CSCanada
Canadian Social Science
Vol. 10, No. 1, 2014, pp. 103-110
DOI:10.3968/j.css.1923669720141001.3046

ISSN 1712-8056[Print]
ISSN 1923-6697[Online]
www.cscanada.net
www.cscanada.org

Efficiency Measurement in Turkish Coal Enterprises Using Data Envelopment Analysis and Data Mining[1]

Fehim Bakirci[a]; Emre Yakut[b],*; Ayhan Demirci[c]; Murat Gündüz[d]

[a]Associate professor. PhD. Faculty of Economics and Administrative Sciences, Ataturk University, Erzurum, Turkey.
[b]Assistant professor. Faculty of Economics and Administrative Sciences, Osmaniye Korkut Ata University, Osmaniye, Turkey.
[c]PhD. Analyst.
[d]Assistant professor. Faculty of Economics and Administrative Sciences, Usak University, Usak, Turkey.
*Corresponding author.

Received 5 October 2013; accepted 19 February 2014

Abstract

Gradual population growths, skyrocketing technological developments and inter-State competitions increase the energy demands continuously. Although the countries try to diverse their energy sources in order to sustain their developments, they also have to pay attention to protect their energy independences. Thus, it is very important to develop their self-resources. Coal is the most common natural source which can meet our energy needs. However, coal mine enterprises have to be administrated cost-effectively in order get minimum energy costs. In this study, the efficiency of Turkish coal enterprises between the years 2003-2010 is measured by using Data Envelopment Analysis (DEA). Then, indicators which are the most important in estimating the efficiency were determined by using the efficiency scores obtained by DEA in the Data Mining technique.

Key words: Coal enterprises; Efficiency; Data envelopment analysis; Data mining

Fehim Bakirci; Emre Yakut; Ayhan Demirci; Murat Gündüz (2014). Efficiency Measurement in Turkish Coal Enterprises Using Data Envelopment Analysis and Data Mining. *Canadian Social Science*, *10*(1), 103-110. Available from: http://www.cscanada.net/index.php/css/article/view/j.css.1923669720141001.3046
DOI: http://dx.doi.org/10.3968/j.css.1923669720141001.3046

INTRODUCTION

It is seen that countries are continually planning new industry and technology investments and working to carry them out in order to achieve economic and social development and elevate their welfare level. Energy carries a critical significance as the main input in terms of industrialization and technological development. For the sake of acquiring energy, having low costs, being easy and continuously attainable from a safe source are the primary subjects. For this reason, countries principally prefer local resources. Even though our country has sufficient local resources, importation has been given precedence in the recent years. Especially natural gas acquired from abroad with the "take-or-pay" method takes place on the top. Ignoring local sources, however, carries important risks in terms of energy safety. Energy crises from around the world that broke out due to several reasons are pushing countries to be more sensitive and rational. Therefore, local resources in our country must be given the due importance and developed (Tamzok & Torun, 2005, p.1).

The most important local source to provide the energy need in our country is coal. Compared to other fossil fuels, its cost, ease of transportation, convenience of stocking feasibility, being safe and secure in terms of easy use, the cheapness of supply to customer and its sustainability are among the qualities that make it a noteworthy energy resource. In order to reduce foreign dependency on energy, coal mining must be carried out in a more productive way. For this purpose, the efficiency of coal mining in our country between 2003 and 2010 are analysed in this study. The method used for analysing the efficiency is Data Envelopment Analysis (DEA). Data mining techniques were used in order to determine which indicators are important in the estimation of efficiency according to DEA.

[1] This study was presented as a paper at the 13th International Econometrics, Operational Research and Statistics Conference held on 24-26 May 2012 in the Turkish Republic of Northern Cyprus.

Copyright © Canadian Academy of Oriental and Occidental Culture

1. PERFORMANCE ANALYSIS

In current competitive market conditions, efficiency and productivity on every level of the production is a highly important notion that confronts us. Productivity is defined as the ratio between inputs used during production and outputs acquired after process; whereas, the notion of efficiency reflects a broader meaning and can be defined as "the sustainability of economic goals presented by decision making units (DMU)" (Sinha, 2008, p.7-27). From this viewpoint, DMUs have to control their performances at all times and level up as possible as it gets. A performance of a DMU can only be accepted as on a 100% level only in the conditions below (Aydagün, 2003, p.9);

a. None of the outputs can be increased unless;
   i. One or more inputs are increased or
   ii. Some of the other outputs are decreased.
b. None of the inputs can be decreased unless;
   i. Some of the outputs are decreased or
   ii. Some other inputs are increased.
c. Any DMU is assumed to have reached 100% relative efficiency only if other related DMUs can bring evidence on the inefficiency of input or output use.

There are different methods used for performance evaluation. Among these, ratio analysis is known to be the easiest method to perform. This method can be used widely for DMUs that produce single input and single output. The method is carried out by figures acquired by proportioning outputs to inputs, sorting the performance from the best to the lowest. It is clear that only the DMUs that are being evaluated can be sorted and the reasons for inefficiency for the inefficient DMUs cannot be determined through this method.

Another performance evaluation method is regression analysis. Being a parametric method, regression analysis is widely used in the cases where more than one input and output are acquired. In this method, DMUs are evaluated according to average performance. Therefore, it does not enable the improvement of the best DMUs, even shows the tendency of dragging them to average as a result.

There will be a need for a new method considering that today's production qualities of DMUs do not fit in single input-single output or multiple input-single output qualities, production processes are taking a more complex form that uses multiple inputs and produce multiple outputs and both methods that are introduced above are insufficient in terms of performance evaluation of today's DMUs. Non-parametric methods that do not look for any parameter as a prerequisite have been useful for removing the weaknesses of the methods mentioned above.

1.1 Data Envelopment Analysis

The history of DEA that has been widely used by both public and private sector in recent years, started with the doctoral dissertation "City and Public" by Edwardo Rhodes of Cornegie Mellon University. In this study, the performances of groups that participated in Program Follow Through and those who did not were compared. The desire of estimating Farrell's single input-output technical efficiency evaluation in 1957 and the relative technical productivity of 70 schools by ignoring prices with multiple inputs and outputs, brought forth DEA proportional formula which is also known as CCR (Charnes, Cooper, Rhodes) and the first article about this subject was published at Journal of Operations Research in 1978.

Introduced and developed by Charnes, Cooper, Rhodes based on DEA models, CCR model measures overall efficiency under the assumption of fixed return to scale. Introduced and developed by Banker's and Banker, Charnes, Cooper's works, BCC model measures only technical efficiency by comparing units on similar scale under the assumption of variable return to scale. Thus BCC model enables research, under the assumption of DMU's variable or fixed return to scale in the case of multiple inputs and multiple outputs.

Data Envelopment Analysis (DEA) was developed to relatively evaluate the economic DMU's efficiency that are similar in the production process and is an evaluation technique that is non-parametric and widely used as it calculates independently in units of measure. Without depending on any predetermined functional relation, DEA can analyse using multiple inputs and multiple outputs, it can determine the inefficiency of each DMU by quantity and source, therefore helping generate the improvement policies. Thanks to these qualities, it has a popular use among several material and service production fields (Bakırcı, 2006, p.167).

With the help of an ordinary statistical method, relative efficiency evaluation can be carried out by drawing on DMU's central tendency evaluation. DEA, on the opposite, is an end-point method and evaluation is carried out by comparing DMUs with the best one.

There are mainly three stages in efficiency evaluation with DEA (Golany & Roll, 1989, pp.237-250);

1. The qualification and election of DMUs that are going to be analysed,
2. The determination of convenient input and output factor variables for the evaluation of relative efficiency of chosen DMUs,
3. The application of DEA models and analysis of results.

1.2 Mathematical Structure of DEA Technique

DEA weighs multiple input and output values together in a linear way. Thus, weighted total input that shows linear weighted sum of the company inputs are calculated as follows:

$$\text{Weighted Total Input} = \sum_{i=1}^{I} v_i x_i \tag{1}$$

Here, v_i is the weight specified for x_i input at the time of conjugation.

Copyright © Canadian Academy of Oriental and Occidental Culture

Similarly, company's weighted total output is calculated by linear weighted aggregate of all outputs.

$$\text{Weighted Total Output} = \sum_{j=1}^{J} u_j y_j \quad (2)$$

Here, u_j is the weight specified for y_j output. Efficiency of weighted total inputs and outputs and efficiency of DMUs, which convert inputs to outputs, are defined as the ratio of inputs to outputs and it is formulated as follows (Ramanathan, 2003, p.39):

Efficiency = Weighted Total Output/Weighted Total Input

$$= \frac{\sum_{j=1}^{J} u_j y_j}{\sum_{i=1}^{I} v_i x_i} \quad (3)$$

Considering that the DMU with the best performance will get 1,0 value, the following limitations should be added to the formula.

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{s} v_i x_{ij}} \leq 1; \qquad j = 1, 2, \ldots\ldots, n$$

$$u_r, v_i \geq o, \qquad r = 1, \ldots\ldots, s \ , \quad i = 1, \ldots\ldots, m \quad (4)$$

Here, y_{rj} and x_{ij} being positive, show j^{th} KVB's input and output amount. In the same way, u_r, $v_i \geq 0$ equation shows weights of these inputs and outputs.

1.3 Strengths and Weaknesses of DEA Technique

The facts that it is supported by economic theories and methods, it focuses on relative efficiency rather than absolute efficiency, it is able to include multiple inputs and outputs co-ordinately in calculation and its ability to determine the optimum sample and specify it as the target make DEA possible to be used widely (Kontodimonopoulos vd., 2007, pp.5-14).

DEA analyse performance based on the optimum not on the statistical population average. For each DMU, an optimum sample and a border are determined. In line with this border, it is decided whether each unit is efficient or inefficient according to the coordinate. With these features, DEA is preferred as an appealing and useful method (Mok vd., 2007, pp.259-274).

Another advantage of DEA is that there isn't any specific functional structure or a behavioural prerequisite. Technological infrastructure between DMUs is completely indefinite and variable. Analysis structure of activities in linear aggregate is natural. Each DMU reaches different production amounts by using separate production process and production plan (Biesebroeck, 2007, pp.529-569).

A significant advantage of DEA as a non-parametric deterministic method used in efficiency measurement is that it can reach a conclusion with a few numbers of observation sets (Pasiouras, 2008, pp.1121-1130).

In comparison with the above mentioned advantages, DEA also has some weaknesses. The perception of the results obtained by DEA that the efficiency of a DMU with a relative efficiency score of 1.0 among the production units in a set cannot increase more is an important obstacle for performance increase. However, this score points out that the DMU in question was determined to be efficient within the existing data set. Another weakness is that there is no hypothesis test for DEA due to its un-parametric nature, thus significance levels of the observed differences cannot be explained statistically (Pereira, 2006, pp.308-315).

DEA is criticized for the deterministic structure of the method, which causes deviations from efficiency border that are accepted as inefficiency. This method is sensitive to measuring errors and modelling mistakes in data. (Hansson, 2007, pp.77-88).

Original DEA model is not able to sort efficient DMUs in a specific way (Zzadeh vd., 2008, pp.1352-1357).

DEA only provides a result within the examined set. This means that there may be a more efficient DMU outside this set. Therefore, the result of the analysis reveals not the most efficient DMU but the most efficient DMU within the existing data set.

2. DATA MINING

Data mining is the process of extracting hidden patterns from a huge amount of data (Kantardzic, 2003).

Data mining is the process of discovering significant and useful relations and models from a data set. As this process has a "discovery" oriented nature, some sources name data mining as "the process of information discovery in databases" (Piramuthu, 2004, p.483).

Data mining is about acquiring the "valuable" information from large scale data. Thus, it is possible to present the relations between data and make assumptions for the future when it is necessary. In this sense, data mining can be evaluated as the process of revealing all hidden information that may be available or may emerge in the future by all the data using certain methods. Classic statistical practices are operated on sufficiently organized and mainly summary data. In data mining, on the other hand, millions, even billions of data and many more variables are handled (Özkan, 2008, p.38).

Advancements in technology enable companies to stock wide range of information on work which is generated every day and compile them together. Therefore, wide-scale databases have emerged and we need something to convert raw data to useful information. Data mining is used for analysing information stored in computers. Data mining techniques have a wide range of application including banking, retailing, insurance and telecommunication (Olson & Delen, 2008, pp.3-8).

Copyright © Canadian Academy of Oriental and Occidental Culture

It is possible to divide data mining into two categories, supervised and unsupervised. An auditor divides related classes according to predetermined criteria, and various examples are given for each class in supervised learning, also named as learning from the sample. The aim of the system is finding qualities of each class with the given examples and stating these qualities with norm sentences (Akpınar, 2000, pp.1-22).

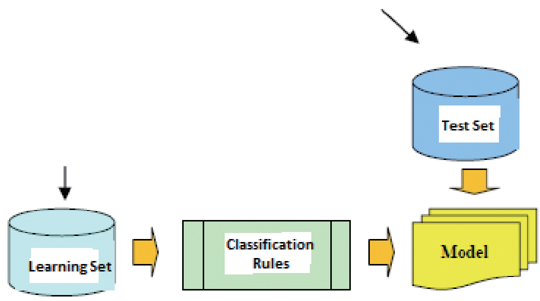

Figure 1
Supervised Learning
Source: Akpınar, 2000, p.14

Unsupervised methods mainly used for understanding, knowing and exploring the data and provide an insight for the future methods to be applied. They also aim to discover the disguised information of input variables (Tsiptsis & Chorianopoulos, 2009, pp.3-4). On the other hand, supervised methods are used to obtain information and result from the data.

Main data mining methods can be classified as follows depending on being supervised or unsupervised (Koyuncugil & Özgülbaş, 2008, p.5):

Supervised
- k-Nearest-Neighbour
- K-means clustering
- Regression models
- Rule induction
- Decision trees
- Neural networks

Unsupervised
- Hierarchical clustering
- Self-organized maps

Decision trees as a method of supervised data mining are classification methods which are easier to understand and interpret when compared to other methods. The model created by using decision tress looks like an upside-down tree. This tree is composed of nodes showing decision making points and branches connecting these nodes. At the top is the root node. This node is tested for a number of features and branches are created according to the results of this test. Each branch is connected to a new decision node and branches are created there as well by testing a number of new features. At the bottom of the tree structure are leaf nodes which no more reproduce new branches (Seyrek & Ata, 2010, p.72). Commonly used decision tree models (algorithms) include ID3 (Iterative Dichotomiser 3), C4.5, C5.0, CART (Classification and Regression Tree), CHAID (Chi-squared Automatic Interaction detection) and QUEST (Quick, Unbiased and Efficient Statistical Tree). ID3 later evolved into C4.5 (Quinlan, 1993), and this was an important with regard to the splitting rule and the calculation method. C5.0 is a commercial version of C4.5, and is available as a closed-source product (Han and Kamber, 2007, pp.291-310).

As it is easy to understand and interpret, this study used decision tree as the classification method and worked with C5.0 classification method, which is the new developed version of C4.5 algorithm.

3. APPLICATION TO COAL ENTERPRISES

This study aimed to examine efficiency of a total number of 8 DMUs including 4 establishment directorates which are affiliated to TKİ (Turkish Coal Enterprises) between the years of 2003 and 2010.

DMUs included in the scope of this study are as follows: Aegean Lignite Enterprise Establishment Directorate (ELI), South Aegean Lignite Enterprise Establishment Directorate (GELI), West Lignite Enterprise Establishment Directorate (GLI), Seyitomer Lignite Enterprise Establishment Directorate (SLI), Can Lignite Enterprise Directorate (CLI), Yenikoy Lignite Enterprise Directorate (YLI), Ilgin Lignite Enterprise Directorate (ELI) and Bursa Lignite Enterprise Directorate (BLI). Data used for efficiency measurement was obtained from TKI's official web site. Data used in the study are shown in Table 1.

Table 1
Input and Output Data

Input data	Output data
X1: Area (Hectare)	Y1: Production Amount(ton)
X2: Coal Reserve (1000 tons)	Y2: Sellable Amount (ton)
X3: Total number of Personnel	Y3: Sales Amount (ton)
X4: Number of vehicles	
X5: Investment (TL)	

In this study, Data Envelopment Analysis, which is an efficiency measurement method without a parameter that is widely used in cases with multiple input and multiple output production, was used. By applying this method, efficiency scores of DMUs were found and afterwards data mining was performed in order to determine which variables contribute the most to these scores. Frontier Analyst Professional package program was used for data envelopment analysis measurement while C5.0 algorithm was used for data mining.

3.1 Efficiency Scores by CCR Model

Efficiency was measured using CCR model, which measures fixed return to scale, and observed efficiency scores are shown in Table 2.

Copyright © Canadian Academy of Oriental and Occidental Culture

Table 2
Efficiency Scores According to CCR Method Analysis

Coal enterprises	2003	2004	2005	2006	2007	2008	2009	2010
ELI	84,55	100	57,85	64,92	65,1	64,08	100	54,14
CLI	30,1	56,89	62,13	83,74	100	90,38	99,11	97,76
GELI	75,74	74,89	81,93	100	83,88	78,48	96,58	74,29
YLI	100	100	100	100	100	100	100	100
GLI	95,54	48,16	33,08	41,06	44,35	35,9	99,54	40,08
ILI	100	69,29	25,8	16,55	4,6	8,86	97,36	66,42
SLI	100	100	100	100	100	100	100	100
BLI	34,95	42,62	17,25	46,13	92,59	43,75	100	41,99

According to the results of data envelopment analysis based on CCR method, YLI and SLI among 8 affiliate DMUs of TKI achieved full efficiency score in all years. Furthermore, CLI managed to increase its efficiency regularly over the years.

Taking into consideration the last year of the obtained data, detailed information on DMUs potential improvement in line with data envelopment analysis based on CCR models is shown in Table 3. This table shows which inputs should be decreased how much or which outputs should be increased how much in order to make these inefficient DMUs become efficient. As a result of this analysis of the year 2010, SLI managed to become reference for 6 times and YLI for once.

Table 3
Potential Improvement Results for 2010 as per CCR Method

	ELI	CLI	GELI	YLI	GLI	ILI	SLI	BLI
X1	-45	-2	-78	-	-68	-94	-	-90
X2	-57	-22	-25	-	-59	-97	-	-58
X3	-52	-36	-38	-	-76	-66	-	-64
X4	-45	-51	-45	-	-63	-71	-	-76
X5	-82	-83	-94	-	-97	-33	-	-89
Y1	18	104	1	-	19	-	-	65
Y2	18	95	-	-	18	-	-	56
Y3	-	-	5	-	-	21	-	-
Reference Set	YLI SLI	SLI	SLI	-	SLI	SLI	-	SLI

3.2 Efficiency Scores by BCC Model

Efficiency was measured using BCC model measuring variable return to scale and observed efficiency score are shown in Table 4.

According to data envelopment analysis based on BCC method; ELI, CLI and ILI were also found to be fully efficient in addition to YLI and SLI.

Detailed information on DMUs potential improvement according to data envelopment analysis based on BCC method for the same year is shown in Table 5. The table shows which inputs should be decreased how much or which outputs should be increased how much in order to make these inefficient DMUs become efficient. As a result of this analysis of the year 2010, CLI managed to become reference for 3 times, SLI for 3 times and YLI for once.

Table 4
Efficiency Scores as per Analysis Based on BCC Method

Coal enterprises	2003	2004	2005	2006	2007	2008	2009	2010
ELI	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
CLI	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
GELI	76,28	76,75	93,01	100,00	87,92	83,99	73,64	88,38
YLI	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
GLI	97,71	49,02	51,22	47,95	50,31	44,37	41,30	48,17
ILI	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
SLI	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
BLI	41,93	50,72	100,00	100,00	100,00	83,67	100,00	91,09

Copyright © Canadian Academy of Oriental and Occidental Culture

Table 5
Potential Improvement Results for 2010 as per BCC Method

	ELI	CLI	GELI	YLI	GLI	ILI	SLI	BLI
X1	-	-	-75	-	-54	-	-	-82
X2	-	-	-11	-	-51	-	-	-8
X3	-	-	-22	-	-73	-	-	-9
X4	-	-	-25	-	-51	-	-	-22
X5	-	-	-88	-	-92	-	-	-8
Y1	-	-	-	-	-	-	-	77
Y2	-	-	-	-	-	-	-	73
Y3	-	-	18	-	1	-	-	77
Reference set	-	-	CLI SLI	-	CLI YLI SLI	-	-	CLI SLI

3.3 Classification by Data Mining

At this stage of the study, 5 input variables and 3 output variables, which were used in data envelopment analysis, were entered into C5.0 algorithm along with a total of 64 records of efficiency scores. As stated before, C5.0 is a supervised algorithm and it, first of all, requires the decision tree to be trained by a data set with a known target variable. That's why 47 data was, which were randomly selected among 64 data sets, were used as a training data set in order to create a model. Model was trained by this 80% and then tested with the remaining %20. Two decision trees were created based on BCC method and CCR method.

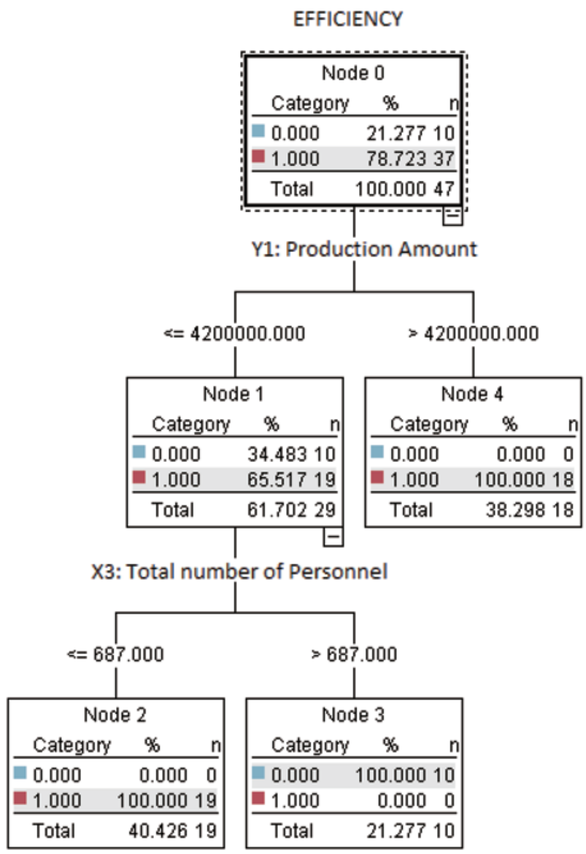

Figure 2
Decision Tree as per CCR Method

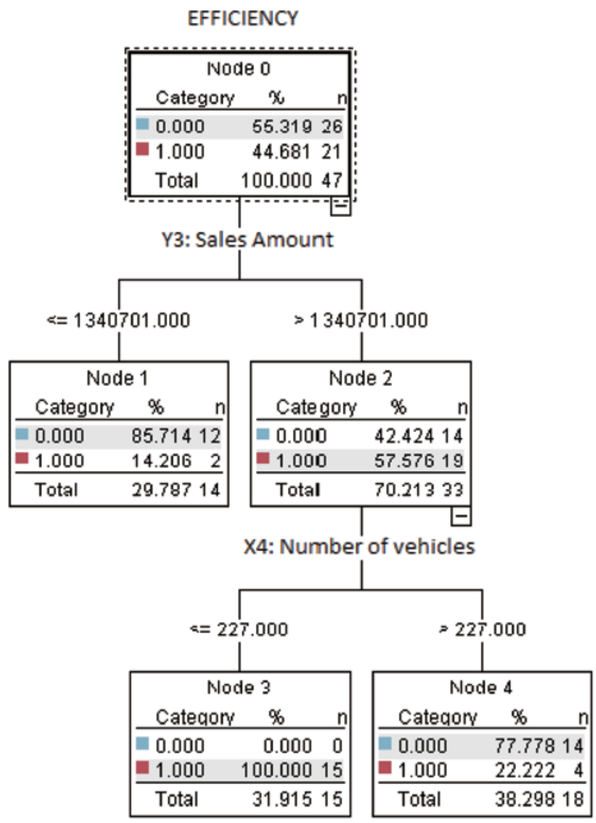

Figure 3
Decision Tree as per BCC Method

3.3.1 Decision Tree Created According to DEA Method

The variable that has the strongest effect on calculating the efficiency in root of the decision tree is Y3 variable, which is "sales amount". In the root of the tree, which included a training set of 47 companies, 21 companies were efficient while 26 companies were inefficient. According to Figure 2, among 47 efficiency scores, production amount of 29.8% is equal to or less than 1,340,701 tons while production amount of 70.2% is more than 1,340,701 tons. According to decision tree, if the sales amount of a coal enterprise is equal to or less than 1.340.701, that company is 85.7% inefficient. Among companies, which were selected as sample, 12 (85.7%) of them were inefficient while 2 (14.3%) of them were efficient. Sales amount of

Copyright © Canadian Academy of Oriental and Occidental Culture

33 enterprises are more than 1,340,701 tons. In this case, 19 (57.6%) enterprises are efficient while 14 (42.4%) of them are inefficient.

With the prerequisite of having a sales amount more than 1,340,701 tons, X4 variable, which is "number of vehicles", is equal to or less than 227, that enterprise is 100% efficient. The study shows that all 15 of the total 15 enterprises, which have such efficiency scores, are efficient. Considering the prerequisite, there are 18 enterprises with more than 227 vehicles. In this case, 14 (77.8%) of them are inefficient while 4 (22.2%) of them are efficient.

Data mining was performed according to efficiency scores obtained from the analysis of 64 enterprises by using CCR method. Data mining estimated 14 out of 17 inefficient enterprises and 41 out of 47 efficient enterprises correct. Results of the analysis are shown in Table 6.

Table 6
Data Mining Analysis as per CCR Method: Success of Efficiency Classification

Data mining analysis		Estimated group			
		Inefficient	Efficient	Total	Accuracy percentage
Observed group	Inefficient	**14**	3	17	82.4
	Efficient	6	**41**	47	87.2
	Total	20	44	**64**	85.9

Table 6 indicates that accuracy of classification by data mining method is 82.4% for inefficient enterprises and 87.2% for efficient enterprises. Overall success of classification by CCR method is 85.9%.

3.3.2 Decision Tree Created According to BCC Method

The variable that has the strongest effect on calculating the efficiency in root of the decision tree is Y1 variable, which is "production amount". In the root of the tree, which included a training set of 47 companies, 37 companies were efficient while 10 companies were inefficient. According to Figure 3, among 47 efficiency scores, production amount of 61.7% is equal to or less than 4,200,000 tons while production amount of 38.3% is more than 4,200,000 tons. According to decision tree, if the production amount of a coal enterprise is equal to or less than 4.200.000 tons, that company is 65.5% inefficient. Among companies, which were selected as sample, 19 (65.5%) of them were inefficient while 10 (34.5%) of them were efficient. Production amount of 18 enterprises are more than 4.200.000 tons. In this case, 18 of the total 18 enterprises (100%) are efficient.

With the prerequisite of having a production amount equal to or less than 4,200,000 tons, if X3 variable, which is "total number of personnel", is more than 687, that enterprise is 100% inefficient. The study shows that 10 of the total 10 enterprises, which have such efficiency scores, are inefficient. Considering the prerequisite, there are 19 enterprises, total personnel number of which are equal to or less than 687. In this case, 19 of the total 19 enterprises are efficient.

Data mining was performed according to efficiency scores obtained from the analysis of 64 enterprises by using BCC method. Data mining estimated 12 out of 17 inefficient enterprises and 47 out of 47 efficient enterprises correct. Results of the analysis are shown in Table 7.

Table 7
Data Mining Analysis as per BCC Method: Success of Efficiency Classification

Data mining analysis		Estimated group			
		Inefficient	Efficient	Total	Accuracy percentage
Observed group	Inefficient	**12**	5	17	70,6
	Efficient	-	**47**	47	100
	Total	12	52	**64**	92,2

Table 7 indicates that accuracy of classification by data mining method is 70.6% for inefficient enterprises and 100% for efficient enterprises. Overall success of classification by BCC method is 92.2%.

CONCLUSION

In this study, efficiency of DMUs subject to analysis in the period of 2003-2010 was measured by data envelopment analysis. Both CCR method measuring fixed return to scale and BCC method measuring variable return to scale was used for the purpose of analysis, and efficiency scores obtained by both methods were compared.

As a result of efficiency analyses, 8 DMUs were analysed only according to data used in analysis, and YLI and SLI proved to be fully efficient for all the years, which were analysed. When analysis was repeated with BBC method; ELI, CLI and ILI also proved to be fully efficient in addition to these two DMUs.

The main difference between these two methods is

Copyright © Canadian Academy of Oriental and Occidental Culture

that BCC method is more flexible than CCR method when it comes to determining input and output weights thanks to its ability to measure variable return to scale. Thus, it is deduced that it provides more reliable results in measurement.

In the second analysis, decision trees were created according to BCC and CCR methods based on C5.0 algorithm, which is a data mining method, and efficiencies were estimated.

According to BCC method, the most significant variable in the root of decision tree was Y1 variable, i.e., "production amount", followed by X3 variable, i.e., "total number of personnel". According to CCR method, the most significant variable in the root of decision tree was Y3 variable, i.e., "sales amount", followed by the branch of X4 variable, i.e., "number of vehicles".

Classification success of BCC method was 70.6% for inefficient enterprises and 100% for efficient enterprises. Overall classification success was 99.2%. Classification success of CCR method was 82.4% for inefficient enterprises and 87.2% for efficient enterprises. Overall classification success was 85.9%.

It can be stated that classification success of BCC method based on C5.0 algorithm, which is used for estimating efficiency statuses, gives more reliable results than CCR method.

REFERENCES

Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği [Knowledge discovery in databases and data mining]. *İstanbul Üniversitesi İşletme Fakültesi Dergisi* [*Istanbul University Faculty of Management Journal*], *29*(1), 1-22. (In Turkish).

Aydagün, A. (2003). Veri Zarflama Analizi. [Data envelopment analysis]. Hava Harp Okulu Havacılık ve Uzay Teknolojileri Enstitüsü, HUTEN Yıl Sonu Semineri, İstanbul. [The Air Force Academy Institute of Aeronautics and Space Technology, HUT Year-end Seminar, Istanbul]. (In Turkish).

Bakırcı, F. (2006). *Üretimde Etkinlik ve Verimlilik Ölçümü, Veri Zarflama Analizi, Teori ve Uygulama* [*Production efficiency and productivity measurement, data envelopment analysis, theory and application*]. İstanbul: Atlas Yayınları. (In Turkish).

Biesebroeck, J. V. (2007). Robustness of productivity estimates. *The Journal of Industrial Economics*, *LV*(3), 529-569.

Golany, B., & Roll, Y. (1989). An aplication procedure for DEA. *International Journal of Management Science*, *17*(3), 237-250.

Han, J., & Kamber, M. (2007). *Data mining: Concepts and techniques*. Amsterdam: Elsevier.

Hansson, H. (2007). The links between management's critical success factors and farm level economic performance on dairy farms in Sweden. *Food Economics, Acta Agriculturae Scandinavica, Section C*, *4*(2), 77-88.

Kantardzic, M. (2003). *Data mining: Concepts, models, methods, and algorithms*. New Jersey: John Wiley & Sons.

Kontodimonopoulos, N., Moschovakis, G., Aletras, V., & Niakas, D. (2007). The effect of environmental factors on technical and scale efficiency of primary health care providers in greece, cost effectiveness and resource allocation. BioMed Central.

Koyuncugil, A. S., & Özgülbaş N. (2008). İMKB'de İşlem Gören Kobi'lerin Güçlü ve Zayıf Yönleri: CHAİD Karar Ağacı Uygulaması. [Appication of strengths and weaknesses of SMEs: CHAID decision tree application]. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Dergisi* [*Dokuz Eylul University Faculty of Economics and Administrative Sciences Journal*], *23*(1), 1-21.

Mok, V., Yeung, G., Han, Z., & Li, Z. (2007). Leverage, technical efficiency and profitability: An application of DEA to foreign-ınvested toy manufactoring firms in China. *Journal of Contemporary China*, *16*(51), 259-274.

Olson, D. & Delen, D. (2008). *Advanced data mining techniques*. Heidelberg: Springer-Verlag.

Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*, 1. Baskı [*Data mining methods* (1st ed.)]. İstanbul: Papatya Yayıncılık. (In Turkish).

Pasiouras, F., Liadaki, A., & Zopounidis, C., (2008). Bank efficiency and share performance: Evidence from Greece. *Applied Financial Economics*, *18*, 1121-1130.

Pereira, A. (2006). Economies of scale in blood banking: A study based on data envelopment analysis. *Vox Sanguinis*, *90*, 308-315.

Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research*, *156*(2), 483-494.

Quinlan, J. R. (1993). C4.5: Programs for machine learning. San Fransisco: Morgan Kaufman.

Ramanathan, R. (2003). *An ıntroduction to data envelopment analysis: A tool for performance measurement*. California USA: Sage Publications.

Seyrek, İ. H., & Ata, H. A. (2010). Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü. [Data envelopment analysis and data mining efficiency measurement in banks]. *BDDK Bankacılık ve Finansal Piyasalar* [*BRSA Banking and Financial Markets*], *4*(2), 67-84. (In Turkish).

Sinha, R. P. (2008). Business efficiency of public sector commercial banks: A data envelopment approach. *ICFAI Journal of Applied Economics*, 7-27.

Tamzok, N., & Torun, M. (2005, December 21-23). Türkiye Enerji Politikaları İçerisinde Kömürün Önemi, TMMOB V [Importance of coal energy policy in Turkey, Chamber V]. *Enerji Sempozyumu, TMMOB* [*Energy Symposium, Chamber*], Ankara.

Tsiptsis, K., & Chorianopoulos, A. (2009). Data mining techniques in CRM ınside customer segmentation. USA: John Wiley & Sons Publishing.

Zzadeh, A., Ghaderi, S. F., Javaheri, Z., & Saberi, M. (2008). A fuzzy mathematical programming approach to DEA models. *American Journal of Applied Sciences*, *5*(10), 1352-1357.

Copyright © Canadian Academy of Oriental and Occidental Culture