw. Dysponując informacjami badania ankietowego 121 przedsiębiorstw Województwa Śląskiego, których oddziaływanie na rozwój regionalny i konkurencyjność gospodarki jest zauważalny, za pomocą drzew decyzyjnych przeprowadzono statystyczną analizę decyzji wyboru najkorzystniejszych działań z punktu widzenia poziomu rozwoju technologii informacyjno-komunikacyjnej oraz rozwiązań logistycznych.

II. DRZEWA KLASYFIKACYJNE

Różnorodność i złożoność procesów ekonomicznych wpływa na ciągle poszukiwanie metod SAW za pomocą których można wyczerpująco i przejrzysto opisać złożoność zjawisk ekonomicznych, min. relacje między czynnikami kształtującymi te zjawiska oraz siłę oddziaływania poszczególnych czynników na rezultaty zjawisk ekonomicznych. Wśród bogatej liczby metod w pracy wykorzystano drzewa klasyfikacyjne. Metodę tą spopularyzowali Breiman, Friedman, Olsen i Stone wydając w 1984 roku książkę oraz opracowując program CART (Classification and Regression Trees). (Gatnar (1998), Gatnar, Walesiak (2004)).

Drzewo jest graficzną prezentacją podziału przestrzeni m wymiarowej X^m na rozłączne fragmenty (segmenty przestrzeni, wielowymiarowe kostki). Procedura podziału ma charakter rekurencyjny. W każdym kolejnym kroku wyjściowy fragment przestrzeni jest rozdzielany w optymalny sposób na dwie lub więcej części za pomocą jednej zmiennej z przestrzeni X^m . Zmienna oraz miejsce podziału są tak wybierane aby zoptymalizować jednorodność uzyskanego podziału ze względu na wybraną cechę y – zmienną zależną. Tak opisany model można zapisać wzorem (Gatnar, Walesiak (2004)):

$$y = \sum_{k=1}^K a_k I\{\mathbf{x}_i \in R_k\} \quad (1)$$

gdzie:

\mathbf{x}_i – wielowymiarowa obserwacja, element przestrzeni X^m

$R_k - k=1, \dots, K$, segment przestrzeni X^m ,

a_k – parametry modelu

$I(q)$ – funkcja wskaźnikowa: $I(q) = \begin{cases} 1 & \text{gdy } q \text{ jest prawdziwe} \\ 0 & \text{gdy } q \text{ jest nieprawdziwe} \end{cases}$

Model (1) jest połączeniem modeli lokalnych, które w zależności od charakteru zmiennych \mathbf{x}_i można zapisać (Gatnar, Walesiak (2004)):

– w przypadku, gdy \mathbf{x}_i mają charakter metryczny równaniem:

$$I\{\mathbf{x}_i \in R_k\} = \prod_{j=1}^m I(v_{kj}^{(d)} \leq x_j \leq v_{kj}^{(g)}) \quad (2)$$

– w przypadku, gdy \mathbf{x}_i mają charakter niemetryczny równaniem:

$$I\{\mathbf{x}_i \in R_k\} = \prod_{j=1}^m I(x_j \in B_{kj}) \quad (3)$$

gdzie

– $v_{kj}^{(d)}$ i $v_{kj}^{(g)}$, to dolna i górna granica odcina w j-tym wymiarze przestrzeni,

– B_{kj} , to podział zbioru wartości zmiennej x_j .

W przypadku gdy zmienna zależna y jest zmienną nominalną, model (1) nazywany jest klasyfikacyjnym. Parametry modelu są ustalane zgodnie z zasadą:

$$a_k = \arg \max_l \{p(l/k)\} \quad (4)$$

gdzie $p(l/k)$ oznacza prawdopodobieństwo, że wybrany obiekt z segmentu R_k należy do klasy l .

W przypadku gdy zmienna zależna y jest zmienną mierzoną na jednej ze skal mocnych, model (1) nazywany jest regresyjnym. Parametry modelu wyznaczamy ze wzoru (Gatnar, Walesiak (2004)):

$$a_k = \frac{1}{N(k)} \sum_{i=1}^{N(k)} y_i \quad (5)$$

gdzie $N(k)$ oznacza liczbę obiektów, które znajdują się w R_k .

Nieparametryczny model (1) ma tę przewagę nad klasycznymi metodami SAW, że:

- nie wymaga spełnienia warunków co do rozkładów zmiennych x_i charakteryzujących badane obiekty,
- słabe skale pomiarowe zmiennych x_i nie stanowią problemu w ujęciu ich w modelu,
- jest odporny na występowanie zmiennych nieobserwowalnych,
- jest odporny na występowanie obserwacji nietypowych (Gatnar, Walesiak (2004)).

Na efektywność modelu (1) wpływa między innymi dobór zmiennych. Do oceny efektywności modelu związanej z wyborem zmiennych wykorzystuje się miary jakości podziału, które badają stopień zróżnicowania segmentów ze względu na wartości zmiennej y , innymi słowy poziom heterogeniczności. W niniejszym opracowaniu w tym celu wykorzystano wskaźnik zróżnicowania Giniego (Gatnar, Walesiak (2004)):

$$H(S_k) = 1 - \sum_{l=1}^L p^2(l/k), \quad (6)$$

Efektywność modelu (1) nie jest tylko związana z doбором zmiennych. W wyniku podziału przestrzeni X^m , chcielibyśmy ponad homogeniczność podziału otrzymać w miarę proste drzewo. Wymaga to z drugiej strony zatrzymania rekurencyjnego podziału przed uzyskaniem pełnej homogeniczności segmentów. Problem ten wymaga odpowiedniego kryterium przzerwania podziału. Wśród najpopularniejszych reguł stopu można wymienić:

- regułę przycinania krawędzi przy błędzie złej klasyfikacji,

$$e(D) = \frac{1}{n} \sum_{i=1}^n I\{D(x_i) \neq y_i\} \quad (7)$$

– regułę przycinania krawędzi przy odchyleniu bazującą na zasadzie największej wiarygodności

$$D^{**} = \arg \min_{W(D^*)} \{e^T(D^*) + S(e^T(D^*))\} \quad (8)$$

gdzie D^* – najmniejsze drzewo, liczone na podstawie próby testowej (e^T) lub sprawdzianu krzyżowego,

– regułę bezpośredniego zatrzymania FACT (Fast Algorithm for Classification Trees) dla zadanej frakcji obiektów.

III. DRZEWY KLASYFIKACYJNE PRZEDSIĘBIORSTW WOJEWÓDZTWA ŚLĄSKIEGO

Bazując na wynikach klasyfikacji czynników kształtujących wirtualne łańcuchy dostaw w skład których wchodzi: stosowana technologia informacyjno-komunikacyjna, kompetencje logistyczne, branża transportowo-spedycyjno-logistyczna (tabela 1) (Trzpiot, Ganczarek), w następnym etapie badania przedsiębiorstw Województwa Śląskiego, podjęto próbę oceny stopnia oddziaływania tych czynników na kształtowanie się wirtualnego łańcucha dostaw.

Dysponując informacjami badania ankietowego 121 przedsiębiorstw Województwa Śląskiego, których oddziaływanie na rozwój regionalny i konkurencyjność gospodarki jest zauważalny, za pomocą drzew decyzyjnych przeprowadzono statystyczną analizę decyzji wyboru najkorzystniejszych działań z punktu widzenia poziomu rozwoju technologii informacyjno-komunikacyjnej oraz rozwiązań logistycznych.

W artykule na bazie modelu (1) zaprezentowano kształtowanie się wirtualnego łańcucha dostaw zmiennej y w zależności od sklasyfikowanych wcześniej czynników kształtujących ten łańcuch. W tabeli 2 przedstawiono tabele liczebności zmiennej y reprezentującej wirtualny łańcuch dostaw. Ideą tworzenia łańcucha dostaw jest tu zachowanie odrębności prawnej przez indywidualne podmioty, które działają na rzecz optymalizacji przepływów fizycznych między sobą.

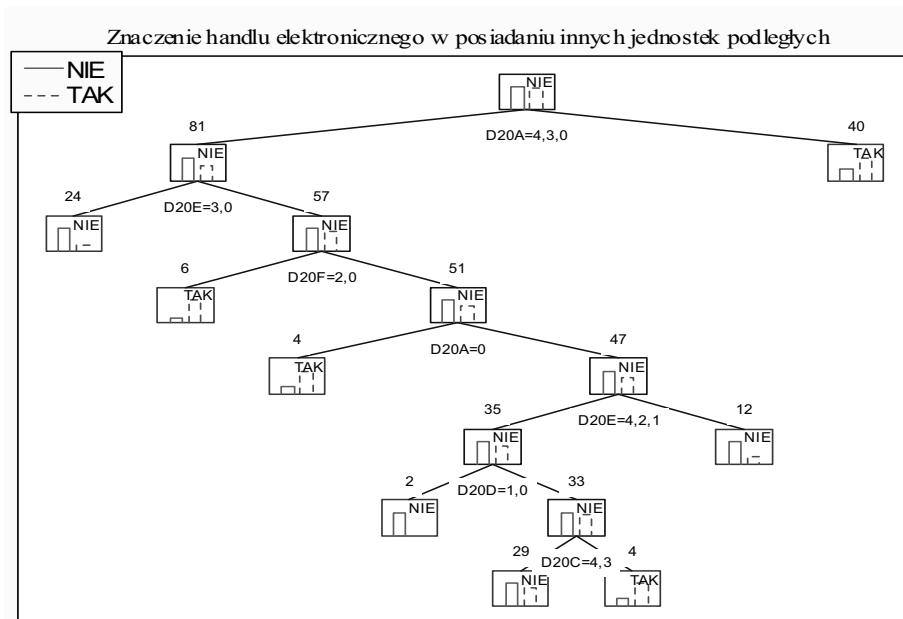
Tabela 1. Wyniki klasyfikacji czynników kształtujących wirtualne łańcuchy dostaw – zmienne x modeli (1)

Odległość euklidesowa	11,68138	17,26794	21,52996	24,60771	25,42024
Zmienne x	Znaczenie handlu elektronicznego	Procesy transportowe	Ocena wykorzystania rozwiązań informacyjno-komunikacyjnych	Planowanie zastosowania technologii informacyjno-komunikacyjnych w różnych obszarach działalności firmy	Wykorzystane do tej pory Rozwiązania informacyjno-komunikacyjne
Obj. Nr 1	D20A	D18A (obecnie)	E21 A	A1D	A2B
Obj. Nr 2	D20B	D18E (za 3 lata)	E21 B	B7I	A4B
Obj. Nr 3	D20D	D18C (obecnie)	E21 C	A2A	D19A (obecnie)
Obj. Nr 4	D20C	D18G(za 3 lata)	E21 D	A4A	D19A (obecnie)
Obj. Nr 5	D20F	D18B (obecnie)	E21 E	B8B	B6A
Obj. Nr 6	D20E	D18F (za 3 lata)	E21 F	B6C	B6B
Obj. Nr 7		D18D (obecnie)	E21 T	D19B(za 3 lata)	B8C
Obj. Nr 8		D18H(za 3 lata)	E21 U	D19B(za 3 lata)	B9F
Obj. Nr 9			E21 G	B7A	B9D
Obj. Nr 10			E21 H	B7G	B9E
Obj. Nr 11			E21 I	B7H	B9A
Obj. Nr 12			E21 J	B7B	B9B
Obj. Nr 13			E21 K	B7E	B9C
Obj. Nr 14			E21 L	B7F	B11A
Obj. Nr 15			E21 O	B7C	B11B
Obj. Nr 16			E21 M	B7D	B11C
Obj. Nr 17			E21 S		
Obj. Nr 18			E21 N		
Obj. Nr 19			E21 P		
Obj. Nr 20			E21 R		
Obj. Nr 21			E21 Q		

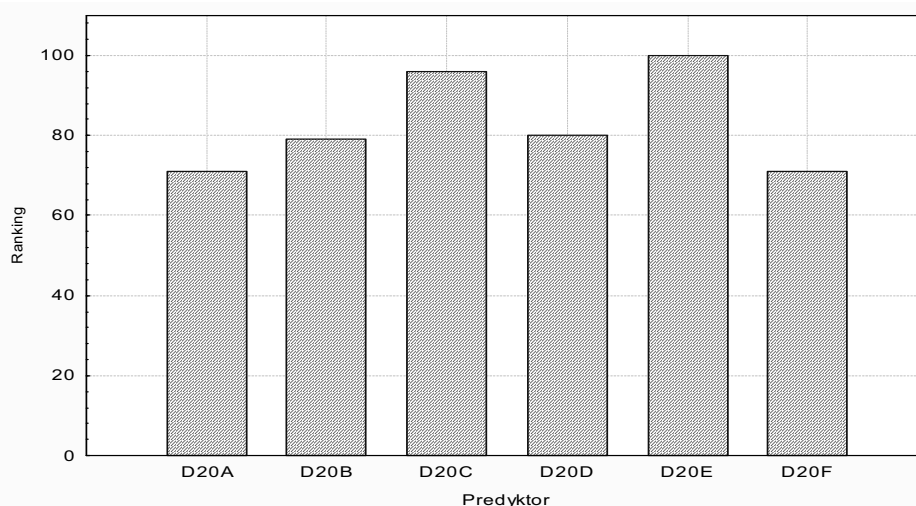
Tabela 2. Rozkład liczebności zmiennej y modelu (1)

Odpowiedz/Pytanie	Czy Państwa firma posiada inne jednostki podległe (np. filie, oddziały, itp.)?	
	Liczność	Procent
NIE	62	51
TAK	59	49

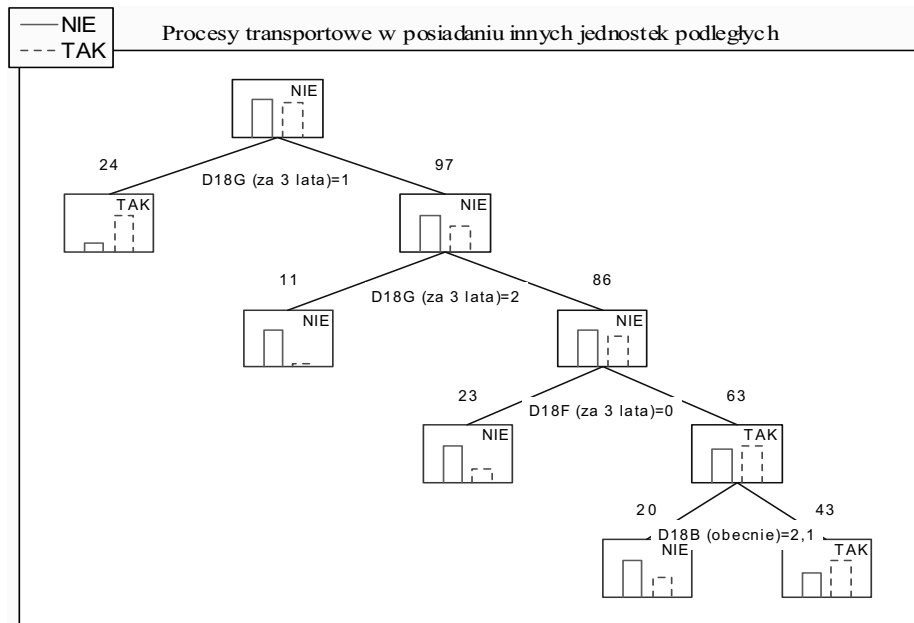
Na rysunkach (1–10) przedstawiono wyniki pięciu rozwiązań modelu (1) w zależności od dobranych w drodze klasyfikacji predyktorów. W tabeli 3 zamieszczono informacje dotyczące wykorzystanych metod rekurencyjnego podziału zbioru danych oraz reguły stopu. Dodatkowo zamieszczono informacje o odsetku błędnie sklasyfikowanych wartości zmiennej y .



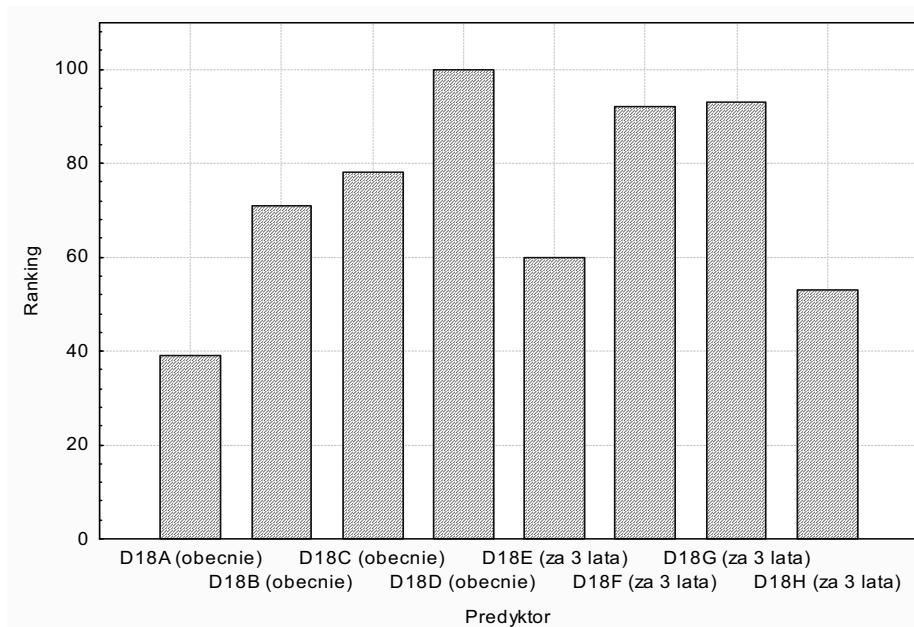
Rys. 1. Znaczenie handlu elektronicznego w posiadaniu innych jednostek podległych



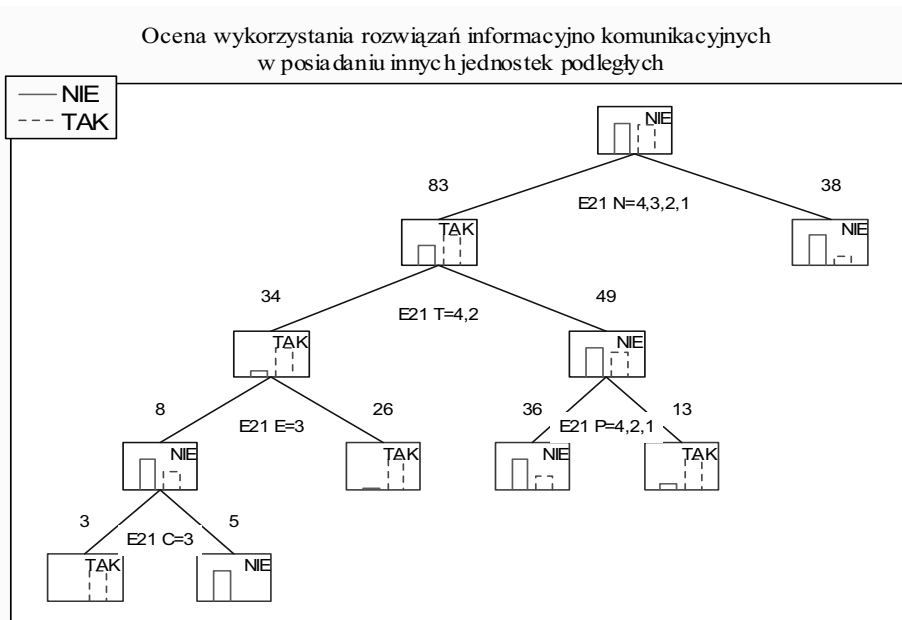
Rys. 2. Ranking znaczenia handlu elektronicznego w posiadaniu innych jednostek podległych



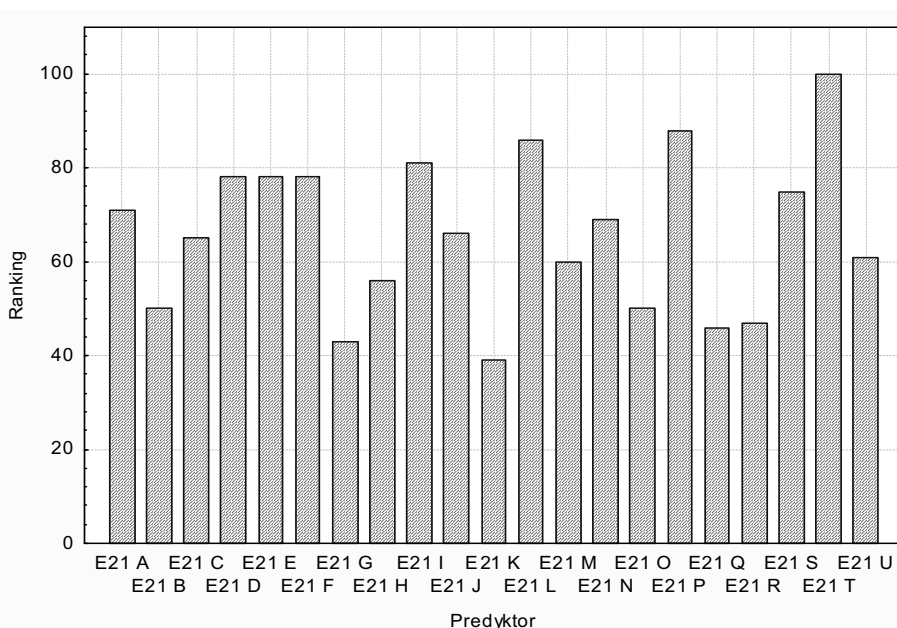
Rys. 3. Procesy transportowe w posiadaniu innych jednostek podległych



Rys. 4. Ranking procesów transportowych w posiadaniu innych jednostek podległych

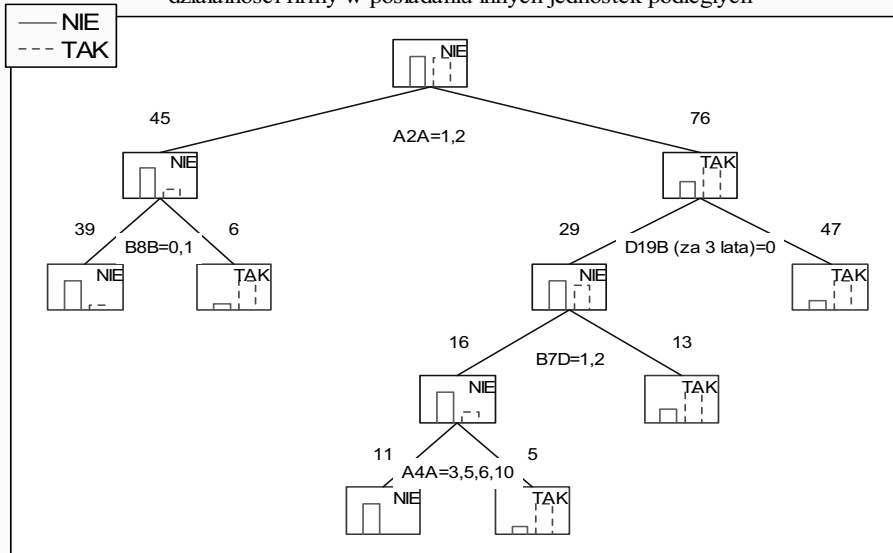


Rys. 5. Ocena wykorzystania rozwiązań informacyjno-komunikacyjnych w posiadaniu innych jednostek podległych

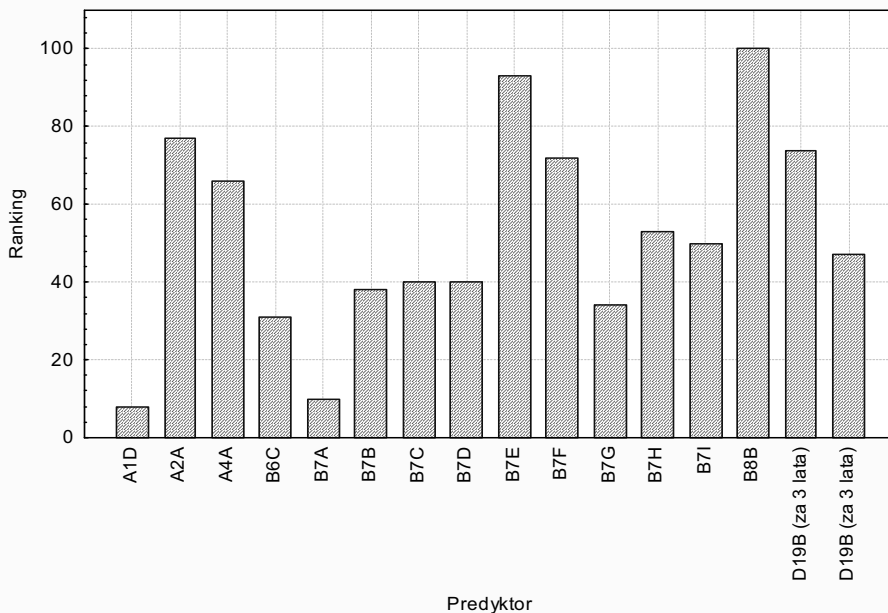


Rys. 6. Ranking oceny wykorzystania rozwiązań informacyjno-komunikacyjnych w posiadaniu innych jednostek podległych

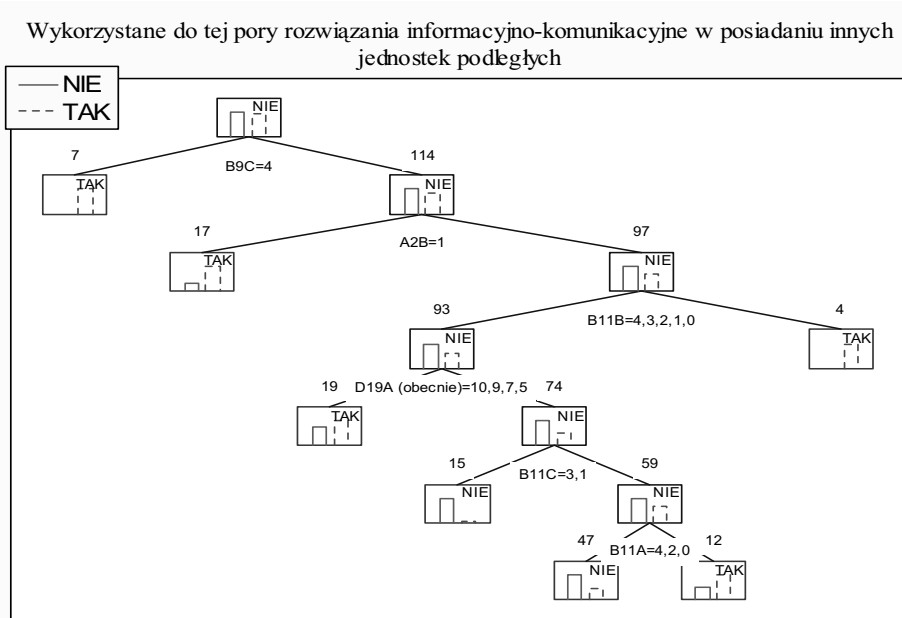
Planowanie zastosowania technologii informacyjno-komunikacyjnej w różnych obszarach działalności firmy w posiadaniu innych jednostek podległych



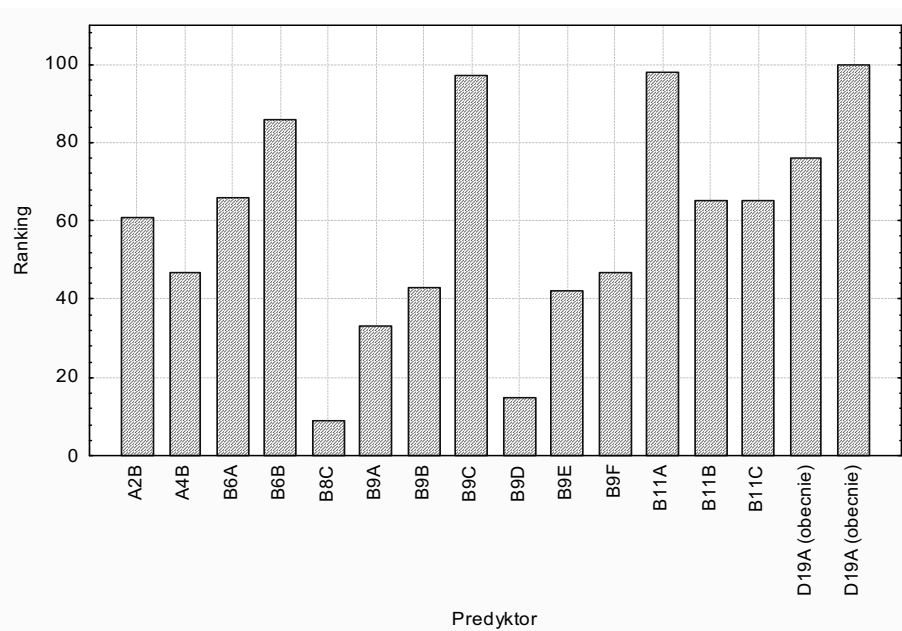
Rys. 7. Planowanie zastosowania technologii informacyjno-komunikacyjnej w różnych obszarach działalności firmy w posiadaniu innych jednostek podległych



Rys. 8. Ranking planowanego zastosowania technologii informacyjno-komunikacyjnych w różnych obszarach działalności firmy w posiadaniu innych jednostek podległych



Rys. 9. Wykorzystane do tej pory rozwiązania informacyjno-komunikacyjne w posiadaniu innych jednostek podległych



Rys. 10. Ranking wykorzystanych do tej pory rozwiązań informacyjno-komunikacyjnych w posiadaniu innych jednostek podległych

Tabela 3. Metody wyboru i błędy oceny wpływu poszczególnych czynników (zmiennych x) w posiadaniu innych jednostek podległych (zmienna y)

Rys.	Zmienne x	metoda	Reguła stopu	Błąd złej klasyfikacji	Globalny koszt	Odchylenie standardowe
Rys. 1.	Znaczenie handlu elektronicznego	Metoda C&RT wyczerpującego poszukiwania podziałów jednowymiarowych	Bezpośrednie zatrzymanie typu FACT Frakcja obiektów 0,25	0,51	0,51435	0,04488
Rys. 3.	Procesy transportowe	Metoda C&RT wyczerpującego poszukiwania podziałów jednowymiarowych	Reguła przycinania krawędzi przy błędzie złej klasyfikacji	0,44	0,44095	0,04388
Rys. 5.	Ocena wykorzystania rozwiązań informacyjno-komunikacyjnych	Metoda C&RT wyczerpującego poszukiwania podziałów jednowymiarowych	Reguła przycinania krawędzi przy błędzie złej klasyfikacji	0,44	0,44177	0,04305
Rys. 7.	Planowanie zastosowania technologii informacyjno-komunikacyjnych w różnych obszarach działalności firmy	Dyskryminacyjny podział jednowymiarowy dla predyktorów nominalnych i porządkowych	Reguła przycinania krawędzi przy odchyleniu standardowym	0,40	0,39202	0,04085
Rys. 9.	Wykorzystane do tej pory rozwiązania informacyjno-komunikacyjne	Metoda C&RT wyczerpującego poszukiwania podziałów jednowymiarowych	Bezpośrednie zatrzymanie typu FACT Frakcja obiektów 0,25	0,43	0,42961	0,04501

Interpretując „Znaczenie handlu elektronicznego” (rysunek 1) w kształtowaniu się wirtualnego łańcucha dostaw możemy powiedzieć, że największy wpływ na posiadanie jednostek podległych przedsiębiorstwa ma zmienna D20E (rysunek 2), która reprezentuje „stopień postrzegania handlu elektronicznego jako narzędzia walki z konkurencją”. Natomiast najmniejszy wpływ zmienna D20A, która reprezentuje „stopień postrzegania wpływu handlu elektronicznego na zysk firmy”.

Interpretując wyniki zaprezentowane na rysunku 1 możemy powiedzieć, że wśród 121, 40 przedsiębiorstw w większości posiada jednostki podległe. Wśród tych przedsiębiorstw są zarówno takie, które zgadzają się z opinią, że „handel elektroniczny ma kluczowy wpływ na zysk firmy” (D20A=5), są też jednak podmioty, które nie zgadzają się z tą opinią (D20A=1 lub D20A=2). Wśród 81 przedsiębiorstw, które nie mają zdania na temat wpływu handlu elektronicznego na zysk firmy, albo zgadzają się ze stwierdzeniem, że taki wpływ istnieje, większość nie posiada jednostek podległych. Wśród nich jednak można wyliczyć takie, które posiadają jednostki podległe. Są to między innymi przedsiębiorstwa, które w „handlu elektronicznym nie widzą możliwość pozyskania nowych dostawców” (D20F=2) oraz „nowych klientów” (D20C=1,2). Jednak w tej grupie są również przedsiębiorstw, które zgadzają się z opinią, że „handel elektronicznym ma wpływ na pozyskanie nowych klientów” (D20C=5) i takie, które nie mają zdania na ten temat (D20C=0).

Wyniki modelu (1) dla pierwszej grupy zmiennych można zinterpretować w ten sposób, że przedsiębiorstwa posiadające jednostki podległe (np. zakłady produkcyjne, magazyny, centra dystrybucji, punkty przeładunkowe, punkty sprzedaży detalicznej, itp.) charakteryzują się mniejszą skłonnością do korzystania z handlu elektronicznego. Przedsiębiorstwa te w większym stopniu korzystają z klasycznych kanałów dystrybucji. Natomiast przedsiębiorstwa o dużym obszarze oddziaływania, zarówno w obrębie kraju jak i za granicą, które nie posiadają jednostek podległych, wykazują większą skłonność do korzystania z rozwiązań opartych na handlu elektronicznym.

IV. PODSUMOWANIE

Podsumowując, drzewa klasyfikacyjne wykorzystano do podziału przedsiębiorstw ze względu na stosowane technologie oraz czynnik określający wirtualny łańcuch dostaw. Uzyskane podziały nie są jednak jednoznaczne i charakteryzują się dużymi błędami klasyfikacji. W związku z tym, aby w dalszej kolejności wykorzystać wyniki modelu (1) do zaproponowania przedsiębiorstwom działań umożliwiających efektywny rozwój wirtualnych łańcuchów dostaw, należy zastanowić się nad nowym doбором czynników kształtujących łańcuch. Tu warto byłoby rozważyć klasyfikację zmiennych, biorąc pod uwagę mniejsze odległości między zmiennymi i budować modele (1) w oparciu o mniejsze zbiory predyktorów.

BIBLIOGRAFIA

- Adamczyk M. (2005), Charakterystyka organizacji wirtualnej, „Gazeta IT”, 2005, nr 9 (39).
- Gatnar E. (1998), Symboliczne metody klasyfikacji danych, PWN, Warszawa.
- Gatnar E., Walesiak M. (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław.
- Jajuga K. (1993), *Statystyczna Analiza wielowymiarowa*, PWN, Warszawa.
- Trzpiot G., Ganczarek A. *The classification of risk on the Polish Power Exchange*, Zeszyty Naukowe Katedry Ekonometrii, red. nauk. Józef Dziechciarz, Wydawnictwo AE, Wrocław, w druku.
- Trzpiot G., Ganczarek A. *Metody klasyfikacyjne danych w analizach wirtualnych łańcuchów dostaw*, Praca przygotowana na Konferencję SwPGS 2007.

*Grażyna Trzpiot
Alicja Ganczarek*

**DECISION TREES IN STATISTICAL ANALYSIS –
A VIRTUAL SUPPLY CHAIN EXAMPLE****Abstract**

The aim of this article is to measure the influence of: IT and communication systems logistic competences and the transport-spedition-logistic branch on the development of virtual supply chain. Moreover, actions enabling the effective formation of virtual supply chains in the enterprises are proposed. Based on results of the survey of Silesian businesses having significant influence on regional development and the competitiveness of the economy, a statistical analysis of decisions taken in the field of logistics was carried out by applying decision trees technique.