VERSITA

BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 1

Sofia • 2013

Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2013-0008

Using Decision Trees for Identification of Most Relevant Indicators for Effective ICT Utilization

Silvana Tomic Rotim[1], *Jasminka Dobsa*[2], *Zdravko Krakar*[1,2]

[1] *ZIH d.o.o., Mazuranic Square 8, Zagreb*
[2] *Faculty of Organization and Informatics, Pavlinska 2, Varazdin*
Emails: stomic@zih.h jasminka.dobsa@foi.hr zdravko.krakar@foi.hr

Abstract: *This paper offers a brief overview of the research of ICT utilization and benefits of its usage. The results of several important studies conducted worldwide are presented. One of them is a study by the World Economic Forum that we use as the basis of our research. This study covers 134 countries, NRI (Network Readiness Index) is used as a parameter to distinguish the readiness of different countries to adopt ICT. NRI consists of 68 indicators that are organized into three groups. Each group describes one of the three main factors crucial for effective utilization of ICT: Environment, Readiness and Usage. The observed countries are divided into four groups (leaders, followers, league III and league IV) and classification by a decision tree is conducted. The decision tree method is applied to each of the three main factors and the results are presented by means of* F_1 *measure.*

Keywords: *NRI-Network Readiness Index, decision trees, effectiveness of ICT utilization.*

1. Introduction

There is no doubt that these days informatization is a global process of great importance for each country and its activities, as well as for each business system. Thanks to its rapid development, ICT has been exerting a pervasive influence on our work and life. Investments in ICT, although colossal, are constantly growing. However, research and practice indicate that the effects of ICT are often undersized and do not meet the expectations. Research in this area is therefore critical in order to attain the necessary knowledge that would facilitate the increase of these effects. This phenomenon is evident at global and macro level, as well as on business systems level.

On this background this paper discusses the research by the World Economic Forum that was conducted for 134 countries. This study includes the data for 68 indicators whereby a country's readiness to adopt ICT is assessed. In [13] a clustering of the European countries is conducted, based on the indicators observed in the research by the World Economic Forum. In order to identify the relevant indicators for effective utilization of ICT, analysis of the variance for the obtained clusters is conducted.

The impact of ICT investment on individual national economies is extensively covered in literature. The studies concerning developed countries, like USA [3], Belgium [8], France [2], etc., show that the investment in ICT is highly correlated with GDP and that it contributes to the increase of productivity and prosperity of a country.

On the other hand, the situation in the developing countries, where the issue of ICT investment started to be considered fairly recently, is quite different. One of the studies covering this subject, conducted in Poland, is discussed in [12]. It indicates lack of research that would deal with the contribution of ICT to growth and productivity in post-communist transition economies. A similar study on China [7] has shown that the relationship between productivity growth and ICT capital had been examined and the estimation of the return on ICT investments has been made. Studies concerning those issues in Croatia [9] have yielded results similar to those in [7] and [12]. In general, it has been shown that the developing countries are using the ICT potential poorly, so the correlation between the investments in ICT and GDP for these countries is rather low.

However, the existing studies suggest that this relationship is more complex since the effective usage of ICT and the benefits that companies and countries can have, depend on a number of factors of the global environment in which ICT is utilized, not solely on the amount of the investment in ICT. Certain factors, such as social, economic and legal environment, the level of technological infrastructure, etc., are most important. It is therefore necessary to continuously and persistently build a positive environment for the penetration of ICT into the national economy and the public sector.

There are several ways to categorize the success of particular countries in using ICT, the most common among which are: NRI, E-readiness, MultiFactor of Productivity (MFP), analysis of the value chain, etc.

Their application makes it possible to determine the performance order of the individual countries in using the potential of ICT on the global level. Based on a previous research we divided 134 countries observed into the following leagues according to Network Readiness Index (NRI): leaders ($NRI \geq 5.1$, 20 countries), followers ($4.2 \leq NRI < 5.1$, 22 countries), league III ($3.3 \leq NRI < 4.2$, 56 countries) and league IV ($NRI < 3.3$, 36 countries).

It is possible to influence the speed and the effects of ICT diffusion by making an impact on a country's attitude towards the development of maturity in ICT. In the diffusion of ICT there are many implementation barriers and appropriate policies on a national level to be developed, that are the catalyst of this diffusion. Therefore, ICT national policy makers must have awareness of their roles and

responsibilities so as to be able to devise clear guidelines and incentives to affect the speed and efficiency of this diffusion.

2. Decision trees

A decision tree is a rule for predicting the class of an example based on the values of its predictor variables. It this research we use the CART (Classification And Regression Tree) method developed by B r i e m a n et al. [1]. A good introduction to the theory of decision trees based on Quinlan's ID3 system can be found in M i t c h e l l [11]. A tree is constructed using training examples. A test example is sorted through a tree from a root to some terminal node and a class of the terminal node is assigned to the example. The class of the terminal node t is j_0 if the proportion of class j_0 examples in the terminal node $P(j_0 \mid t)$ maximizes the proportion of j class examples in node $t\, P(j \mid t)$, where j goes through the set of all classes. Each node in a tree specifies a test of some variable based on the condition of the form $X \le c$. The tree is split according to this condition: the examples that satisfy the condition are on the left, while the examples which do not satisfy the condition are on the right. The tree algorithm searches through the set of all variables at each node and finds the best split for each variable. Then it compares the splits for all variables and finds the best ones. Generally, the aim is to select a split at each node so that the data in each descendant node is purer than the data in the parent node.The impurity of the node is largest when all classes are equally mixed together and smallest when the node contains the data from only one class. The variable used in a splitting condition is selected according to the splitting criterion in order to maximize the decrease in impurity. As a splitting rule, or the criterion according to which the best variable is selected, we use the Gini diversity index which is expressed as

$$i(t) = \sum_{j \neq i} P(j \mid t) P(i \mid t).$$

The Gini diversity index is interpreted as the probability of misclassification;thereby the goodness of the split criterion is to execute a split at any node at which the probability of misclassification is most reduced.

A characteristic of the classification trees is that, if no limits are placed on the number of splits that are performed, pure classification will be eventually achieved with each terminal node containing only one class of examples. However, such pure classification is usually unrealistic, especially in cases of noisy data. One of the controlling options when splitting stops is to allow the splitting to continue until all terminal nodes are pure or contain no more examples than a specified minimum fraction of the examples from other classes. This option is referred to as the FACT style direct stopping criteria [10].

The purpose of a classification tree is not only to obtain accurate classification of the examples but also to provide an insight into the predictive structure of the data. The variables appearing in the splitting criteria near the top of a decision tree are most important for data discrimination. However, we might be interested in

ranking all variables according to their importance, not only those appearing in the splitting criteria. Namely, some variables do not yield the best split at the node, but may still give the second or third best split. The purpose of a measure of importance is therefore to recognize the importance of those variables which may be masked by others. The measure of importance of the variable x_m $M(x_m)$ is defined by using the so called surrogate splits [1]. Here the measure of importance of a variable is normalized to fit within the range from 0 up to 100 by the formula

$$100\, M(x_m) / \max_m M(x_m) .$$

3. Experiment

3.1. Data description

In order to classify the observed countries by NRI, applying the decision tree method, it was necessary to design and create a database with attributes and particular parameters according to which the readiness to use ICT was examined. As the main source for design of this database the World Economic Forum research was used. The original research, entitled "The Global Information Technology Report 2008-2009" [4], covers 134 countries. Our research includes data for 68 indicators according to which the readiness of some countries to adopt ICT is assessed. These indicators are grouped into three main groups and 9 subgroups (Fig. 1). In the calculation of NRI index, all groups and subgroups are given the same weight. This reflects the assumption that all of them provide similar contribution to the overall NRI.

ICT cannot be developed and used efficiently in a vacuum. Factors of environment, readiness and usage are therefore very important. The Environment group considers the friendliness of a country's environment for ICT development through 30 indicators grouped into three different subgroups. The second group, Readiness, is related to the extent, to which a country's main stakeholders are interested and prepared to use technology in their everyday activities. It is measured through 23 different indicators. The last component of NRI, Usage, gauges the actual usage of ICT, with a particular focus on the impact of ICT in terms of efficiency and productivity gains. This group consists of 15 indicators.

For each of the 134 countries observed in [4], key indicators, such as population, per capita GDP, percentage of Internet users, bandwidth Internet connection in Mb/s and percentage of mobile phone users, are listed.

3.2. Results of experiment

The observed data and ranking of the countries show that Europe has a relevant position indeed in the global network map, since 12 out of the top 20 performers are from Europe, with Denmark and Sweden at the very top. According to NRI, it is possible to recognize the following leagues: leaders (NRI $\geq$ 5.1, 20 countries), followers (4.2 $\leq$ NRI < 5.1, 22 countries), league III (3.3 $\leq$ NRI < 4.2, 56 countries) and IV(NRI < 3.3, 36 countries).Our experiment was conducted applying Wolfram

Research Statistica 10.0 software. We applied the decision tree method using the aforementioned classification into four leagues as a fact. Hereafter we present the results of the decision trees for three main factors (Environment, Readiness, Usage) by analyzing the structure of the obtained trees and the importance of the indicators.

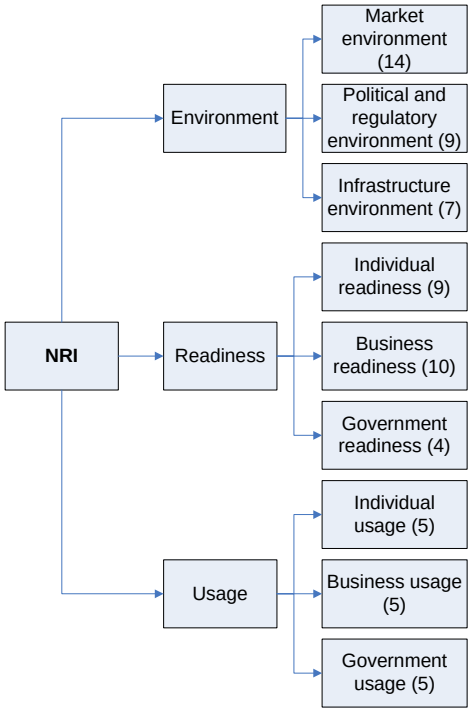

Fig. 1. NRI structure [4]

Table 1 shows the importance of all indicators measured within the Environment factor (average value for tree folds). The six most important indicators (corresponding to the number of splits in the decision tree) are marked in bold.

The example of the decision tree for the Environment factor (for one of the folds, and with the stopping criterion FACT=0.15) is shown in Fig. 2. At the top of the tree the countries are divided by the criterion related to the "Accessibility of digital content" indicator, which is one of the two most important indicators according to Table 1. Three indicators selected by the splitting rule in the decision tree correspond to the six most important indicators listed in Table 1. By means of the decision tree the rules for classification into each one of the four classes (leaders, followers, league III and IV) are defined. For example, the rule for classifyinga country into the leaders' class is:

If ("Accessibility of digital content" >5.815) and ("Quality of companies in the ISP sector" >4.9) then class=leader.

For the league III class there are tree terminal nodes. The rule for classifying a country into that class is:

If (("Accessibility of digital content" ≤ 5.815) and ("Intellectual property protection" ≤ 3.91) and ("Laws relating to ICT" >3.365))

or (("Accessibility of digital content" ≤ 5.815) and ("Intellectual property protection" ≤ 3.91) and ("Laws relating to ICT" ≤ 3.365) and ("Utility patents" > 0.365))

or (("Accessibility of digital content" ≤ 5.815) and ("Intellectual property protection" >3.91) and ("Intensity of local competition" ≤ 5.01)) then class=league III.

Table 1.Importance of different indicators for the Environment factor

Importance of the indicator	Indicator
91.33	**Venture capital availability**
80.67	**Financial market sophistication**
88.00	**Availability of latest technologies**
67.67	State of cluster development
80.00	**Utility patents**
32.67	High-tech exports
27.33	Burden of government regulation
22.67	Extent and effect of taxation
21.00	Total tax rate
30.33	Time required to start a business
20.67	Number of procedures required to start a business
68.67	Intensity of local competition
40.67	Freedom of the press
98.33	**Accessibility of digital content**
44.67	Effectiveness of law-making bodies
98.33	**Laws relating to ICT**
62.33	Judicial independence
77.67	Intellectual property protection
62.67	Efficiency of legal framework for disputes
75.67	Property rights
62.67	Quality of competition in the ISP sector
24.67	Number of procedures to enforce a contract
24.33	Time to enforce a contract
67.33	Telephone lines
75.00	Secure Internet servers
72.00	Electricity production
50.33	Availability of scientists and engineers
67.67	Quality of scientific research institutions
44.33	Tertiary enrolment
31.00	Education expenditure

Fig. 3 shows the decision tree for the second factor – Readiness (for one of the folds, and with the stopping criterion FACT= 0.15). It has six splits and seven terminal nodes. The decision tree shows that the indicator "Internet access in schools" is most relevant for classifying different countries regarding their readiness for effective ICT utilization. Table 2 shows all measured indicators pertaining to the Readiness factor, and the six most important ones (corresponding to the number of splits in the decision tree) are marked in bold. Four of the six most important indicators are used as a splitting criterion in the decision tree. Two indicators used in the decision tree as splitting criteria, which are not among the six most important ones according to Table 2, are on the lower tree levels (i.e., the third and fourth level of the right subtree).

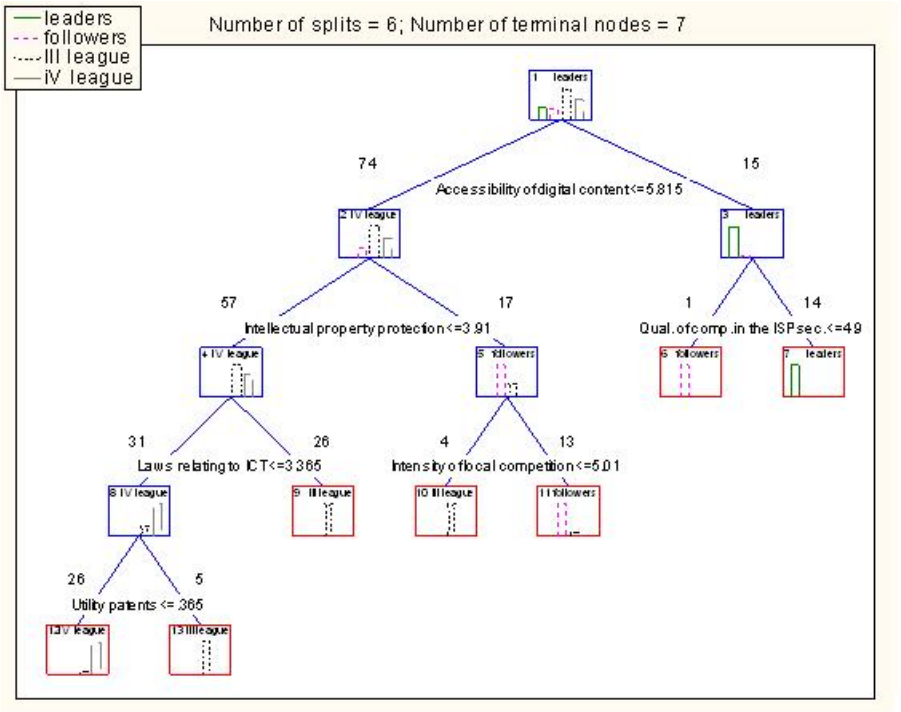

Fig. 2. Decision tree for the Environment factor

According to the criterion "Internet access in schools", the decision tree is divided into two subtrees. The leaders and followers classes are assigned to the terminal nodes in the right subtree, while classes of leagues III and IV are assigned to the terminal nodes in the left subtree. For example, for the leaders'class there are two terminal nodes and the rule for classifying a country into that class is

If (("Internet access in schools" > 4.155) and ("Local availability of specialized research and training services" >4.935) and ("Importance of ICT to government vision of the future" >4.61))

or (("Internet access in schools" >4.155) and ("Local availability of specialized research and training services" >4.935) and ("Importance of

ICT to government vision of the future" ≤ 4.61) and("High-speed monthly broadband subscription" ≤ 1.355)) then class=leader.

Table 2. Importance of different indicators for Readiness factor

Importance of indicator	Indicator
60.00	Quality of math and science education
69.00	Quality of the educational system
99.67	**Internet access in schools**
88.00	**Buyer sophistication**
62.33	Residential telephone connection charge
58.67	Residential monthly telephone subscription
82.00	High-speed monthly broadband subscription
79.00	Lowest cost of broadband
69.33	Cost of mobile telephone call
78.00	Extent of staff training
85.33	**Local availability of specialized research and training services**
63.33	Quality of management schools
75.00	Company spending on R&D
82.67	**University-industry research collaboration**
58.33	Business telephone connection charge
63.67	Business monthly telephone subscription
86.67	**Local supplier quality**
60.67	Local supplier quantity
26.33	Computer, communications, and other services imports
55.00	Government prioritization of ICT
57.67	Government procurement of advanced technology products
68.33	Importance of ICT to government vision of the future
86.00	**E-Government Readiness Index**

Fig. 4 shows the decision tree for the third factor – Usage (for one of the folds, with the stopping criterion FACT=0.15). It also consists of six splits and seven terminal nodes. The "Extent of business Internet use" indicator is selected by splitting the rule to divide the tree at the top into two subtrees. Usage is a unique criterion which differentiates the class of the leading countries from countries in other classes. So the classification rule for the leaders' class is simple:

If ("Extent of business Internet use" >5.32) then class=leader.

Table 3 shows all indicators important for defining the Usage factor and the six most important ones (corresponding to the number of splits in the decision tree) are marked in bold. Five of these six indicators are also selected by splitting the rule in the decision tree.

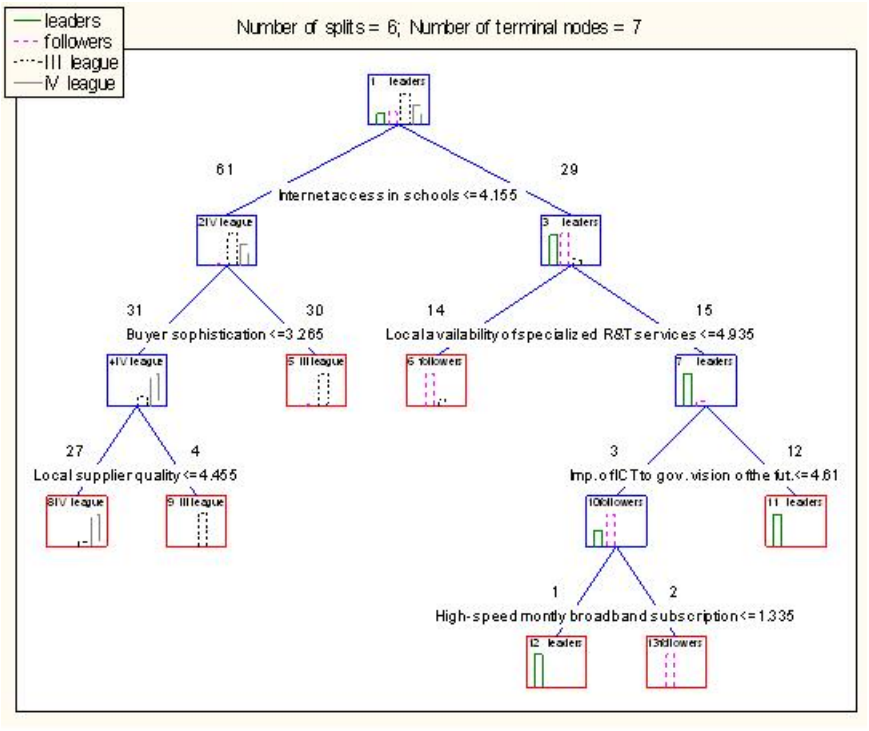

Fig. 3. Decision tree for Readiness factor

Table 3. Importance of different indicators for Usage factor

Importance of indicators	Indicator
86.00	Mobile telephone subscribers
87.33	Personal computers
93.67	**Broadband Internet subscribers**
92.00	**Internet users**
67.00	Internet bandwidth
95.33	**Prevalence of foreign technology licensing**
95.00	**Firm-level technology absorption**
69.33	Capacity for innovation
83.00	Availability of new telephone lines
93.67	**Extent of business Internet use**
52.67	Government success in ICT promotion
82.00	Availability of government online services
79.67	ICT use and government efficiency
91.67	**Presence of ICT in government offices**
49.67	E-Participation Index

Overall classification results are presented in Table 4. The performance of the classification is measured by F_1 measure, which is a combined measure computed as $F_1 = 2pr/(p+r)$, where p is precision of classification and r is recall of classification. Precision is the proportion of countries correctly classified into a specific class from all the countries for which this class is predicted, while recall is

the proportion of countries predicted to be in a specific class from all the countries which are observed to be in this class.

The classification is obtained by 3-fold cross validation and for two levels of stopping criteria for each class: FACT05 (meaning that the proportion of elements in classes different from the predicted is less than 5%) and FACT15 (meaning that the proportion of elements in classes different from the predicted is less than 15%). The decision trees obtained by FACT05 stopping criterion are significantly more complex than those obtained by FACT15 criterion, which are more interpretative due to their simplicity. As can be seen from Table 4, the values of F_1 measure are higher for FACT05 criterion than for the FACT15 criterion only for the Environment factor (the better result for every class and the factor is marked in bold). For the other two factors the results of F_1 measure are similar for the two examined stopping criteria. Therefore, the classification by more complex decision trees with purer terminal nodes did not result in better classification performance, which indicates the presence of overfitting in the case of classification using stricter stopping criteria.

Generally speaking, the classification performance is best for the leaders'class, which means that the leading countries are easily recognized according to the used indicators. The classification performance is not so good for the followers' class, because some of them are inclined to the adjacent classes, and in some cases it is difficult to split them correctly.

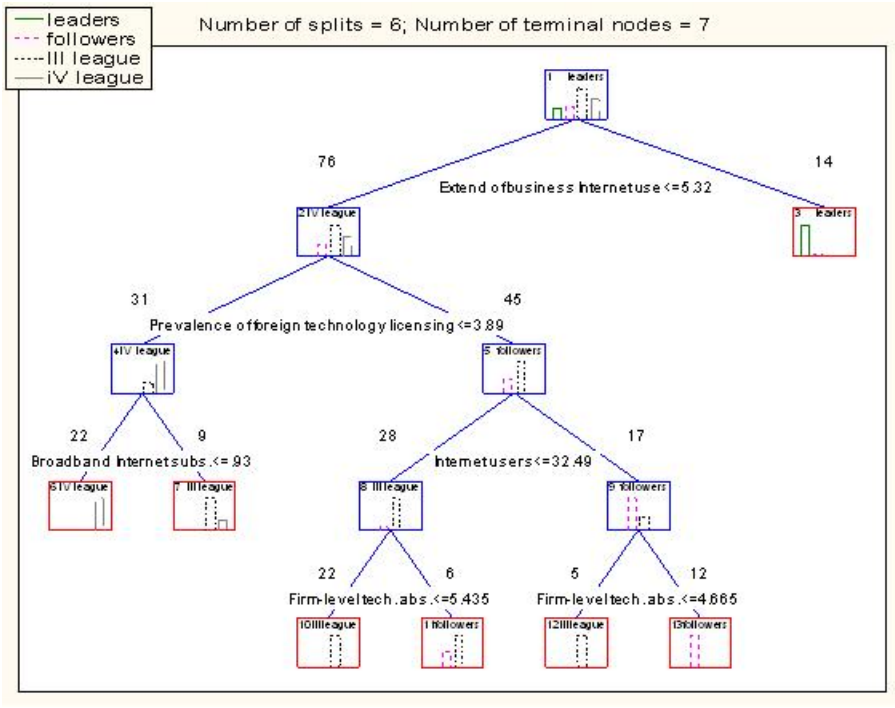

Fig. 4. Decision tree for Usage factor

Table 4. Classification results

Factor	Stopping criteria	F_1 measure			
		leaders	followers	league III	league IV
Environment	FACT05	**0.88 ± 0.06**	**0.61 ± 0.15**	**0.87 ± 0.04**	**0.91 ± 0.04**
	FACT15	0.84 ± 0.05	0.50 ± 0.16	0.74 ± 0.03	0.89 ± 0.03
Readiness	FACT05	**0.83 ± 0.04**	0.49 ± 0.19	**0.72 ± 0.15**	**0.80 ± 0.05**
	FACT15	0.81 ± 0.07	**0.56 ± 0.09**	**0.72 ± 0.03**	0.78 ± 0.03
Usage	FACT05	**0.88 ± 0.04**	**0.57 ± 0.08**	**0.73 ± 0.06**	0.76 ± 0.05
	FACT15	**0.88 ± 0.04**	**0.57 ± 0.08**	0.71 ± 0.07	**0.78 ± 0.06**

4. Conclusion and discussion

The main objective of this paper was to identify the relevant indicators for efficient utilization of ICT. For this purpose we classified countries represented by 68 indicators by means of a decision tree. The most relevant indicators are those which appear as splitting criteria in the decision tree, especially those which are near the top of the tree. The list of indicators according to their measure of relevance is also provided.

For the Environment, the most relevant indicators are related to the availability of venture capital, latest technologies and utility patents, accessibility of digital content, legal infrastructure related to ICT and intellectual property. For the Readiness factor, the most relevant indicators are to a great extent related to education and collaboration between industry and university. Other relevant indicators are buyers' sophistication, local supplier quality and government readiness measured by the E-Government Readiness Index. For the Usage factor the most relevant indicators are related to the extent of Internet use (private and business), firm-level technology absorption and the presence of ICT in government offices.

This approach provides a mechanism for tracking the main indicators for differentiation between countries. Therefore the obtained results could be used in developing an improvement plan aimed at enhancing the utilization of ICT on the country level. Every government should ensure the necessary improvements in this area reflected in the key indicators for measuring the effectiveness of the undertaken action.

Acknowledgement: This paper was supported by the Development of metric for ICT management (016-0161199-1718) Project, funded by the Croatian Ministry of Science, Education and Sport.

References

1. Brieman, L., J. H. Friedman, R. A. Olshen, C. J. Stone. Classification and Regression Trees.Wedsworth, Belmont, 1984.
2. Cette, G., J. Mairesse, Y. Kocoglu. The Contribution of Information and Communication Technology to French Economy Growth. – Banquet de France Bulletin Digest, June 2001, No 90, 23-37.

3. Dedrick, J., V. Gurbaxani, K. L. Kraemer. Information Technology and Economic Performance: A Critical Review of the Empirical Evidence. – ACM Computing Surveys, Vol. **35**, March 2003, No 1, 1-28.
4. Dutta, S., I. Mia. The Global Information Technology Report 2008-2009. World Economic Forum, Accessed 15 May, 2009.
5. Emrouznejad, A., E. Cabanda, R. Gholami. An Alternative Measure of the ICT-Opportunity Index. – Information & Management, Vol. **47**, 2010, 246-254.
6. Hanafizadeh, M. R., A. Saghaei, P. Hanafizadeh. An Index for Cross-Country Analysis of ICT Infrastructure and Access. – Telecommunications Policy, Vol. **33**, 2009, 385-405.
7. Heshmati, A., W. Yang. Contribution of ICT to the Chinese Economic Growth. The RATIO Institute and Techno-Economics and Policy Program College of Engineering, Seoul National University, 2006.
8. Huveneers, C**.** ICT Diffusion and Firm-Level Performance**.** Case Studies for Belgium, Planning & Working Papers, 2003.
9. Krakar, Z., S. Tomić Rotim. Assessment of Croatia's Readiness for Using the ICT Potentials. – In: Proceedings of 20th Central European Conference on Information and Intelligent Systems, Varaždin, 2009.
10. Loh, W.-Y., N. Vanichsetakul. Tree-Structured Classification via Generalized Discriminant Analysis (With Discussion). – Journal of the American Statistical Association, Vol. **83**, 1988, 715-728.
11. Mitchell, T. M. Machine Learning. McGraw-Hill, 1997.
12. Piatkowski, M. The Contribution of ICT Investment to Economic Growth and Labor Productivity in Poland. – Tiger Working Paper Series, July 2003, No 43, Warsaw, 1-23.
13. Tomić Rotim, S., J. Dobša. Clustering of European Countries by Their Readiness for Effective Utilization of ICT. – In: Proceedings of 33rd International Conference on Information Technology Interfaces ITI'2011, Cavtat, 2011.