PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 65
Ekonometria 25 2009

Grażyna Trzpiot, Alicja Ganczarek-Gamrot
University of Economics in Katowice

DECISION TREES FOR A VIRTUAL SUPPLY CHAIN

Summary: The aim of this article is to measure the influence of logistic competences and the transport on the development of a virtual supply chain for different groups of enterprises from the Silesian region, Poland. Moreover, we propose actions enabling the effective formation of virtual supply chains in the enterprise. Based on the results of the survey of Silesian businesses having significant influence on the regional development and the competitiveness of the economy, a statistical analysis of decisions taken in the field of logistics was carried out by applying the decision trees technique.

Key words: Silesian businesses, virtual supply chain, decision trees.

1. Introduction

In the new millennium we can observe the dynamic development of new information technologies and organizational environments. The Virtual Supply Chain (VSC) plays a huge role in these processes. The supply chain is an integrated business model for logistics management. It consists of the logistic and information elements ranging from the demands of the marketplace to the specific product delivery to the customer [Gunasekaran, Ngai 2004; Steward 1995]. VSC is based on business alliances using Internet technologies and has four phases: plan, source, make and deliver. Each phase has its own information system modules that are integrated with other phases of the virtual chain. The integration of the supply chain can be achieved by: communication, information system, trading and education, performance measures and rewards, and strategic planning [Gunasekaran, Ngai 2004].

The aim of this article is to measure the influence of logistic competences and the transport on the development of a virtual supply chain for different groups of enterprises from the Silesian region, Poland.

2. Methodology

A lot of empirical results show that the economic and sociological variables do not have normal distribution and are often described by nominal values, while distributions of variables often have outlier observations. Thus, it is often the case that we cannot apply the classic methods to the classification of the empirical variables. In

this paper, we apply the nonparametric method – the classification trees – to the classification of the influence of logistic competences and the transport on the development of the virtual supply chain. This method is based on the recursive partitioning of the m-dimension space X^m into homogenous subsets concerning the dependent variable y. When the dependent variable y is nominal, equation (1) is a classification tree [Breiman et al. 1984; Gatnar, Walesiak 2004]:

$$y = \sum_{k=1}^{K} a_k I\{\mathbf{x_i} \in R_k\} \tag{1}$$

where: $\boldsymbol{x}_i$ – multivariate variable, element of X^m,

R_k – $k = 1, \ldots, K$, are disjoint regions in the m-dimensional feature space, segment of X^m,

a_k – the parameters,

$I(q)$ – an indicator function: $I(q) = \begin{cases} 1 & \text{if } q \text{ is true} \\ 0 & \text{if } q \text{ is false} \end{cases}$.

We can divide the measures of the quality of model (1) by homogenous measures and impurity measures, but the second group of measures is used much more often than the first one. The first group of measures is based on the estimate of the similarity degree of variables inside one segment. The second group of measures is based on the estimate of the degree of differences between all segments of $R_{k.}$. In our analysis we used three measures:

- Gini volatility coefficient [Breiman et al. 1984]:

$$H(S_k) = 1 - \sum_{j=1}^{J} p^2(j/k), \tag{2}$$

- Chi-squared statistics:

$$\chi^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(N_j(k) - E_j(k))^2}{E_j(k)}, \tag{3}$$

- G-squared statistics based on Likelihood quotient [Mingers 1989]

$$G = 2N\ln(2)\Delta E(S, x_m), \tag{4}$$

where: $\Delta E(S) = E(S) - \sum_{k=1}^{K} E(S_k)p(k)$, $S_k = \{x_i, i = 1,\ldots m: \quad x_i \in R_k)$,

J – a number of different values of dependent value y (j = 1,... J),

K – a number of disjoint subsets of set S ($S = S_1 \cup S_2 \cup \ldots \cup S_K$; k = 1,... K),

N – the size of a sample ($N = \sum_{j=1}^{J} N_j = \sum_{k=1}^{K} N_j(k)$),

$E_j(k)$ – an expected number of y equal j in k^{th} subset of $S\left(E_j(k) = \frac{N_j \times N(k)}{N} \right)$,

p(j/k) – a probability that the dependent variable y equals j is in segment R_k, where j is the most frequent value of y.

The aim of equation (1) is, on one hand, to create homogenous groups of variables and, on the other hand, to build a tree that is simple and easy to interpret. Thus, the question is when we should stop the recursive partitioning of the m-dimension space X^m.

The methods that are the most frequently used to do it are:

- Cost-complexity pruning method:

$$e(D) = \frac{1}{n} \sum_{i=1}^{n} I\{D(x_i) \neq y_i\}. \tag{5}$$

- One Standard Error rule (1SE rule) [Breiman et al. 1994, p. 69]:

$$D^{**} = \arg\min_{W(D^*)} \{e^T(D^*) + S(e^T(D^*))\}, \tag{6}$$

where D^* is the smallest tree, which was estimated based on a testing sample or cross validation.

- FACT (Fast Algorithm for Classification Trees).

3. The data set

The paper presents the classification of data obtained in the survey of 112 Silesian enterprises. The survey was aimed at production companies, trading companies and service providers. The main objective was to classify the factors determining the virtual supply chain consisting of communication systems, logistic competences and the transport-spedition-logistic branch.

We used the quantitative variables, such as the number of employees at the end of 2005, turnover and percentage of domestic and foreign sales, characterized by huge volatility and a very strong right skewness. Moreover, the data set contains outliers. The data do not have normal distribution.

Figure 1 shows that the classification of sectors depends on the above variables. In order to present different groups of respondents, we chose the C&RT recursive

partitioning method, FACT rule and G-measure of classification errors. Table 1 (in the next section) presents the errors of the recursive partitioning of all enterprises.

Production companies, for example, can be divided into three groups. One of them represents the enterprises with more than 500 employees (variable G3 is presented in a soft scale) and less than a PLN 1bn turnover at the end of 2005 (variable G4 is presented in a soft scale). The second group of production companies represents the enterprises with less than 500 employees (G3), enterprises which do not use the virtual exchange (M15C) and the percentage of sales outside Poland but inside the EU was higher than 5.5% (M10B) and outside the European Union it was not higher than 7.5% (M10C). The last group of production companies employed less than 500 people (G3) and their sales outside Poland but still inside the EU were lower than 5.5% of total sales (M10B). These companies did not use bar codes (M15F) and their turnover was higher than PLN 40m (G4).

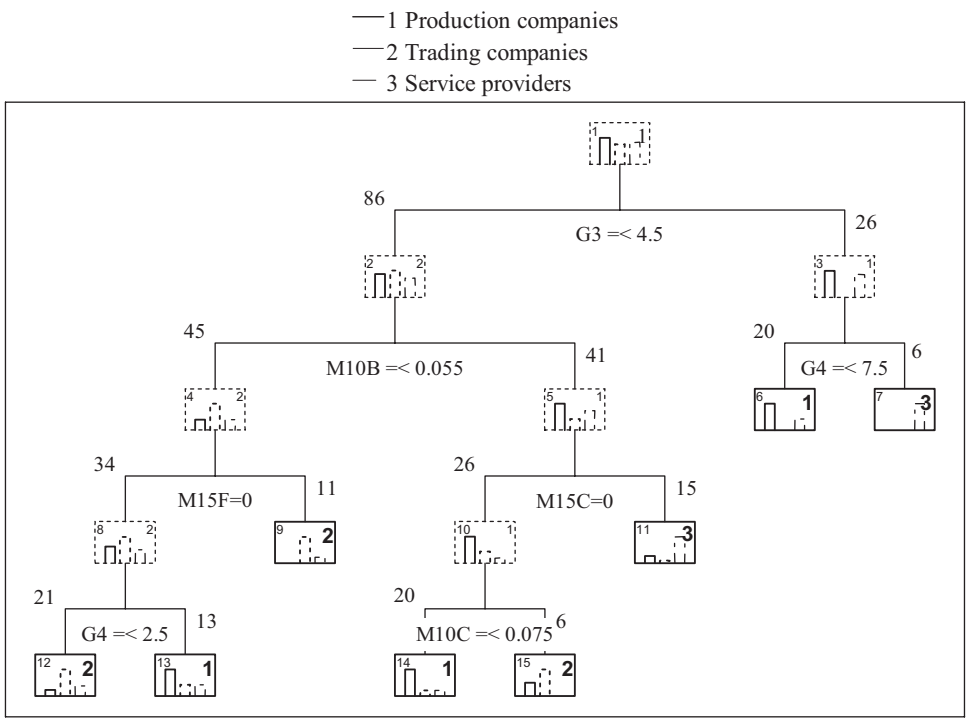

Fig. 1. Classification of sectors

Source: own calculation in STATISTICA.

The information about the opinion on logistic competences and the transport and their influence on the development of the virtual supply chain is described in the Likert scale (Likert 1932 Ph.D on Columbia University).

The opinions about various logistic competences and transport are very similar in this group of enterprises and they were generally good. But a huge group of respondents do not have a clear opinion. A very low percentage of enterprises have a negative opinion about logistic competences and transport, and they are very careful about giving strong opinions on this topic.

Figure 2 shows the results of the classification of enterprise sectors computed using opinions about logistic competences and transport. To present different groups of respondents we chose the C&RT recursive partitioning method, FACT rule and G-measure of classification errors. Table 1 (in the next section) presents the errors of the recursive partitioning of all enterprises.

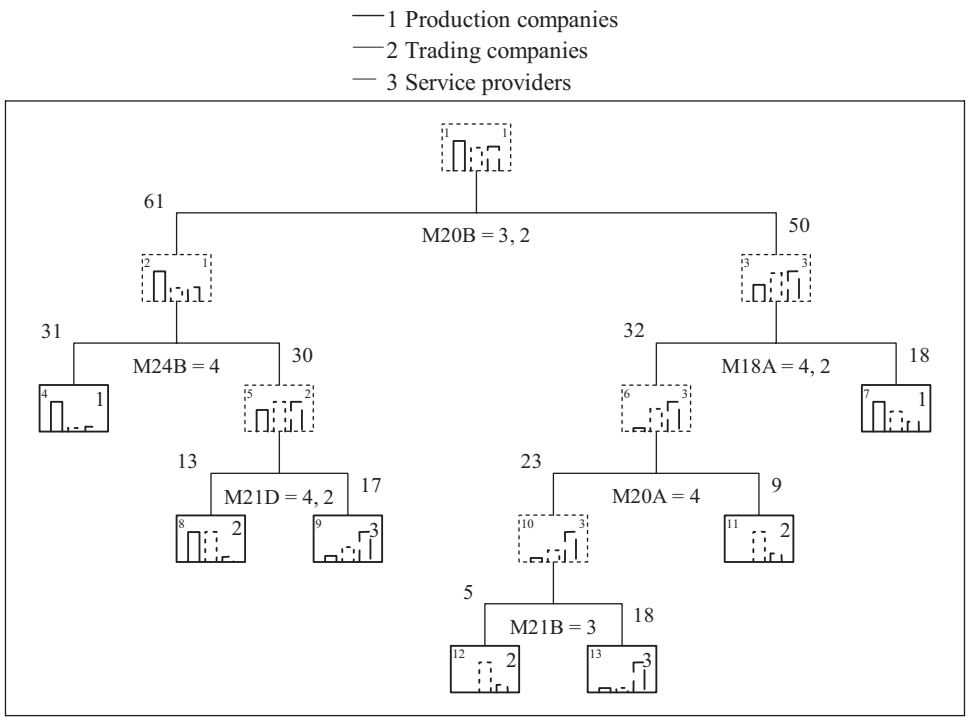

Fig. 2. Classification of sectors

Source: own calculation in STATISTICA PL.

Production companies can be divided into two groups. The respondents from one group disagree or neither agree nor disagree with the opinion that their enterprise is well prepared for potential problems inside the company (M20B). They have a positive opinion about the availability of the production and business infrastructure in Poland (M24B). The other group of respondents, which has a good opinion about their readiness for potential problems inside the company (M20B), has no opinion about regular monitoring of logistic costs and logistic effect (M18A).

The trading companies can be divided into three groups. The respondents from the first group disagree and neither agree nor disagree with the opinion that their enterprise is well prepared for potential problems inside the company (M20B). They have positive and negative opinions about the availability of production and business infrastructure in Poland (M24B) and partnership with suppliers and customers (M21D).

The second group of respondents, which has a good opinion about their readiness for potential problems inside the company (M20B), has a different opinion about the regular monitoring of logistic cost and logistic effect (M18A). This group of respondents can be divided into two groups. One group, which has various opinions about the information flow inside the company (M20A) and the other one, which has a good opinion about it (M20A) and has no opinion about their readiness for potential problems outside the company (M21B).

The service providers can be divided into two groups. The respondents from one group disagree and neither agree nor disagree with the opinion that their enterprise is well prepared for potential problems inside the company (M20B). They have positive and negative opinions about the availability of production and business infrastructure in Poland (M24B) and have no opinion about partnership with suppliers and customers (M21D). The other group of respondents, who have good opinion about their readiness for potential problems inside the company (M20B), have various opinions about the regular monitoring of logistic cost and logistic effect (M18A). This group of respondents has a good opinion about the information flow inside the company (M20A) and different opinions about their readiness for potential problems outside the company (M21B).

4. Empirical analysis

We analyzed model (1) where the influence of logistic competences and the transport logistic branch are the dependent variables y. In model (1) we took into consideration the qualitative and quantitative variables which we recognized as factors describing the development of the virtual supply chain. We used the C&RT (Classification and Regression Trees) recursive partitioning method proposed by Breiman [Breiman et al. 1984] with three different measures of volatility between groups and available in the STATISTICA PL package. To stop the recursive partitioning, we used three pruning methods: cost-complexity pruning, one Standard Error (1SE) rule and FACT-Fast Algorithm for Classification Trees. The choice of the best results of our classifications is presented in Table 1 and in Figures 3-6. We chose the following dependence variable y for model (1):

I) The opinion about logistic competences compared with competitors:

- M17A – The company can reduce time between order and delivery.
- M17F – The company can adapt to customer needs regarding terms of delivery:

1. The company is much worse than competitors.

2. The company is worse than competitors.
3. The company is not worse or better than competitors.
4. The company is better than competitors.
5. The company is much better than competitors.

II) The opinion about logistic activity of the company:

- M18D – The company regularly monitors and values benefits of logistic activity:
  1. The company very seldom monitors and values benefits of logistic activity.
  2. The company seldom monitors and values benefits of logistic activity.
  3. The company monitors and values benefits of logistic activity neither seldom nor regularly.
  4. The company regularly monitors and values benefits of logistic activity.
  5. The company very regularly monitors and values benefits of logistic activity.

III) The opinion about internal and external co-operation in logistic activity:

- M20C – The company uses an information system to facilitate decision making in an effective way:
  1. Strongly disagree.
  2. Disagree.
  3. Neither agree nor disagree.
  4. Agree.
  5. Strongly agree.

Table 1. Classification errors

Fig.	Recursive partitioning method	Measure	Pruning	Classification error	Cross--validation	Standard deviations
1	C&RT	G	FACT 0,2	0.5714	0.5714	0.0468
2	C&RT	G	FACT 0,2	0.5625	0.5676	0.0470
3	C&RT	G	FACT 0,05	0.5268	0. 5315	0.0473
4	C&RT	χ^2	FACT 0,3	0.4286	0.4324	0.0470
5	C&RT	χ^2	1SE rule	0.3125	0.3333	0.0447
6	C&RT	$H(S_k)$	FACT 0,2	0.4554	0,5315	0.0474

Source: own calculation.

We estimated six models given by equation (1). The two first models (Figure 1 and Figure 2) are described in the section above. The next four models (Figure 3-6) describe the influence of particular factors on logistic competences of the companies researched. We used 6 ordinal predictors and 35 nominal predictors from

the independent factors describing the development of the virtual supply chain. Table 1 presents classification errors. In our analysis we wanted to identify factors which have an influence on logistic competences of the company. To do it, we tried to find answers to a few questions. The first question is what kind of factors has an influence on the opinion that the company can reduce time between order and delivery. We present the result of our analysis on the decision tree (Figure 3). Based on this result, we can say that the company is worse than competitors in reducing time between order and delivery (y = 2), if the company has problems with meeting the terms and quantity of the delivery (M17B) and in many cases seldom (but in a few cases very regularly) monitors an influence of logistic activity outside the company (M18E).

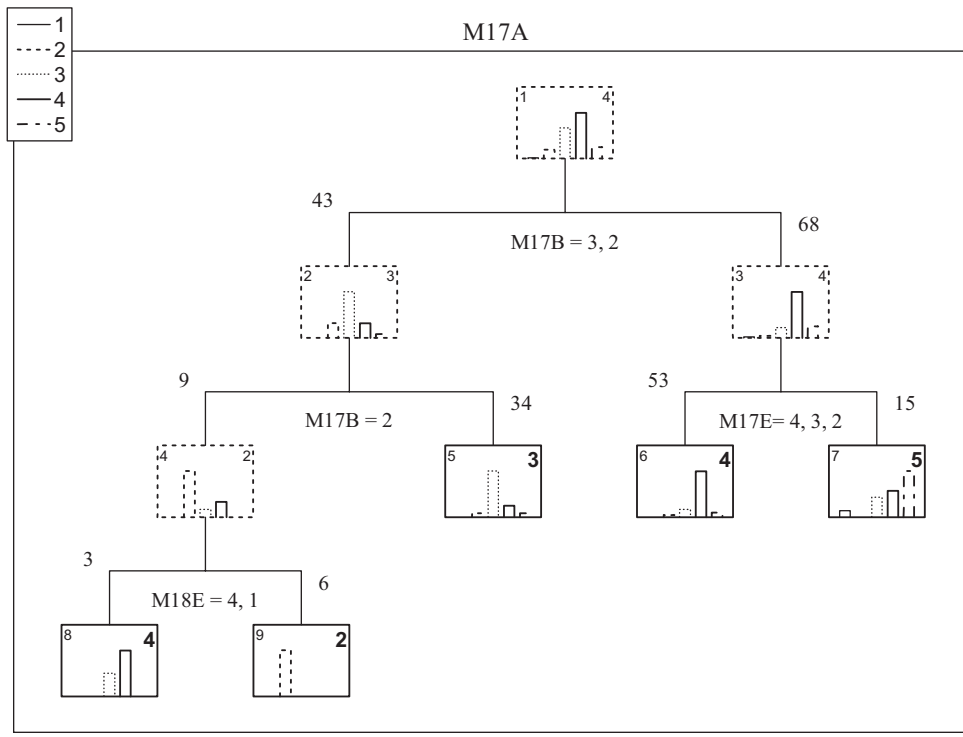

Fig. 3. The ability to reduce time between order and delivery

Source: own calculation.

The companies which have problems with meeting the terms and quantity of the delivery (M17B), but in many cases frequently (but in a few cases very seldom) monitor an influence of logistic activity outside the company (M18E), are better than competitors in reducing time between order and delivery (y = 4).

The companies are better than competitors in reducing time between order and delivery (y = 4), if they have no problem with meeting the terms and quantity of the delivery (M17B) and much better than competitors (y = 5) if they can regularly update terms and quantity of the delivery (M17E).

The next question is what kind of factors has an influence on the opinion that the company can adapt to customer needs regarding terms of delivery. We present the result of our analysis on a decision tree (Figure 4).

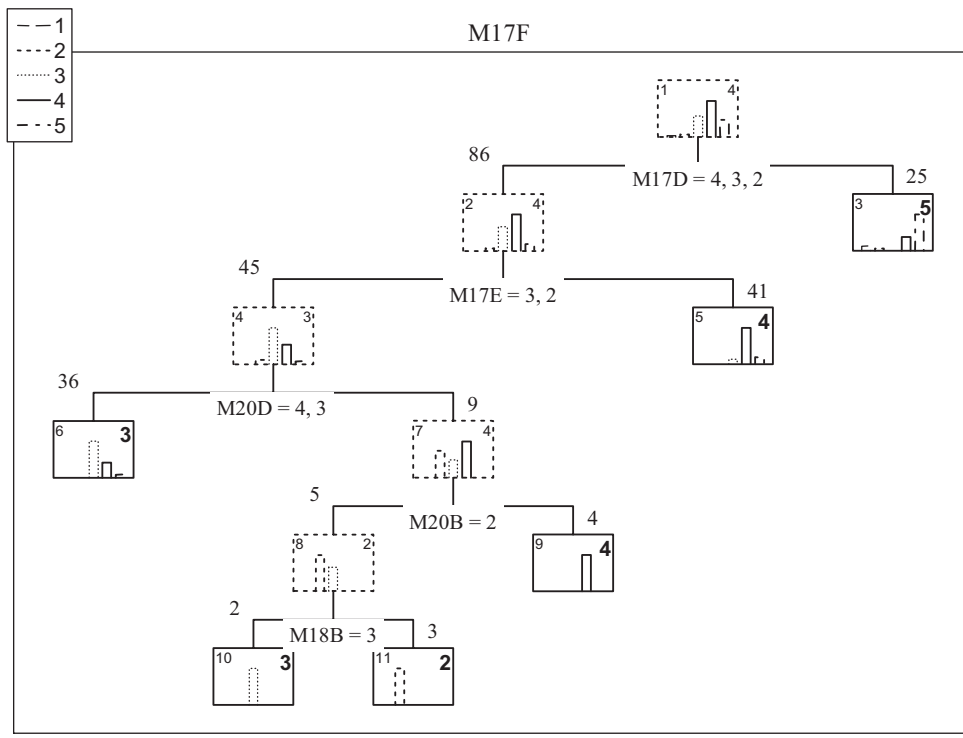

Fig. 4. The ability to adapt to customer needs

Source: own calculation.

Based on this result, we can say that the company is worse than competitors in adapting to customer needs regarding terms of delivery (y = 2), if the company has a problem with maintaining contact with customers (M17D) and updating terms and quantity of delivery (M17E). Additionally, these companies have a poor system of strategic planning (M20D) and preparation for potential problems (M20B) and neglect the regular monitoring and evaluation of logistic costs and activity of their suppliers and customers (M18B).

The companies are better than competitors in adapting to customer needs regarding terms of delivery (y = 4), if they have no problems with factors described

above and much better than competitors (y = 5) if they have no problems with maintaining contact with customers (M17D).

The next question is what kind of factors has an influence on the opinion that the company regularly monitors and values benefits of logistic activity. We present the result of our analysis on a decision tree (Figure 5). Based on this result, we can say that the monitoring of the influence of logistic activity outside the company (M18E) is the most important factor which influences the regular monitoring and measuring of the benefits of logistic activity inside the company. If the companies neglect monitoring logistic activity outside the company, they neglect it inside the company (y = 1 or y = 2) and, on the contrary, if they regularly monitor logistic activity outside the company, they also do it inside the company (y = 4 or y = 5).

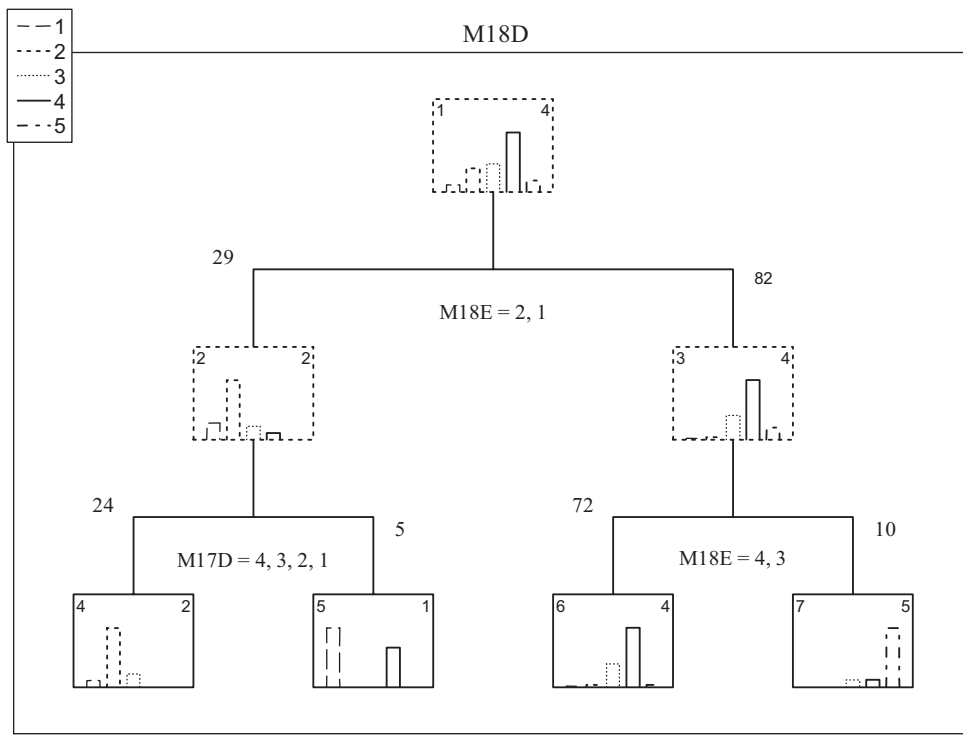

Fig. 5. The monitoring of logistic activity

Source: own calculation.

The last question is what kind of factors has an influence on the opinion that the company uses an information system to facilitate decision making in a correct way. We present the result of our analysis on a decision tree (Figure 6). The companies which have a poor system of strategic planning (M20D) do not have any opinion on

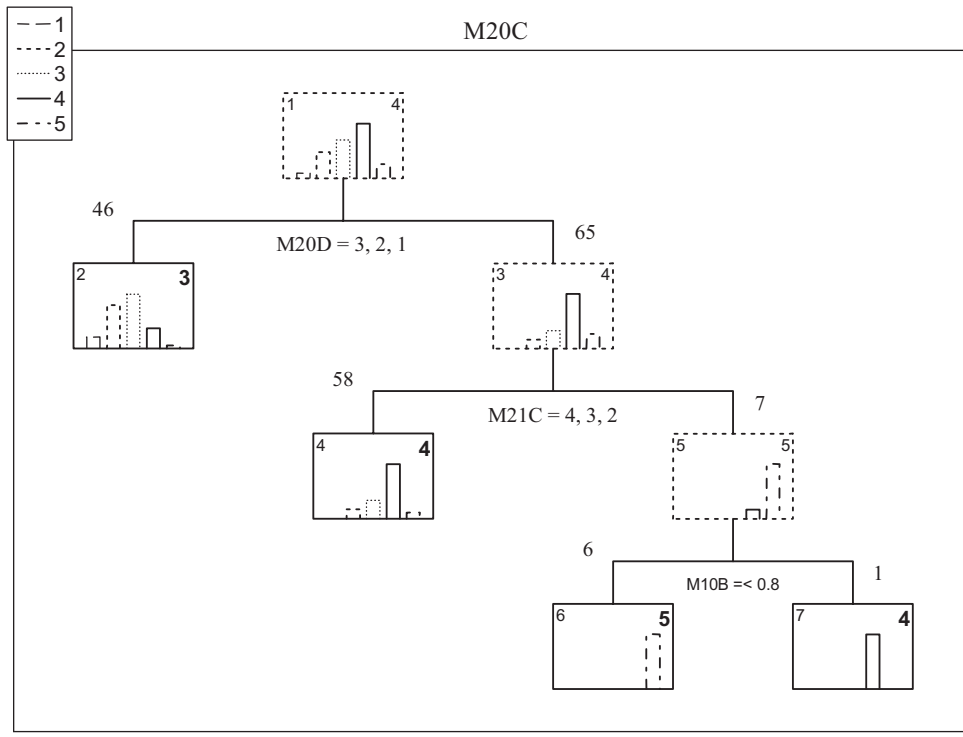

Fig. 6. The use of an information system in the decision making process

Source: own calculation.

this issue (y = 3). So the group of companies which have a good system of strategic planning (M20D) have a good opinion on this issue (y = 4 or y = 5). Additionally, if the company uses an information system to facilitate contacts with suppliers and customers (M21C) and their percentage of sales outside Poland but inside the EU is lower then 80% (M10B), the company strongly agrees with the opinion that an information system provides operations managers with the appropriate information to facilitate decision making (y = 5).

5. Conclusion

The aim of this article is to measure the influence of logistic competences and the transport logistic branch on the development of a virtual supply chain for different groups of enterprises from the Silesian region, Poland. We used decision trees to divide the Silesian enterprises taken into consideration: communication, information system, trading, performance measures and rewards and strategic planning, which are described by quantitative and qualitative variables. As a result, we obtained two

trees with the division of the enterprises into three sectors dependent on different levels of logistic competences and general characteristics of these sectors. Based on the result of Figure 1, we can say that the service providers are the most developed, production companies rank next and the last ones are trading companies, but the opinion about logistic competences in every group of enterprises is very similar (Figure 2).

Based on the classification of logistic competences (Figure 3-6), we can divide these enterprises into two groups: one with poor logistic conditions and the other one with good logistic conditions. But in every example of the classification, the actual logistic conditions are less important than the opinion about them. For example, if we look at an information system in these companies, we can say that it is starting to develop, so, perhaps, the answer is that at the moment the information system is not so important in the development of the virtual supply chain.

Variables are assigned to groups with respect to the level of correlation between them and y. So, the classification in decision tree can be analyzed from the causality perspective [Gatnar 2003]. Based on results of our paper in economical sense we can encourage enterprises to develop logistic competences gradually. They should develop one leading of competence, for example: communication, information system, trading, performance measures and rewards and strategic planning. The choice of one competence should help to develop all of them.

In our paper we use nonparametric classification trees method, because this method − as opposed to classical methods − as it does not require assumptions on variable distributions, it can be used for variables described by nominal values, it can be used for non observational variables and variables with outlier observations and it is robust on added variables [Gatnar, Walesiak 2004]. But this method has disadvantages too. First of all this method is not objective. The result of recursive classification depends on the choice of: recursive partitioning method, measures of volatility between groups, pruning methods. Special care should be taken when interpreting these results due to small sample size and low variability of variables.

Literature

Breiman L., Friedman J., Olshen R., Stone C., *Classification and regression trees*, CRC Press, London 1984.

Gatnar E., *Statystyczne modele struktury przyczynowej zjawisk ekonomicznych*, AE, Katowice 2003.

Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.

Gunasekaran A., Ngai E.W.T., *Virtual supply-chain management*, "Production, Planning&Control" 2004, vol. 15 no 6, p. 584-595.

Mingers J., *An empirical comparison of pruning methods for decision tree induction,* "Machine Learning" 1989, no 4.

Steward G., *Supply chain performance benchmarking study reveals keys to supply chain excellence,* "Logistics Information Management" 1995, 8(2), p. 38-44.

DRZEWA DECYZYJNE W ANALIZIE WIRTUALNYCH ŁAŃCUCHÓW DOSTAW

Streszczenie: Celem artykułu jest określenie stopnia oddziaływania kompetencji logistycznych przedsiębiorstw województwa śląskiego na rozwój wirtualnych łańcuchów dostaw, a także 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 zauważalne, za pomocą drzew decyzyjnych przeprowadzono statystyczną analizę decyzji wyboru najkorzystniejszych działań z punktu widzenia poziomu rozwoju technologii informacyjno-komunikacyjnej oraz rozwiązań logistycznych.