Omega 41 (2013) 131–142

Contents lists available at SciVerse ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Using Data Envelopment Analysis (DEA) for monitoring efficiency-based performance of productivity-driven organizations: Design and implementation of a decision support system

Sergey Samoilenko [a,*], Kweku-Muata Osei-Bryson [b]

[a] Department of CIS/CS, Virginia Union University, Richmond, VA23220, USA
[b] Department of Information Systems, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Available online 6 February 2012

Keywords:
Productivity
Dynamic environments
Efficiency
DEA
DSS
Data mining
Clustering
Decision Tree

ABSTRACT

The competitive nature of the business environment requires the productivity-driven organization to be aware of its relative level of effectiveness and efficiency vis-à-vis its competitors. This suggests the need, first, for an effective mechanism that allows for discovering appropriate productivity models for improving overall organizational performance, and, second for a feedback-type mechanism that allows for evaluating multiple productivity models in order to select the most suitable one. In this paper our focus is on organizations that consider the states of their internal (e.g., possibly exemplified by *resource-based view*) and external (e.g., possibly exemplified by *positioning*) organizational environment in the formulation of their strategies. We propose and test a DEA-centric Decision Support System (DSS) that aims to assess and manage the relative performance of such organizations.

1. Introduction

Modern organizational entities typically operate in dynamic, competitive environments. Within this context, the critical issues of organizational survival and advancement often lead to calls for improvements in the levels of effectiveness and efficiency [64]. However, due to the relativity of the concepts of efficiency and effectiveness, productivity-driven organizations must take into consideration the performance of their competitors. For the dynamic nature of the business environment will cause the levels of performance of competing organizations to change over time, and if the efficiency of the competitors has improved, then a productivity-driven organization must respond with its own improvements in efficiency.

Although some improvements in productivity do not require any drastic structural transformations but simply call for a gradual type of improvements in the level of performance (e.g., TQM, BPI, etc.), significant changes in the levels of effectiveness and efficiency often require structural reorganizations (e.g., ERP, BPR, etc.) that could result in periods of unstable behavior, which, if not managed, could escalate and become chaotic [52]. Resultantly, in a dynamic business environment any static model that is used to describe the relationship between inputs and outputs will have limited usefulness and feasibility in periods of instability. This suggest the need, first, for an effective mechanism that allows for discovering appropriate productivity models for improving overall organizational performance [24] and, second for a feedback-type mechanism that allows for evaluating multiple productivity models in order to select the most suitable one.

The overall goal of this investigation is to propose and test a Decision Support System (DSS) that aims to assess and manage the relative performance of organizations. We focus on organizations that consider the states of their internal (e.g., possibly exemplified by *resource-based view*) and external (e.g., possibly exemplified by *positioning*) organizational environment in the formulation of their strategies, such that the achievement of an organizational goal is dependent on the level of performance that is commonly measured in terms of the levels of the efficiency of utilization of inputs, effectiveness of the production of outputs, and efficiency of conversion of inputs into outputs. This suggests that an important component technique of our DSS is Data Envelopment Analysis (DEA), which is widely used by researchers and practitioners for the purposes of measuring productivity and relative performance [74, 7, 17, 15, 73, 26, 63]. However, other techniques are also required for providing answers to several questions that are relevant to the organization's search for the productivity model that is most suitable with respect to survival and advancement. In this investigation we focus on the following questions related to system requirements:

We present our investigation as follows. Part One outlines the functionality and composition of the proposed system. Part Two

* Corresponding author. Tel.: +1 804 342 7381.
E-mail addresses: svsamoilenko@vuu.edu (S. Samoilenko), KMOsei@vcu.edu (K.-M. Osei-Bryson).

0305-0483/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.omega.2011.02.010

offers an overview of the structural elements of the proposed DSS. Part Three outlines the design of DSS. Part Four offers an illustrative example of the DSS in action. A brief conclusion follows.

2. The functionality and composition of the DSS

The dynamic nature of the business environment suggests the presence of a concept that is central to a productivity-driven organization, namely, that of the *superior stable configuration*. Given the goal of achieving a high level of efficiency of conversion of inputs into outputs, a superior stable configuration in the context of a productivity-driven organization may imply *a model of conversion of inputs into output (input–output model) characterized by a high level of efficiency*. Consequently, we put forward the following propositions:

Proposition 1. *Stability of the performance of a productivity-driven organization is dependent on the presence of the stable input–output model.*

Proposition 2. *Accomplishment of the organizational goal of a productivity-driven organization is dependent on the creation and implementation of a stable input–output model characterized by the high level of efficiency.*

Proposition 3. *In order to monitor performance of a productivity-driven organization, DSS must be able to create and identify superior stable configurations, represented by the input–output models characterized by the high level of efficiency.*

We suggest that the design of the proposed DSS must include two sets of functionalities: *externally-oriented*, and *internally-oriented*. The *externally-oriented* functionality of this DSS is directed towards evaluating the external competitive environment of a productivity-driven organization, as well as identifying the differences between the current state of the organization and the states of its competitors. The *internally-oriented* functionality, on the other hand, is directed towards the optimization of the level of productivity of the organization, as well as towards an identification of the factors impacting the efficiency of the input–output process. We suggest that such a DSS could be implemented using a combination of parametric and non-parametric data analytic and data mining techniques including Data Envelopment Analysis (DEA), Cluster Analysis (CA), Decision Tree (DT), Neural Networks (NN), and Multivariate Regression (MR). The suggested functionality of this DSS is presented in Table 2.

While the five data analytic techniques that we use in the design of the proposed DSS have been utilized in IS research in a stand-alone fashion, they are also very frequently used in combination. For example, DEA is widely employed for the purpose of evaluating productivity and performance (e.g., [35, 59, 57, 6, 73, 2, 34, 24, 38]), but it has also been used to complement other data analytic techniques: cluster analysis (e.g., [61, 30, 36, 41]), neural network induction (e.g. [54, 10, 27, 42, 71]), decision tree induction (e.g., [55, 53, 71]), regression analysis (e.g., [19, 6, 47, 56]), and other methods([37, 26, 50]).

3. Overview of the structural components of the DSS

3.1. Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a nonparametric method for measuring the efficiency of a *decision-making unit* (DMU). Any group of entities that receives the same set of the inputs and produces the same set of outputs could be designated as a DMU; it could be a group of people, companies, hospitals, schools, industries, or countries. To determine the relative efficiency of each DMU in the group, DEA collapses inputs and outputs defined by the model into a ratio of a single meta-input and meta-output, and uses methods of linear programming to calculate the efficiency score for each DMU, where obtained score is reflective of the performance [60, 8, 41, 74]. This comparison results in a ranking of the DMUs in terms of their *relative efficiency*, where the highest-ranking DMUs are considered relatively efficient and assigned a perfect score of 1, while the rest of the DMUs in the sample are considered to be relatively inefficient. Resultantly, DEA 'envelops' the data set with the *efficiency frontier* consisting of the relatively efficient DMUs. The two commonly mentioned orientations of DEA models are the *Input-Oriented* and the *Output-Oriented* [12]. An *Input-Oriented* model is concerned with the minimization of the use of the inputs for achieving a given level of the output [14]. A relatively efficient DMU under *input-orientation* cannot reduce its levels of inputs any further to achieve a given level of output, while the relatively inefficient DMUs (with the scores of greater than "0" but less than "1") could. An *Output-Oriented* DEA model, conversely, is concerned with the maximization of the level of the outputs per given level of inputs. A relatively efficient DMU under *output-orientation* cannot increase its levels of outputs any further while relying on a given level of inputs, while the relatively inefficient DMUs (with the scores of greater than "1") could. Thus, while in both cases a relatively efficient DMU is assigned a score of "1", a relatively inefficient DMU will receive a score of greater than "1" under *output-orientation*, and a score in the [0, 1) interval under *input-orientation*.

DEA is a flexible method [16, 65, 1, 37] that can be applied under different underlying economic assumptions about the returns to scale [58] yield different DEA models [25]. An assumption of the *constant return-to-scale* (CRS) model reflects the situation where the changes in output are in the same proportion as the changes in inputs (e.g., changes of 50% in inputs correspond to the changes of 50% in outputs), while assumptions of the *variable returns-to-scale* (VRS) model reflects increasing (e.g., changes of 25% in inputs correspond to the changes of 50% in outputs), and *non-increasing returns-to-scale* (NIRS) model reflects decreasing (e.g., changes of 50% in inputs correspond to the changes of 25% in outputs) returns to scale. We direct the interested reader to the comprehensive presentations of the theoretical underpinnings of the DEA by Cook and Zhu [18] and Cooper et al. [20].

3.2. Cluster analysis (CA)

Clustering is a popular non-directed learning data mining technique for partitioning a dataset into a useful set of mutually exclusive clusters such that the similarity between the observations within each cluster (i.e., subset) is high, while the similarity between the observations from the different clusters is low (e.g., [45, 49, 44, 67, 22]). There are different reasons for doing clustering, and one of them is to find a set of natural groups (i.e., segmentation), and the corresponding description of each group. This is relevant if there is the belief that there are natural groupings in the data. Jain et al. [32] noted that there are three approaches for assessing cluster validity: (1) external assessment which involves comparing the generated segmentation (i.e., set of clusters) with an a priori structure, typically provided by some domain experts; (2) internal assessment which attempts to determine if the generated set of clusters is "intrinsically appropriate" for the data; and (3) relative assessment which involves comparing two segmentations (i.e., two sets of clusters) based on some performance measures and measure their relative performance. Our use of cluster analysis is based on the assumption that there are natural groupings in the data, and will involve the use of external assessment to assess cluster validity. There are numerous algorithms available for doing clustering. They may be categorized in various ways such as: hierarchical (e.g., [43, 68]) or partitional (e.g., [40]), deterministic or probabilistic (e.g., [5]), hard or fuzzy (e.g., [3, 23]). A hybrid partitional/hierarchical approach provided by SAS *Enterprise Miner* is used to generate the clusters.

3.3. Decision tree induction (DT)

A decision tree (DT) is a tree-structure representation of the given decision problem (e.g., [51, 62, 72]). Construction of a DT involves a recursive partitioning of the training data resulting in a DT such that each *non-leaf node* of the tree is associated with one of the input variables, each branch from a non-leaf node is associated with a subset of the values of the corresponding input variable, and either each leaf node is associated with a single class or further partitioning of the given leaf would result in at least its child nodes being below some specified threshold. Associated with each leaf node of the DT is an IF–THEN rule that associates values of input variables with a class assignment.

The splitting method is the component of the DT induction algorithm that determines both the variable that is selected for a given node of the DT and also the partitioning of the values of the selected variable into mutually exclusive subsets such that each subset uniquely applies to one of the branches that emanate from the given node. Typically the variable that is selected as the root node of the DT is the most important predictor variable of that DT.

3.4. Neural networks (NN)

An artificial neural network (NN) is a black box model that can be used in directed learning to represent unknown complex relationships in the data. Neural network induction is particularly useful for prediction and classification problems where there is no known mathematical formula that relates inputs to outputs, and prediction is more important than explanation. This data mining method has been used extensively in the field of IS research [70, 46, 31, 69, 66, 39, 21, 28, 13, 9, 29]. For directed learning problems the NN consists of an *input layer* (for the input variables) *hidden layer*, and *output layer* (for the target variable); each layer containing one or more *nodes*. This brief overview of such complex subject as NN cannot do justice to the topic. Thus, we direct the interested reader to Bishop [4] for a comprehensive treatment of the subject. It is interesting to note that both NN induction and DEA provide black box models that relate inputs and outputs Fig. 1.

4. Structure of the DSS

The overview of the purpose of each technique used in our DSS is provided in Table 3, and depicted graphically in Fig. 2. Table 4 describes the questions that this DSS is designed to answer.

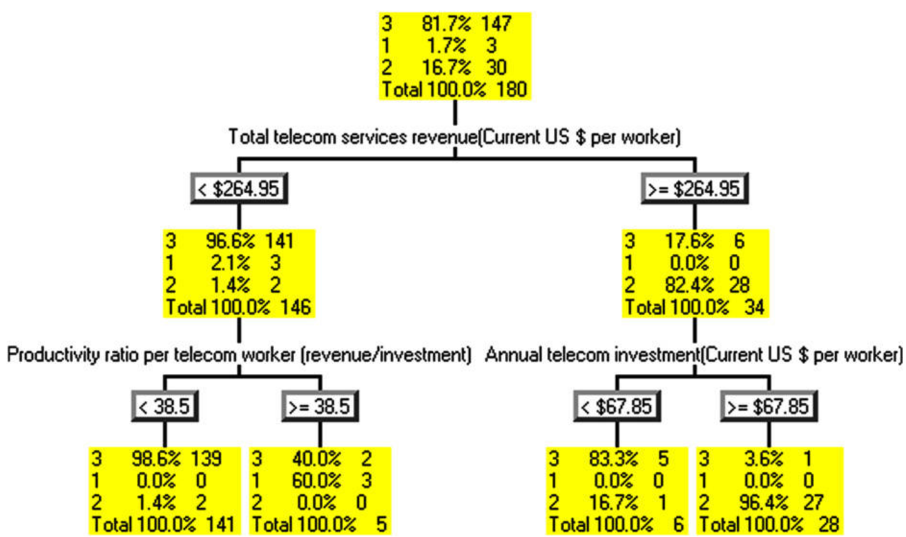

Fig. 1. Example of a DT.

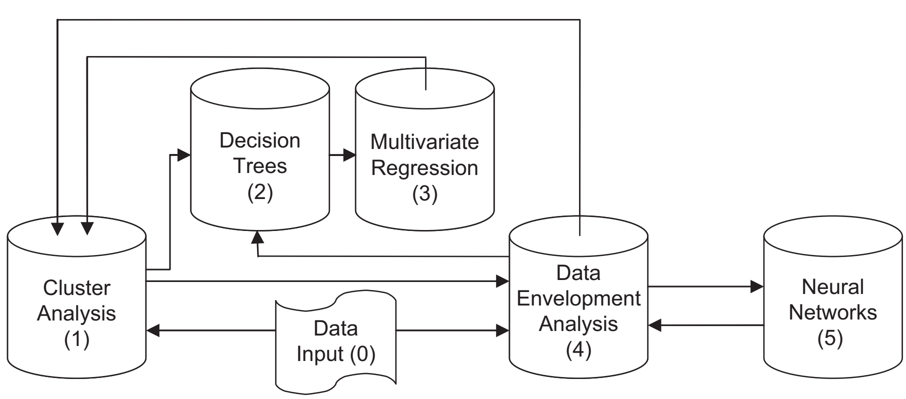

Fig. 2. Design of the Proposed DEA-centric DSS.

5. Illustrative example

Our DSS can be applied at different organizational levels, including the country level and the firm level. This illustrative example involves the country level and deals with the efficiency and effectiveness of the impact of investments in Telecoms, a type of investments that is common to almost all of the economies in the world. The context is represented by the following 18 transition economies: Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Czech Republic, Estonia, Hungary, Kazakhstan, Kyrgyz Republic, Latvia, Lithuania, Moldova, Poland, Romania, Slovakia, Slovenia, and Ukraine. The time-series data covering the period from 1993 to 2002 were obtained from the *World Development Indicators* database (web.worldbank.org) and the International Telecommunication Union' *Yearbook of Statistics* (www.itu.int) Figs. (3)–(8).

Within the context of the sample, we view 18 TEs as business entities with the same general type of "investments in Telecoms-to-revenues in Telecoms" business process. TEs compete for the limited pool of investments funds, in the form of Foreign Direct Investments (FDI) and private investments, under a condition that for a given economy the level of incoming investments is dependent on the level of productivity in regard to "investments-to-revenues" process. We pose the following general research question consistent with the topic of this paper, namely:

How could a given TE improve its level of productivity with regards to its investments in Telecoms?

This question, in turn, is expanded into three efficiency-based sub-questions consistent with the design of DEA-centric DSS, namely:

1. *How could a given TE improve its level of efficiency of utilization of investments in Telecoms?*
2. *How could a given TE improve its level of efficiency of production of revenues from Telecoms?*
3. *How could a given TE improve its level of efficiency of the process of conversion of investments into revenues from Telecoms?*

In this illustrative example we will demonstrate how the proposed design of our DEA-centric DSS, geared toward answering eleven questions listed in Section 1, can contribute to answering these three efficiency-based questions. We present the illustrative example in step-by-step fashion following the sequential order of system requirements outlined in Tables 1 and 3. We use SAS' *Enterprise Miner* data mining software to conduct CA, DT, NN, and MR, and *OnFront* to conduct DEA. Because the design of the proposed DSS systems is DEA-centric, one of the prerequisites for using it is associated with identifying a DEA model that is to be used in evaluating productivity of the organizational entities in the sample. Table 5 lists the set of variables used in the illustrative example.

5.1. Step 1: System Requirement 1.1

The purpose of the first system requirement is to offer the decision maker a capability to inquire into the nature of the competitive business environment in regard to the presence of the multiple heterogeneous groups of business entities. In order to implement this functionality we incorporated CA into the design of our DSS.

The variables used in CA are listed in Table A1 of Appendix; the variables were standardized prior to CA. We began the cluster analysis by using "Automatic" setting, which did not require a specification of the exact number of clusters by the analyst. This setting produced a five-cluster solution that was considered to be the starting point in the analysis. By sequentially reducing the number of clusters we derived a two-cluster solution which was considered to be final. The membership of the two clusters is provided in Table 6. What is the basis for accepting this two clusters segmentation? The reader may recall that earlier we stated that we would use an external assessment approach to assess cluster validity. Further domain expert opinion can be considered to provide external confirmation of the validity of this segmentation. Such domain expert support is provided in the research of Piatkowski [48], who concluded that in the period "*between 1995 and 2000 ICT capital has most potently contributed to output growth in the Czech Republic, Hungary, Poland, and Slovenia*." Thus, it could be suggested that we were able to separate 18 transition economies (TEs) into the two groups: the *Leaders* group that consists of TEs which benefited the most from the investments in telecom, and the *Followers* group that consists of TEs where the benefits are less pronounced.

The results of the CA provide evidence that Cluster1 is different from Cluster2 in terms of the two dimensions: *Investments* and *Revenues from Telecoms*. Consequently, with regard to these dimensions, for a given TEs its own cluster will represent a peer context, while members of the other cluster will comprise a

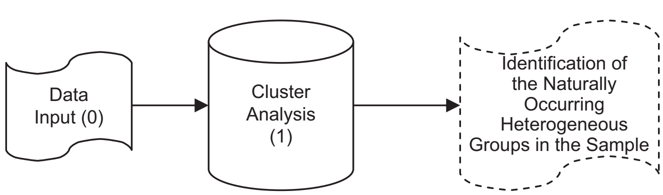

Fig. 3. *SR1.1* Detection of changes in the external competitive environment – CA.

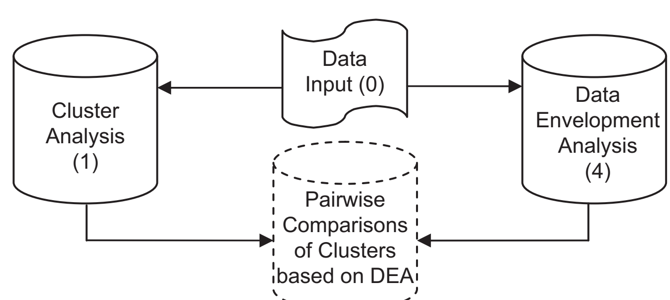

Fig. 5. *SR1.3* Identification of the relative efficiency of the business entity relative to its competitors – CA and DEA.

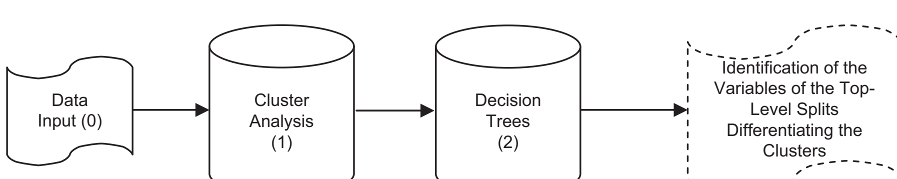

Fig. 4. *SR1.2* Identification of the possible factors that resulted in changes – CA and DT.

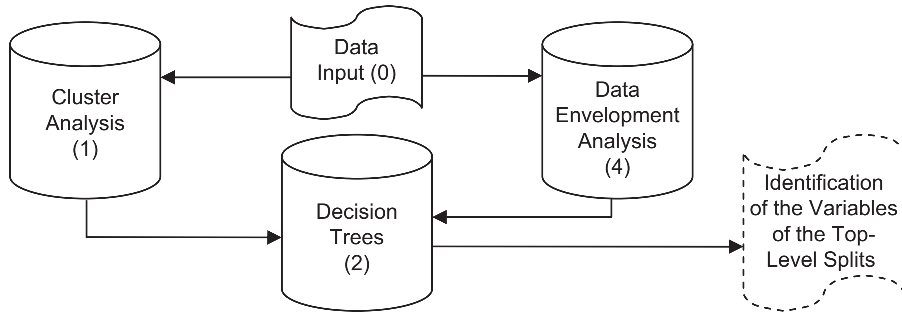

Fig. 6. *SR1.4* Identification of the factors associated with the differences in the relative efficiencies of the competitors – CA, DEA & DT.

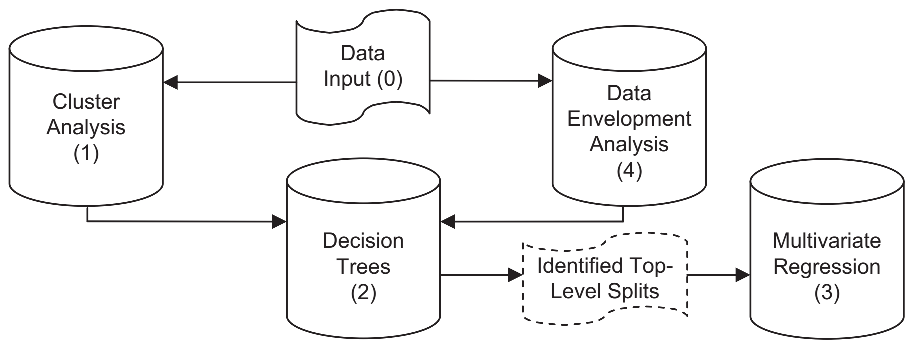

Fig. 7. *SR2.1* Identification of the factors impacting the current level of the relative efficiency of the input–output process.

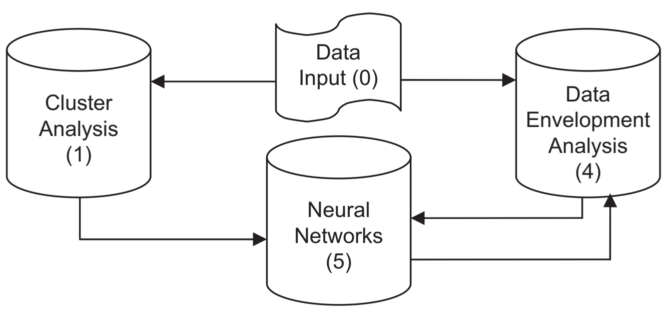

Fig. 8. ***SR2.2*** Identification of the most effective ways of increasing the level of efficiency of the input–output process.

non-peer context. Step 1 allows the decision maker to answer Question 1, namely

Q1: *What are the naturally occurring groups of business entities that exist in the current competitive environment, and what is the membership of the peer group of the given entity?*

as follows:

A1: *The context of 18 TEs is comprised of two groups, membership of each group is provided in* Table 6.

5.2. Step 2: System Requirement 1.2

However, even if the decision maker identified the presence of multiple context within a given business environment, it is not clear what differentiates the peer context from the non-peer context; this question will be answered by the functionality of our DSS that addresses System Requirement 2.

The results of the CA allow us to introduce a target variable "ClusterNumber" to serve as an identifier of a given group in the sample. By using this variable in DT analysis we can determine, based on the top-level split, the dimension that differentiates two groups the most. Clearly, when conducting DT analysis we do not have to be limited to the set of variables that was used for CA. The results of DT analysis in the form of the decision rules are presented in Table 7. "*N*" is the number of training observations associated with the given decision rule.

These results allows the decision maker(s) to identify the relevant dimension that most differentiates the peer from the non-peer context for business organizational entity in the sample. For while the two TE clusters differ in terms investments and revenues from Telecoms, but the single most important dimension that differentiates two clusters is associated with the level of investments in Telecoms per Telecom worker. Step 2 allows the decision maker to answer Question 2:

Q2: *What are the differences between the groups of the entities in regard to the values of the variables relevant to business?*

as follows

A2: *The two groups of 18 TEs differ most significantly in terms of the respective levels of investments in Telecoms per Telecom worker.*

5.3. Step 3: System Requirement 1.3

Once the decision maker is able to identify the presence (or absence) of heterogeneous groups of the competitors within the business environment and to pinpoint the most relevant differentiating variable, she/he will benefit greatly from knowing how the groups differ in terms of the efficiency of utilization of investment and production of revenues. This additional information is obtained by means of incorporating DEA in the design of the proposed DSS.

Completing Step 3 involves running DEA and calculating the scores of the relative efficiency for each business entity in the sample. It should be noted that DEA is not applied separately to each of the groups (i.e., cluster) that were generated during the cluster analysis (CA) step. Rather DEA is applied to the entire sample, and so the relative efficiency scores are not determined based on cluster

Table 1
System requirements-related questions.

Q no.	Description
Q1	What are the naturally occurring groups of business entities that exist in the current competitive environment, and what is the membership of the peer group of the given entity?
Q2	What are the differences between the groups of the entities with regard to the values of the variables relevant to business?
Q3	What are the differences between the groups of business entities in terms of the efficiencies of utilization of inputs and production of outputs?
Q4	What is the level of the efficiency-based performance of the peer group of the *organizational* entity relative to other groups?
Q5	What is the level of the efficiency-based performance of the business entity relative to the other *organizational* entities within the same group of peers?
Q6	What are some of the variables that may be responsible for the heterogeneity of the sample with regard to the efficiency-based performance?
Q7	What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the *organizational* entity relative to the entities within the same peer group?
Q8	What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the *organizational* entity relative to the *organizational* entities within other groups?
Q9	What are some of the complementarities that may allow for improving the level of the efficiency-based performance of the *organizational* entity?
Q10	Whether the existing inefficiencies of the business entity are associated with the insufficient levels of inputs or with the inefficient processes of conversion of inputs into outputs?
Q11	Whether the changes in the level of productivity of the *organizational* entity are driven by changes in technology, or changes in efficiency?

Table 2
Possible structural implementation of the functionality of DSS.

Functionality	System requirement	Structural components
1. Externally-oriented	***SR1.1*** Detection of changes in the external competitive environment	Cluster Analysis
	SR1.2 Identification of the possible factors that resulted in changes	Combination of Cluster Analysis and Decision Tree
	SR1.3 Identification of the relative efficiency of the *organizational* entity relative to its competitors	Data Envelopment Analysis
	SR1.4 Identification of the factors associated with the differences in the relative efficiencies of the competitors	Combination of Data Envelopment Analysis, Cluster Analysis, and Decision Tree
2. Internally-Oriented	***SR2.1*** Identification of the factors impacting the current level of the relative efficiency of the input–output process	Multivariate Regression analysis
	SR2.2 Identification of the most effective ways of increasing the level of efficiency of the input–output process	Combination of Data Envelopment Analysis and Neural Networks

Table 3
Design of a DSS – steps & methods used for generating actionable information.

Method	General purpose	Outcome	Specific purpose in the DSS
Data Envelopment Analysis (DEA)	Allows for evaluating the relative efficiency of the conversion of inputs into outputs by each *organizational* entity in the sample	Scores for the relative efficiency for each *organizational* entity in the sample	To determine scores for the relative efficiency of the process of converting inputs into outputs by each *organizational* entity, as well as for each group of the entities in the sample.
Cluster Analysis (CA)	Allows for identifying of naturally occurring groups within a data sample	Groups of the sample that differ with respect to the variables representing the entities in the sample	To identify the presence of groups of *organizational* entities which differ in terms of the levels of the relevant to the *organizational* domains values
Decision Tree (DT)	Allows for obtaining a set of decision rules for separating multiple groups within the sample.	A set of variables (based on the top-level splits) that allow for differentiating multiple groups of the sample	To identify a set of variables that differentiate the *organizational* entities within the sample
Multivariate Regression (MR)	Allows for identifying presence of the interaction effect between the independent variables in the model.	A set of complementarities that impact the dependent variable	To identify a set of input variables that have a complementary positive (or negative) impact on the outputs of *organizational* entities
Neural Networks (NN)	Allows for modeling the relationships between inputs and outputs	A “black-box” model of the process by which inputs are transformed into outputs	To construct a model for the transformation of inputs into outputs for each of the groups (as identified by CA) with different levels of relative efficiency (as calculated by DEA)

membership. This approach allows us to legitimately compare the average relative efficiency score for any pair of clusters. In the case of our illustrative example, CA resulted in two clusters. Our application of DEA resulted in *relative efficiency* and *Malmquist Index* scores for each DMU in the entire set. The application of both techniques allows for the legitimate computation of average *relative efficiency* and *Malmquist Index* scores for each cluster. As a result, the decision maker has information regarding the averaged relative efficiency for the peer- and non-peer groups. For the purposes of our illustrative example we conducted DEA under assumptions of

Table 4
DSS requirements and corresponding sequences of techniques.

Requirement	Question addressing system requirement	Sequence
SR1.1	Q1: What are the naturally occurring groups of organizational entities that exist in the current competitive environment, and what is the membership of the peer group of the given entity?	0–1
SR1.2	Q2: What are the differences between the groups of the entities in regard to the values of the variables relevant to organizational entity?	0–1–2
SR1.3	Q3: What are the differences between the groups of organizational entities in terms of the efficiencies of utilization of inputs and production of outputs?	0–1–4
SR1.3	Q4: What is the level of the efficiency-based performance of the peer group of the organizational entity relative to other groups?	0–1–4
SR1.3	Q5: What is the level of the efficiency-based performance of the organizational entity relative to the other business entities within the same group of peers?	0–1–4
SR1.4	Q6: What are some of the variables that may be responsible for the heterogeneity of the sample in regard to the efficiency-based performance?	0–1–4–2
SR1.4	Q7: What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the organizational entity relative to the entities within the same peer group?	0–1–4–2
SR1.4	Q8: What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the organizational entity relative to the entities within other groups?	0–1–4–2
SR2.1	Q9: What are some of the complementarities that may allow for improving the level of the efficiency-based performance of the organizational entity?	0–1–4–2–3
SR2.2	Q10: Whether the existing inefficiencies of the organizational entity are associated with the insufficient levels of inputs or with the inefficient processes of conversion of inputs into outputs?	0–1–4–5–4
SR2.2	Q11: Whether the changes in the level of productivity of the business entity are driven by changes in technology, or changes in efficiency?	0–1–4

Table 5
List of variables used DEA models.

Role	Variables
Input	GDP per capita (in current US $), Full-time telecommunication staff(% of total labor force), Annual telecom investment per telecom worker, Annual telecom investment(% of GDP in current US $), Annual telecom investment per capita, Annual telecom investment per worker
Output	Total telecom services revenue per telecom worker, Total telecom services revenue(% of GDP in current US $), Total telecom services revenue per worker, Total telecom services revenue per capita

Table 6
Results of CA.

Cluster		Members
ID	Description	
1	The *Leaders*	Czech Rep, Estonia, Hungary, Latvia, Lithuania, Poland, Slovenia, Slovakia
2	The *Followers*	Albania, Armenia, Azerbaijan, Belarus, Bulgaria, Kazakhstan, Kyrgyzstan, Moldova, Romania, Ukraine

constant (CRS), *variable* (VRS) and *non-increasing* (NIRS) *return-to-scale* and averaged the scores for Cluster1 and Cluster2. The results are presented in Table 8.

During Step 3 we also conduct DEA to calculate the Malmquist index (MI) of productivity growth for both clusters in order to measure changes in the productivity and efficiency [11, 33] and to evaluate the relative magnitude of the components of MI, namely, change in efficiency (EC) and change in technology (TC). This allows the decision maker to identify whether the growth in productivity was primarily efficiency, or technology-driven. The results are presented in Table 9.

Results of this step provide the decision maker with important information regarding the relative efficiency of the peer vs. the non-peer group within the competitive business environment. In

Table 7
Results of DT analysis.

Set of decision rules generated by decision tree analysis
IF Annual telecom investment(Current US $ per telecom worker) < $9610 **THEN** Cluster = { **2: 96.4%;** 1: 3.6% } where $N=110$. **IF** Annual telecom investment (Current US $ per telecom worker) ≥ $9610 **THEN** Cluster = { 2: 2.9%; **1: 97.1%** } where $N=70$.

the case of our illustrative example an investigator can easily determine that under any assumption regarding *return-to-scale*, Cluster1 is relatively more efficient than Cluster2 not only in terms of utilization of investments (*input-orientation*), but also in terms of the production of revenues (*output-orientation*). However, we can expect that each cluster will contain relatively efficient business units and relatively inefficient ones. The purpose of System Requirement 4 is to provide the decision maker with the functionality allowing inquiring into the differences between relatively inefficient and relatively efficient peer- and non-peer entities. It should be noted that this step allows the decision maker to answer Questions 3, 4, 5 and 11:

Q3: *What are the differences between the groups of business entities in terms of the efficiencies of utilization of inputs and production of outputs?*

A3: *Cluster1 group of 18 TEs is relatively more efficient that Cluster2 group in terms of the utilization of investments and production of revenues from Telecoms.*

Q4: *What is the level of the efficiency-based performance of the peer group of the business entity relative to other groups?*

A4: *Those TEs that are members of Cluster1 of 18 TEs are, on average, relatively more efficient than the members of Cluster2.*

Q5: *What is the level of the efficiency-based performance of the business entity relative to the other business entities within the same group of peers?*

Table 8
Results of DEA, averaged scores of relative efficiency.

Orientation	Return to scale	Cluster1	Cluster2	Conclusion
Input-oriented	CRS	0.89	0.79	*Cluster1* is relatively more efficient than *Cluster2*
	VRS	0.95	0.88	*Cluster1* is relatively more efficient than *Cluster2*
	NIRS	0.89	0.80	*Cluster1* is relatively more efficient than *Cluster2*
Output-oriented	CRS	1.21	1.44	*Cluster1* is relatively more efficient than *Cluster2*
	VRS	1.18	1.30	*Cluster1* is relatively more efficient than *Cluster2*
	NIRS	1.21	1.38	*Cluster1* is relatively more efficient than *Cluster2*

Table 9
Results of DEA, Malmquist Index and EC and TC components.

Productivity growth	Cluster1	Cluster2	Conclusion
MI	1.23	1.20	*Cluster1* has a relatively higher growth in productivity than *Cluster2*
EC component of MI	1.13	1.09	Growth in productivity of *Cluster1* has a relatively greater contribution from changes in efficiency than *Cluster2*
TC component of MI	1.11	1.13	Growth in productivity of *Cluster2* has a relatively greater contribution from changes in technology than *Cluster1*
Dominant component	EC > TC	TC > EC	Growth in productivity of Cluster1 is efficiency-drivenGrowth in productivity of Cluster2 is technology-driven

A5: *Both groups of 18 TEs, namely, Cluster1 and Cluster2, contain relatively efficient and relatively inefficient entities; scores of the relative efficiency provided by DEA allow for evaluating the level of efficiency-based performance of each TE relative to its peers.*

Q11: *Are the changes in the level of productivity of the organizational entity are driven by changes in technology, or changes in efficiency?*

A11: *The changes in the level of productivity of Cluster1 are driven by changes in efficiency, while the changes in the level of productivity of Cluster2 are driven by changes in technology.*

5.4. Step 4: System Requirement 1.4

In the case of our illustrative example we ended up with two clusters; thus, we can hypothesize the presence of four groups of TEs within our sample: relatively efficient economies of Cluster1, relatively efficient economies of Cluster2, relatively inefficient economies of Cluster1, and relatively inefficient economies of Cluster1. By introducing a target variable *Cluster&Efficiency*, with domain of values [1, 2, 3, 4], we can identify each group of TEs within each cluster and use the target variable in DT analysis to identify the split variables and their values that differentiate the groups. To identify the most meaningful splits the decision maker can opt to display the resulting decision tree in the form of the easy-to-interpret decision rules, and then to concentrate on the rules that have a high probability for the occurrence of the group based on the decision rule. The results of Step 4 are presented in Table 10.

The results provided in Table 10 demonstrate that the functionality provided by System Requirement 4 allows the decision maker to obtain important information regarding some of the factors that differentiate not only efficient and inefficient peers, but also efficient and inefficient non-peers. This information could be useful for the purposes of intra-group benchmarking, as well as for the purpose of formulating strategies for business units that are interested in intergroup transitioning.

Results from this Step allow the decision maker to answer Questions 6, 7 and 8:

Q6: *What are some of the variables that may be responsible for the heterogeneity of the sample in regard to the efficiency-based performance?*

Table 10
Decision rules generated by DT analysis.

Group	Decision rule	Posterior probability
Cluster 1: efficient TEs	*Productivity Ratio per Telecom Worker* ≥ 1.5674014075 & *Annual Telecom Investment* < $836,899,003 & *Full-Time Telecommunication Staff %* ≥ 0.0039016912 & *Annual Telecom Investment per Worker* ≥ $58	0.94
	Productivity Ratio per Telecom Worker ≥ 4.1754445351 & *Annual Telecom Investment per Worker* ≥ $58	1.00
Cluster 2: inefficient TEs	*Annual Telecom Investment* ≥ $836,899,003 & *Full-Time Telecommunication Staff %* ≥ 0.0039016912 & *Productivity Ratio per Telecom Worker* < 4.1754445351 & *Annual Telecom Investment per Worker* ≥ $58	1.00
	Full-Time Telecommunication Staff % < 0.0039016912 & *Productivity Ratio per Telecom Worker* < 4.1754445351 & *Annual Telecom Investment per Worker* ≥ $58	1.00
Cluster 3: efficient TEs	*Full-Time Telecommunication Staff %* < 0.0031414015 & *Productivity Ratio per Telecom Worker* ≥ 3.8043909395 & *GDP per Capita* ≥ $519 & *Annual Telecom Investment per Worker* < $33	1.00
	Total Telecom Services Revenue ≥ 0.0118204323 & *GDP per Capita* < $519 & *Annual Telecom Investment per Worker* < $33	1.00
Cluster 4: inefficient TEs	*Productivity Ratio per Telecom Worker* < 3.8043909395 & *GDP per Capita* ≥ $519 & *Annual Telecom Investment per Worker* < $33	1.00
	Productivity Ratio per Telecom Worker < 2.002357802 & $33 ≤ *Annual Telecom Investment per Worker* < $58	1.00

A6: *Some of the variables responsible for the heterogeneity of the sample in regard to the efficiency-based performance are: Productivity Ratio per Telecom Worker, Annual Telecom Investment, and Full-Time Telecommunication Staff %.*

Q7: *What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the organizational entity relative to the entities within the same peer group?*

A7: *Suggested empirically-justifiable strategies are associated with decreasing the levels of heterogeneity of the variables that are responsible for the heterogeneity of the peer group in regard to the efficiency-based performance such as:*

- *Productivity Ratio per Telecom Worker and Annual Telecom Investment for Cluster1*
- *Full-Time Telecommunication Staff % and Productivity Ratio per Telecom Worker for Cluster2*

Q8: *What are some of the empirically-justifiable strategies that could be employed to improve the level of the efficiency-based performance of the organizational entity relative to the organizational entities within other groups?*

A8: *Suggested empirically-justifiable strategies are associated with decreasing the levels of heterogeneity of the variables that are responsible for the heterogeneity of the sub groups in the sample in regard to the efficiency-based performance such as Productivity Ratio per Telecom Worker and Full-Time Telecommunication Staff %.*

But even if the decision maker could benefit from knowing the variable levels of which could be manipulated in order to obtain improvements in the production of outputs, he/she could benefit even more from identifying complementarities between those variables that produce a synergistic effect on the output. The functionality of the DSS that addresses System Requirement 5 allows the decision maker to identify some of the complementarities that may exist between the relevant to the production process variables.

5.5. Step 5: System Requirement 1.5

In order to demonstrate the functionality of our DSS to identify existing complementarities between the relevant to the production process variables, we present an example with a higher degree of rigor than the average production environment may require. Namely, in the context of our set of TEs we will construct the model of the relationship between the inputs and outputs based on a solid theoretical basis- using a neoclassical framework of growth accounting.

A neoclassical production function relates output and inputs in the following manner: $Y=f\,(A, K, L)$, where Y is an output (most often in the form of GDP), A is the level of technology (TFP), K is a capital stock, and L is a quantity of labor. In the case of our illustrative example, we can relate investments in Telecoms, full-time Telecom employees, and GDP as follows: GDP$=f$ (TFP, investments in Telecoms, full-time Telecom employees). Based on this formulation we can use the following translog formulation of the production function to test for the presence of interaction:

$$\log Y=\beta_0+\beta_1 * \log K+\beta_2 * \log L+\beta_3 * \log K^2+\beta_4 * \log L^2+\beta_5 * \log K * \log L+x,$$

where K is annual investment in Telecoms and L is a quantity of full-time Telecom staff. A test for the presence of the interaction between investments in Telecoms and Telecom staff would involve testing of the following hypothesis:

H0. β_5 is not statistically discernible from 0 at the given level of α.

We present the results of testing of the null hypothesis in Table 11.

Step 5 allows the decision maker to answer Question 9:

Q9: *What are some of the complementarities that may allow for improving the level of the efficiency-based performance of the organizational entity?*

A9: *Annual Telecom Investment and Full-Time Telecommunication Staff % are complementary factors that may allow for improving the level of the efficiency-based performance of the organizational entities in the Leaders group (i.e., Cluster1).*

However, despite obtaining important information regarding the presence of complementarities, the decision maker will still need additional information regarding the best route to improvements in the level of the efficiency of the production process, specifically as it relates to the production of outputs. For example, we determined that the organizational entities of Cluster 1 are, on average, more efficient than the organizational entities of Cluster2; however, we also determined that the levels of investments and revenues of the members of Cluster1 are higher than those of Cluster2. This situation allows for two possible interpretations; first, members of Cluster1 are more efficient than members of Cluster2 because of the superior process of conversion of inputs into outputs, or, the members of Cluster1 are more efficient because they have higher levels of inputs which allows for establishing and maintaining more efficient processes. Consequently, the design of DSS must allow for the functionality allowing determining the most appropriate route to improvement in the production of outputs, namely, whether to increase the level of inputs, or whether to improve the production processes first.

5.6. Step 6: System Requirement 1.6

The functionality of our DSS that addresses this system requirement is based on the capability of NN to model the process of conversion of inputs into outputs. Given a set of input nodes, representing inputs of the production process, and a set of output nodes, representing outputs of the production process, NN analysis allows the decision maker to create a *transformation function* representing the model of the input–output process. If that *transformation* function is saved, then it could be applied to a new set of production inputs with the purpose of generating a new set of the production outputs. Consequently, the decision maker can obtain a set of simulated inputs–outputs that can be subjected to DEA and the results compared to the results of DEA conducted using the original set.

Table 11
Using MR to identify complementarities between the relevant production variables.

Interaction term in the model	Subset	β Estimate	P @ 95%	Test of H0
$\boldsymbol{\beta_5}$ (investments in Telecoms∗Telecom staff)	Cluster 1	57.4954	< .0001	Rejected
$\boldsymbol{\beta_5}$ (investments in Telecoms∗Telecom staff)	Cluster 2	−2.1280	0.0087	Accepted

Table 12
Using Neural Networks to simulate the values of outputs of DEA model.

Inputs	Transformation function	Output	Outcome, Input–Output model	DEA model #
Cluster1	TF1	Cluster1: original	Actual inputs and outputs of Cluster 1	1
Cluster2	TF2	Cluster2: original	Actual inputs and outputs of Cluster 2	2
Cluster1	TF2	Cluster2: simulated	Simulated model of Cluster2, where the outputs are based on the inputs of Cluster1	3
Cluster2	TF1	Cluster2: simulated	Simulated model of Cluster2, where the outputs are based on the process of input–output conversion of Cluster1	4

Table 13
Results of output-oriented DEA, simulated.

Scenario	DEA model	CRS	VRS	NIRS	Interpretation
Simulated, Cluster2 utilizes the level of inputs of Cluster1	Cluster1, actual Cluster2, simulated	2.09 2.30	1.38 2.00	1.38 2.17	Members of Cluster2 should not pursue an increase in the level of inputs as a mean of increasing efficiency of output production.
Outcome	Is there a gain in efficiency for Cluster2 relative to Cluster1?	No, loss of 9.20%	No, loss of 30.87%	No, loss of 36.26%	
Simulated, Cluster2 utilizes the processes of input–output conversion of Cluster1	Cluster1, actual Cluster2, simulated	2.04 1.62	1.79 1.14	1.80 1.14	Members of Cluster2 should pursue the improvements in the production process as a mean of increasing efficiency of output production.
Outcome	Is there a gain in efficiency for Cluster2 relative to Cluster1?	Yes, gain of 25.62%	Yes, gain of 57.32%	Yes, gain of 57.45%	

In the case of our illustrative example NN analysis allows us to generate two transformation functions, *TF1* for Cluster1 and *TF2* for Cluster2. If we apply the inputs of Cluster1 to TF2 we can simulate the level of outputs that members of Cluster2 would have produced if they had the levels of inputs of Cluster1. Conversely, if we apply TF1 to inputs of Cluster2 we can simulate the level of outputs that members of Cluster2 would have obtained if they utilized process of conversion of inputs into outputs of Cluster1. Altogether, given 2 clusters, we end up with four DEA models listed in Table 12.

Once the NN analysis was conducted and the simulated values were saved, we ended up with four DEA models, two DEA models with the original values of inputs and outputs (DEA models 1 and 2), and two DEA models with the simulated values of inputs and outputs (DEA models 3 and 4). Both simulated DEA models were created for the purposes of gaining insights into the most appropriate route of improving the level of efficiency of Cluster2. Next, we re-run DEA analysis using the simulated models (DEA models 3 and 4) and obtained the new relative efficiency scores. By comparing the scores produced by the original models with the scores of the simulated models we can determine whether Cluster2 will get greater gains in efficiency of production of outputs from increasing the level of inputs (DEA model 3), or from improving the efficiency of conversion of inputs into outputs (DEA model 4). Results of this comparison are presented in Table 13.

These results allow the decision maker to answer Question 10:

Q10: *Whether the existing inefficiencies of the organizational entity are associated with the insufficient levels of inputs or with the inefficient processes of conversion of inputs into outputs?*

A10: *The existing inefficiencies of the organizational entities-members of Cluster2 are associated with the inefficient processes of conversion of inputs into outputs.*

6. Conclusion

In this paper we presented a DEA-centric DSS that provides facilities for assessing and managing the relative performance of productivity driven organizations that operate in unstable environments. The design of our DSS was guided by a set of system requirements (see Table 1) that are highly relevant to a productivity driven organization's efforts to identify and evaluate multiple productivity models in order to select the most suitable one for the given organization. These requirements suggested a coupling of the capabilities of DEA with capabilities of multiple data mining techniques as well as established theoretical frameworks (i.e., neo-classical growth accounting). The resulting DSS is applicable to different organizational levels, including the country level and the firm level. In this paper we demonstrated the feasibility and usability of this DSS on country-level organizational entities.

It should be noted that while other studies have combined data mining (DM) techniques with DEA, to the best of our knowledge this is the first study that has provided an integrated DEA-DM decision support model that can address the multiple productivity-related issues listed in Table 1. It should also be noted that while we utilized a specific set of data mining techniques that other techniques could also be utilized. For example, regressions splines could be used instead of regression. Similarly our DSS model allows for the utilization of other theoretical frameworks for addressing the issues such as complementarity. The results of this research suggests that additional exploration of integrated DEA-centric models involving multiple DM techniques and theoretical frameworks for addressing multiple productivity-related issues could be a fruitful area of design science research.

Appendix

See Table A1.

Table A1
Variables used to conduct CA.

Variables
Total telecom services revenue (% of GDP in current US $),
Total telecom services revenue per capita (Current US $),
Total telecom services revenue per worker (Current US $),
Total telecom services revenue per telecom worker (Current US $),
Annual telecom investment per capita (Current US $),
Annual telecom investment (% of GDP in current US $),
Annual telecom investment per worker (Current US $),
Annual telecom investment per telecom worker (Current US $).

References

[1] Andre F, Herrero I, Riesgo L. A modified DEA model to estimate the importance of objectives with an application to agricultural economics. Omega 2010;38(5):371–82.
[2] Avkiran N, Rowlands T. How to better identify the true managerial performance: state of the art using DEA. Omega 2008;36(2):317–24.
[3] Bezdek J. Pattern Recognition with Fuzzy Objective Function Algorithms. New York, NY: Plenum Press; 1981.
[4] Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press; 1995.
[5] Bock H. Probability models in partitional cluster analysis. Computational Statistics and Data Analysis 1996;23:5–28.
[6] Bollou F, Ngwenyama O. Are ICT investments paying off in Africa? An analysis of total factor productivity in six West African countries from 1995 to 2002 Information Technology for Development 2008;14(4):294–307.
[7] Botti L, Briec W, Cliquet G. Plural forms versus franchise and company-owned systems: a DEA approach of hotel chain performance. Omega 2009;37(3):566–78.
[8] Bougnol M-L, Dula JH, Estellita Lins MP, Moreira da Silva AC. Enhancing standard performance practices with DEA. Omega 2010;38(1-2):33–45.
[9] Bu N, Fukuda O, Tsuji T. EMG-based motion discrimination using a novel recurrent Neural Network. Journal of Intelligent Information Systems 2003;21(2):113–26.
[10] Çelebi D, Bayraktar D. An integrated neural network and Data Envelopment Analysis for supplier evaluation under incomplete information. Expert Systems with Applications 2008;35(4):1698–710.
[11] Chang H, Choy H, Cooper W, Ruefli T. Using Malmquist Indexes to measure changes in the productivity and efficiency of US accounting firms before and after the Sarbanes–Oxley Act. Omega 2009;37(5):951–60.
[12] Charnes A, Cooper WW, Lewin AY, Seiford LM. Data Envelopment Analysis: Theory, Methodology and Applications. Norwell, MA: Kluwer Academic Publishers; 1994.
[13] Choi YS, Yoo SI. Text Database Discovery on the Web: Neural Net Based Approach. Journal of Intelligent Information Systems 2001;16(1):5–20.
[14] Cook W, Bala K. Performance measurement and classification data in DEA: Input-oriented model. Omega 2007;35(1):39–52.
[15] Cook W, Hababou M. Sales performance measurement in bank branches. Omega 2001;29(4):299–307.
[16] Cook W, Liang L, Zhu J. Measuring performance of two-stage network structures by DEA: a review and future perspective. Omega 2010;38(6):423–30.
[17] Cook W, Seiford L, Zhu L. Models for performance benchmarking: measuring the effect of e-business activities on banking performance. Omega 2004;32(4):313–22.
[18] Cook WD, Zhu J. Data Envelopment Analysis. Modeling Operational Processes and Measuring Productivity 2008. ISBN/EAN13: 1434830233/9781434830234.
[19] Cooper WW, Tone K. Measures of inefficiency in Data Envelopment Analysis and Stochastic frontier estimation. European Journal of Operational Research 1997;99(1):72–88.
[20] Cooper WW, Seiford LM, Zhu J. Data envelopment analysis: History, models and interpreations. In: Cooper WW, Seiford LM, Zhu J, editors. Handbook on Data Envelopment Analysis, 1–39. Boston: Kluwer Academic Publishers; 2004. Chapter 1.
[21] Crestani F, Rijsbergen CJ van. A model for adaptive information retrieval. Journal of Intelligent Information Systems 1997;8(1):29–56.
[22] Cristofor, D. and Simovici, D. (2002). An information-theoretical approach to clustering categorical databases using genetic algorithms. Proceedings of the SIAM DM Workshop on Clustering High Dimensional Data, pp. 37–46. Arlington, VA.
[23] Dave R. Generalized Fuzzy C-Shells clustering and detection of circular and elliptic boundaries. Pattern Recognition 1992;25:713–22.
[24] Du J, Liang L, Chen Y, Bi G. DEA-based production planning. Omega 2010;38(1-2):105–12.
[25] Dula JH. Data Envelopment Analysis (DEA). In: Pardalos PM, Resende MGC, editors. Handbook of Applied Optimization. New York: Oxford University Press; 2002. p. 531–43.
[26] Eilat H, Golany B, Shtub A. R&D project evaluation: an integrated DEA and balanced scorecard approach. Omega 2008;36(5):895–912.
[27] Emrouznejad A, Shale E. A combined neural network and DEA for measuring efficiency of large scale datasets. Computers and Industrial Engineering 2009;56(1):249–54.
[28] Fu LiMin. Knowledge discovery by inductive neural networks. IEEE Transactions on Knowledge and Data Engineering 1999;11(6):992–8.
[29] Harb H, Chen L. Voice-based gender identification in multimedia applications. Journal of Intelligent Information Systems 2005;24(2–3):179–98.
[30] Hirschberg, J.G., and Lye, J.N. (2001). Clustering in a Data Envelopment Analysis using bootstrapped efficiency scores. Department of Economics – Working Paper Series 800, The University of Melbourne.
[31] Holsheimer M, Siebes APJM. Data mining: the search for knowledge in databases, Report CS-R9406. Amsterdam, The Netherlands: Centre for Mathematics and Computer Science; 1994.
[32] Jain A, Murty M, Flynn P. Data clustering: a review. ACM Computing Surveys 1999;31(3):264–323.
[33] Kao C. Malmquist productivity index based on common-weights DEA: the case of Taiwan forests after reorganization. Omega 2010;38(6):484–91.
[34] Kao C, Hung H. Efficiency analysis of university departments: an empirical study. Omega 2009;36(4):653–64.
[35] Khouja M. The use of Data Envelopment Analysis for technology selection. Computers and Industrial Engineering 1995;28(1):123–32.
[36] Lemos CAA, Lins MPE, Ebecken NFF. DEA implementation and clustering analysis using the K-means algorithm. In Data Mining VI – Data Mining, Text Mining and their Business Applications. Transactions of the Wessex Institute; 2005.
[37] Liu J, Lu W. DEA and ranking with the network-based approach: a case of R&D performance. Omega 2010;38(6):453–64.
[38] Lozano-Vivas A, Pastor J. Do performance and environmental conditions act as barriers for cross-border banking in Europe? Omega 2010;38(5):275–82.
[39] Lu H, Setiono R, Liu H. Effective data mining using Neural Networks. IEEE Transactions on Knowledge and Data Engineering 1996;8(6):957–61.
[40] McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: Lecam, L.M. and Neyman, J. (Eds.): Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.
[41] Morais P, Camanho A. Evaluation of performance of European cities with the aim to promote quality of life improvements. Omega 2011;39(4):398–409.
[42] Mostafa MM. A probabilistic neural network approach for modelling and classifying efficiency of GCC banks. International Journal of Business Performance Management 2009;11(3):236–58.
[43] Murtagh F. A survey of recent advances in hierarchical clustering algorithms which use cluster centers. Computer Journal 1983;26:354–9.
[44] Okazaki S. What do we know about mobile internet adopters? A cluster analysis Information and Management 2006;43(2):127–41.
[45] Osei-Bryson K-M, Inniss T. A hybrid clustering algorithm. Computers and Operations Research 2007;34(11):3255–69.
[46] Pao Y-H, Sobajic DJ. Neural networks and knowledge engineering. IEEE Transactions on Knowledge and Data Engineering 1991;3(2):185–92.
[47] Parthasarathy S, Anbazhagan N. Evaluating ERP projects using DEA and regression analysis. International Journal of Business Information Systems 2008;3(2):140–57.
[48] Piatkowski, M. (2003). Does ICT investment matter for output growth and labor productivity in transition economies? TIGER Working Paper Series, No. 47. December. Warsaw. Retrieved January 20, 2008 from www.tiger.edu.pl.
[49] Rai A, Tang X, Brown P, Keil M. Assimilation patterns in the use of electronic procurement innovations: a cluster analysis. Information and Management 2006;43(3):336–49.
[50] Ramanathan B, Yunfeng J. Incorporating cost and environmental factors in quality function deployment using Data Envelopment Analysis. Omega 2009;37(3):711–23.
[51] Razi M, Athappilly K. A comparative predictive analysis of Neural Networks, nonlinear regression and classification and regression tree (CART) Models. Expert Systems with Applications 2005;29(1):65–74.
[52] Samoilenko S. Information systems fitness and risk in IS development: insights and implications from chaos and complex systems theories. Information Systems Frontiers 2008;10(3):281–92.
[53] Samoilenko S. Contributing factors to information technology investment utilization in transition economies: an empirical investigation. Information Technology for Development 2008;14(1):52–75.
[54] Samoilenko S, Osei-Bryson KM. Strategies for telecoms to improve efficiency in the production of revenues: an empirical investigation in the context of transition economies. Journal of Global Information Technology Management 2008;11(4):56–75.
[55] Samoilenko S, Osei-Bryson K-M. Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and Decision Trees. Expert Systems with Applications 2007;34(2):1568–81.
[56] Samoilenko S, Osei-Bryson KM. An exploration of the effects of the interaction between ICT and labor force on economic growth in transitional economies. International Journal of Production Economics 2008;115(2):471–81.
[57] Samoilenko, S., and Green, L. (2008). Convergence and productive efficiency in the context of 18 transition economies: empirical investigation using DEA. Proceedings of the Southern Association for Information Systems Conference, Richmond, VA, USA March pp. 13–15.
[58] Seiford L, Zhu J. An investigation of returns to scale in data envelopment analysis. Omega 1999;27(1):1–11.

[59] Shao BBM, Lin WT. Measuring the value of information technology in technical efficiency with stochastic production frontiers. Information and Software Technology 2001;43(7):447–56.
[60] Shimshak D, Lenard M, Klimberg R. Incorporating quality into Data Envelopment Analysis of nursing home performance: a case study. Omega 2009;37(3):672–85.
[61] Shin HW, Sohn SY. Multi-attribute scoring method for mobile telecommunication subscribers. Expert Systems with Applications 2004;26(3):363–8.
[62] Sohn S, Moon T. Decision Tree based on Data Envelopment Analysis for effective technology commercialization. Expert Systems with Applications 2004;26(2):279–84.
[63] Stewart T. Goal directed benchmarking for organizational efficiency. Omega 2010;38(6):534–9.
[64] Theodorou P, Florou G. Manufacturing strategies and financial performance – the effect of advanced information technology: CAD/CAM systems. Omega 2008;36(1):107–21.
[65] Tone K, Tsutsui M. Dynamic DEA: a slacks-based measure approach. Omega 2010;38(3–4):145–56.
[66] Vai M, Xu Z. Representing knowledge by neural networks for qualitative analysis and reasoning. IEEE Transactions on Knowledge and Data Engineering 1995;7(5):683–90.
[67] Wallace L, Keil M, Rai A. Understanding software project risk: a cluster analysis. Information and Management 2004;42:115–25.
[68] Ward J. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 1963;58:236–44.
[69] Wiederhold G. Mediators in the architecture of future information systems. IEEE Computer 1992;25(3):8–49.
[70] Wiederhold, G.C.M., Walker, M.G., Blum, R.L. and Downs S.M. (1986). Acquisition of knowledge from data. Proceedings of ACM SIGART ISMIS, pp. 74–84.
[71] Wu D. Supplier selection: a hybrid model using DEA, decision tree and neural network. Expert Systems with Applications 2009;36(5):9105–12.
[72] Wu M-C, Lin S-Y, Lin C-H. An effective application of decision tree to stock trading. Expert Systems with Applications 2006;31(2):270–4.
[73] Yu M, Lin E. Efficiency and effectiveness in railway performance using a multi-activity network DEA model. Omega 2008;36(6):1005–17.
[74] Zhong W, Yuan W, Li S, Huang Z. The performance evaluation of regional R&D investments in China: an application of DEA based on the first official China economic census data. Omega 2011;39(4):447–55.