Machine Learning Clustering Algorithms Based on the DEA Optimization Approach for Banking System in Developing Countries

Mirpouya Mirmozaffari, Azam Boskabadi, Gohar Azeem, Reza Massah, Elahe Boskabadi, Hamidreza Ahady Dolatsara, and Ata Liravian

***Abstract* — The primary purpose of this paper is to combine Data Envelopment Analysis (DEA) optimization approach with machine learning clustering method in datamining in order to introduce the most efficient DEA Decision-Making Units (DMUs) and the best Clustering algorithm respectively. The main goal of this paper in optimization part is to evaluate bank efficiency with cross-efficiency over 2014-2019 with Data Envelopment Analysis (DEA) for 12 banks from two developing countries. The cross-efficiency evaluation is an extension of DEA that provides a ranking method and eliminates unrealistic DEA weighting schemes on weight restrictions, without requiring prior information. Applying cross-efficiency can be beneficial for managers to expand their comparison and evaluation. The ranking of decision-making units (DMUs) is one of the most critical topics in efficiency assessment. To find the superior model, we consider input-oriented BCC-CCR and CCR-BCC models. This study overcomes with some data and methodology issues in measuring the efficiency of developing country's banks and highlights the importance of inspiring increased efficiency through the banking industry comparing new suggested models and the new results. After applying the optimization step, in the second part, in Machine learning step, clustering method has been applied. Clustering is the procedure of grouping similar items together. This group of the items is called the cluster. Different clustering algorithms can be used according to the behavior of data. Farthest First and Expectation Maximization algorithms have been applied. Finally, BCC-CCR and Farthest First algorithms have been proposed as a superior optimization model and machine learning algorithm, respectively.**

***Index Terms* — Machine Learning, Optimization, Data Envelopment Analysis, Data Mining, Clustering, Cross-Efficiency, Banking System.**

I. Introduction

Despite the unprecedented growth in the banking industry in developing countries, research on the performance and efficiency of this industry is almost challenging. Therefore, one of the aims of this study is measuring efficiency levels at the banks, which are an essential topic for administrators, stockholders, and customers.

Svitalkova [1] shows that non-parametric methods are more acceptable than parametric ones for ranking decision-making units (DMUs). Based on Wanke et al. [2], DEA is a critical non-parametric method presently applied for efficiency and productivity evaluation. This method, technologically advanced by Charnes et al. [3], is founded by a scientific way of measuring efficiency. DEA classifies the most efficient DMUs and specifies what inefficient units must do to become efficient. To clarify more, DEA shows the best observe to be recognized from an efficiency frontier [4].

Over time, it has correspondingly progressed to develop more diverse with highlighting on product diversity through novel blends and valuable formulation developments instead of cooperating quality to live only as a low-cost common substitute.

The primary purpose of this paper is based on the evaluation of efficiency with cross-efficiency. We have applied the two following models to find the superior one:

- $CCR_{IO} - BCC_{IO}$
- $BCC_{IO} - CCR_{IO}$

The rest of the paper clarifies as follows:

Part 2 distributes a careful evaluation of literature related to the calculation of productivity and stipulates the potential role of the current study. Part 3 discusses the four steps of research methodology, and evaluation of each step (step1: CCR-BCC, BCC-CCR evaluations), step2: Inputs and outputs description, step3: Evaluation in cross-efficiency, step4: Process of efficiency with cross-efficiency, followed by a discussion and conclusion of the experimental consequences in Part 5 and 6 respectively. Finally, Figure 1 shows the whole process:

Published on June2, 2020.

M. Mirmozaffari (Corresponding Author), Azam Boskabadi, and Gohar Azeem, PhD Candidate, Department of Industrial Manufacturing and Systems Engineering, The University of Texas at Arlington, Arlington, TX, USA. (e-mail: mirpouya.mirmozaffari@mavs.uta.edu).

R. Massah, Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX, USA.

E. Boskabadi, PhD student, Department of Economics, College of Liberal Arts, Auburn University, AL, USA.

H. A. Dolatsara, Assistant Professor, Graduate School of Management, Clark University, Worcester, MA, USA.

A. Liravian, Department of Biomedical Engineering, The University of Texas at Arlington, Arlington, TX, USA.

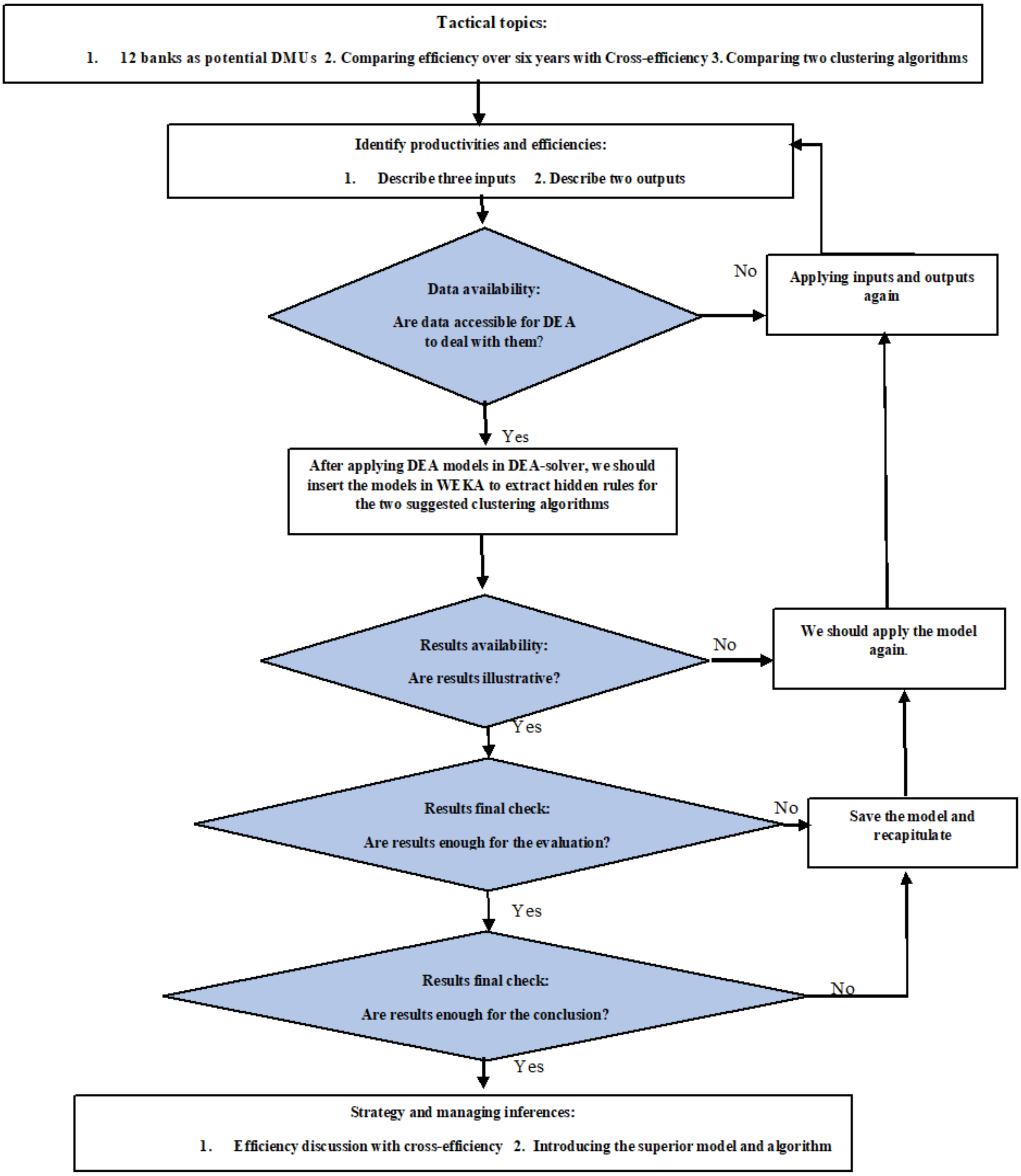

Fig. 1. Assessment procedure of machine learning and Optimization

II. Background and Literature Review

Banks are crucial foundations in a country's budget and economy. Based on Tsolas and Charles [5], the banking part plays an essential role in each country; consequently, difficulties in this area are the central part of the numerous papers. Based on the importance of economic institutes, many previous articles have pursued to assess the performance of banks in various countries [6-13]. Berger and Humphrey [14], in an outstanding study, surveyed 130 pieces of training that examined 21 multiple countries to evaluate bank efficiency base on parametric and non-parametric approaches, which shows the importance of education on efficiency evaluation in bank sectors.

Other related studies can be addressed through various computational and rating methods [15-24].

It is perceived that various DEA models are commonly utilized in different studies to compare, rank, and evaluate energy efficiency. Thus, a comprehensive comparison of several efficiencies delivers insight into the bank's efficiency. This comparison is of considerable significance to bank practitioners who desire to assess efficiency at a proper step of its progression. A unique comparing exclusive four models in cross-efficiency is applied, which eventually results in comparing several efficient and inefficient DMUs. Finally, finding the superior model provide valuable information for bank managers to select the best model. Meanwhile, comparing various bank's companies from different developing countries is one of the novelties of our research, which considers large laboratories at the same time. Thus, it can be beneficial for managers to have superior evaluating, remove unrelated data, and more effective processes.

III. Research Methodology

The objective of this study is to compare companies' efficiency effectively. Using a comparative DEA with cross efficiency is established to determine the features of banks in terms of some DMUs with two suggested models. Finally, the entire progression can be divided into four steps, as follows

A. CCR-BCC and BCC-CCR models:

1) CCR-BCC Model:

Consider manufacturing technology where if it produces X_0 and Y_0 then λX_0 can produce λY_0 only when we have $\lambda \leq 1$. We make a set of production possibilities that include observations and apply the principles of convexity and feasibility. This series will be introduced as follows.

$$T_{CCR-BBC} = T_{NI} = \{(X,Y)|X \geq \sum_{j=1}^{n} \lambda_j X_j \,\&\, Y \leq \sum_{j=1}^{n} \lambda_j Y_j \,\&\, \sum_{j=1}^{n} \lambda_j \leq 1 \,\&\, \lambda \geq 0\} \quad (1)$$

Suppose the purpose of evaluating the DMU with input X and output Y concerning the abovementioned technology will be the following definition:
T is defined as the set of possible production.

The main goal in the input-oriented method is to find a virtual unit in which the input θX_0 is not more than X_0, and the minimum production should be Y_0. In fact:

$$\begin{aligned} &Min\theta \\ &St. \\ &(\theta X_0, Y_0) \in T_{ND} \end{aligned} \quad (2)$$

Based on the T_{ID} structure for $CCR_{IO} - BCC_{IO}$:

$$\begin{aligned} &Min\theta \\ &St. \\ &\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_p \qquad , i = 1, \ldots\ldots, m \\ &\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rp} \qquad , r = 1, \ldots\ldots, s \\ &\sum_{j=1}^{n} \lambda_j \leq 1 \\ &\qquad \lambda_j \geq 0 \qquad , j = 1, \ldots, n \end{aligned} \quad (3)$$

2) BCC-CCR Model:

Consider manufacturing technology where if it produces X_0 and Y_0 then λX_0 can produce λY_0 only when we have $\lambda \geq 1$. We make a set of production possibilities that include observations and apply the principles of convexity and feasibility. This series will be introduced as follows.

$$T_{BCC-CCR} = T_{ND} = \{(X,Y)|X \geq \sum_{j=1}^{n} \lambda_j X_j \,\&\, Y \leq \sum_{j=1}^{n} \lambda_j Y_j \,\&\, \sum_{j=1}^{n} \lambda_j \geq 1 \,\&\, \lambda \geq 0\} \quad (4)$$

Suppose the purpose of evaluating the DMU with input X and output Y in relation to the abovementioned technology will be the following definition:

T is defined as the set of possible production

The goal in the input-oriented method is to find a virtual unit in which the input θX_0 is not more than X_0 and at least produce Y_0. In fact:

$$\begin{aligned} &Min\theta \\ &St. \\ &(\theta X_0, Y_0) \in T_{ND} \end{aligned} \quad (5)$$

Based on the T_{ND} structure for $BCC_{IO} - CCR_{IO}$:

$$\begin{aligned} &Min\theta \\ &St. \\ &\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_p \qquad , i = 1, \ldots\ldots, m \\ &\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rp} \qquad , r = 1, \ldots\ldots, s \\ &\sum_{j=1}^{n} \lambda_j \geq 1 \\ &\qquad \lambda_j \geq 0 \qquad , j = 1, \ldots, n \end{aligned} \quad (6)$$

B. Inputs and outputs Description:

As soon as the six suggested models, the orientation, and the DMUs were established, the next step was to create which variables would be involved in the model. This involvement is the most important step in applying DEA. After applying the DMU j (j=1. . . n) which are banking facilities in eight selected developing countries or decision-making units, we propose the following three inputs and two outputs in our study:

- Xij (i= 1. . . m): Fixed assets
- Ncj (C= 1. . . c): Personnel expenses
- Qoj (O= 1. . . o): Total deposits
- Yrj (r = 1. . . s): Total loans
- Mhj (H = 1. . . h): Total profits

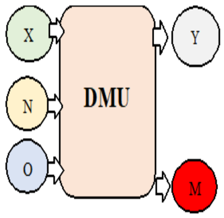

Fig. 2. Three inputs and two outputs for the two suggested models

Drake et al. [25] divide the selecting variables for financial institutes, into two following parts:

- Production: Based on Benston [26], banks are mainly measured to be service providers for customers. The inputs involve physical variables such as staff, capital, and materials. The outputs are generally related to the services available to customers, which may include deposits and loans
- Intermediation: Based on Sealey and Lindley [27], the critical role of banks is to gather assets and change them into investments and other profitable assets. The bank is chiefly playing an essential intermediary among extra managers and a lack of managers.

The production approach is more appropriate for assessing agencies, while the intermediation method is more suggested for bank evaluation. Several papers such as Svitalkova [1], Liu et al. [10] Zimkova [28], and Assaf et al.

[29] have used the same inputs in our study. However, these papers measured the number of employees instead of personnel expenses.

Concerning the outputs, many papers used the total loans as the output such as Drake et al. [25], Liu et al. [10], Assaf et al

[26] and Yilmaz and Güneş [30]. Numerous papers used the intermediation method also used total loan output, like in our study, based on the primary duty of banks, is to take deposits and to lend money. We consider the total profits in our research as a second output too.

The data analysis was directed applying R language and the Benchmarking package, which makes many DEA models available.

The linear and dual evaluation for all the above-mentioned models are presented below:

1) Linear model in $CCR_{IO} - BCC_{IO}$:

$$Max = \sum_{r=1}^{s} u_r y_{rp} + \sum_{h=1}^{H} e_h m_{hp}$$

$$\begin{aligned} St.\quad & \sum_{i=1}^{m} v_i x_{ip} + \sum_{c=1}^{C} f_c n_{cp} + \sum_{o=1}^{O} k_o q_{op} \le 1 \\ & \sum_{r=1}^{s} u_r y_{rj} + \sum_{h=1}^{H} e_h m_{hj} + \sum_{v=1}^{V} g_v d_{vj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{c=1}^{C} f_c n_{cj} - \sum_{o=1}^{O} k_o q_{oj} \le 0 \\ & u_r, e_h, v_i, f_c, k_j \ge 0 \;\; , j = 1, \dots, n \end{aligned} \tag{7}$$

2) Dual model in $CCR_{IO} - BCC_{IO}$:

$$\begin{aligned} & Min\ \theta \\ St.\quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_p x_{ip} \\ & \sum_{j=1}^{n} \lambda_j n_{cj} \le \theta_p n_{cp} \\ & \sum_{j=1}^{n} \lambda_j q_{oj} \le \theta_p q_{op} \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rp} \\ & \sum_{j=1}^{n} \lambda_j m_{hj} \ge m_{hp} \\ & \sum_{j=1}^{n} \lambda_j \le 1 \\ & \lambda_j \ge 0 \quad \theta_p\ free \end{aligned} \tag{8}$$

3) Linear model in $BCC_{IO} - CCR_{IO}$:

$$Max = \sum_{r=1}^{s} u_r y_{rp} + \sum_{h=1}^{H} e_h m_{hp}$$

$$\begin{aligned} St.\quad & \sum_{i=1}^{m} v_i x_{ip} + \sum_{c=1}^{C} f_c n_{cp} + \sum_{o=1}^{O} k_o q_{op} \ge 1 \\ & \sum_{r=1}^{s} u_r y_{rj} + \sum_{h=1}^{H} e_h m_{hj} + \sum_{v=1}^{V} g_v d_{vj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{c=1}^{C} f_c n_{cj} - \sum_{o=1}^{O} k_o q_{oj} + w \le 0 \\ & u_r, e_h, v_i, f_c, k_j \ge 0 \;\; , j = 1, \dots, n \end{aligned} \tag{9}$$

4) Dual model in $BCC_{IO} - CCR_{IO}$

$$\begin{aligned} & Min\ \theta \\ St.\quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_p x_{ip} \\ & \sum_{j=1}^{n} \lambda_j n_{cj} \le \theta_p n_{cp} \\ & \sum_{j=1}^{n} \lambda_j q_{oj} \le \theta_p q_{op} \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rp} \\ & \sum_{j=1}^{n} \lambda_j m_{hj} \ge m_{hp} \\ & \sum_{j=1}^{n} \lambda_j \ge 1 \\ & \lambda_j \ge 0 \quad \theta_p\ free \end{aligned} \tag{10}$$

C. Definition of cross-efficiency:

Suppose there are N DMUs to be evaluated, indexed by j = 1,…,N, and each DMU is assumed to produce s different outputs from m different inputs, denoted by two following relations:

$$x_j = (x_{1j} \dots x_{mj}) \tag{11}$$

$$y_j = (y_{1j} \dots y_{sj}) \tag{12}$$

All the components of the vectors x_j and $y_j$for all the DMUs are assumed to be non-negative. Each DMU is assumed to have at least one strictly positive input and one strictly positive output as follows:

$$x_{io} = (i = 1,2, \dots, m) \tag{13}$$

$$y_{ro} = (r = 1,2, \dots, s) \tag{14}$$

Based on the inputs as mentioned above and outputs the relative efficiency score is defined as bellow:

$$\theta_0(u_r, v_i) = \max \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{15}$$

where $u_r$and v_i are the non-negative vectors that represent the output and input weights, respectively. Additionally, we require that the same weights, when applied to all the DMUs, do not provide any unit with an efficiency greater than one. This condition appears in the following set of constraints:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1 \qquad j = 1,2, \dots . N \tag{16}$$

Cross-efficiency matrix for N DMUs and the average cross-efficiency is available in Table I:

TABLE I: CROSS-EFFICIENCY MATRIX

DMU	Target DMU 1 2 … n	Average cross-efficiency
1	θ_{11} θ_{12} … θ_{1N}	$1/N\sum\theta_{1K}$
2	θ_{21} θ_{22} … θ_{2N}	$1/N\sum\theta_{1K}$
.	. . . .	.
.	. . . .	.
.	. . . .	.
n	θ_{N1} θ_{N2} … θ_{NN}	$1/N\sum\theta_{NK}$

Therefore, the fractional model for obtaining the relative efficiency of DMU_O is as follows:

$$\theta_0 = \max\frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$St.$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad j = 1,2,\dots.N \tag{17}$$

$$v_i \geq £ \quad u_r \geq £ \qquad i = 1,2,\dots,m;\ r = 1,2,\dots,s$$

where £ > 0 is a non-Archimedean construct to ensure strongly efficient solutions. In order to ensure the above model, have finite optimal values, Ali, and Seiford [31] and Mehrabian et al. [32] argued that £ should satisfy the following inequality.

$$£ < 1/max\{\textstyle\sum_{i=1}^{m} v_i\, x_{io}\} \tag{18}$$

In the above model, each unit is evaluated by its best weights. DMU0 is efficient if it lies on the Pareto frontier. Model (3) is a fractional programming problem and can be converted to the following linear programming:

$$Max \textstyle\sum_{r=1}^{s} u_r\, y_{ro}$$

$$St. \tag{19}$$

$$\textstyle\sum_{i=1}^{m} v_i\, x_{io} = 1$$

$$\textstyle\sum_{r=1}^{s} u_r\, y_{ro} - \sum_{i=1}^{m} v_i\, x_{io} \leq 0 \qquad j = 1,2,\dots.N$$

$$\lambda_j \geq 0 \qquad ,j = 1,\dots,n$$

$$v_i \geq £ \quad u_r \geq £ \qquad i = 1,2,\dots,m;\ r = 1,2,\dots,s$$

where £ > 0 is a non-Archimedean construct.

Let $u_r(r = 1,\dots,s)$ and $v_i(i = 1,\dots,m)$ be the optimal output and input weights to the above model, respectively.

$\theta_0 = \sum_{r=1}^{s} u_r\, y_{ro}$is referred to as the CCR efficiency of DMU_0, which is the best relative efficiency that DMU_0can achieve and reflects the self-evaluated efficiency of the DMU_0. $DMU_0$is CCR is efficient if $\theta_0 = 1$; otherwise, $DMU_0$is CCR-inefficient.

The cross-efficiency value of $DMU_j$is defined as $\theta_0 = \max\frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$, which reflects the peer evaluation of $DMU_0$to DMU_j (j = 1,...,N; j ≠0). Consequently, we obtain an N × N matrix in which the diagonal members show the CCR-efficiency scores. We can compute the average of the cross-efficiency scores (the common method) in each row to proceed with a cross-evaluation. See Table 1 for details.

Because alternative input and output weights exist, Sexton et al. [33] suggested introducing a secondary goal to compute the unique efficiency score for DMUs. Doyle and Green [34] proposed aggressive and benevolent models, which minimize or maximize, respectively, the efficiency of the composite DMU constructed for other DMUs as compared to DMU_0. The most commonly used secondary goal is the aggressive model, which minimizes the efficiency of the composite DMU constructed for n 1 DMUs. The aggressive model is given as:

$$Min \sum_{r=1}^{s} u_r \,(\sum_{\substack{j=1\\ j\neq 0}}^{n} y_{rj})$$

$$St. \tag{20}$$

$$\textstyle\sum_{i=1}^{m} v_i\, (\sum_{\substack{j=1\\ j\neq 0}}^{n} x_{ij}) = 1$$

$$\textstyle\sum_{r=1}^{s} u_r\, y_{ro} - \theta_0 \sum_{i=1}^{m} v_i\, x_{io} = 0$$

$$\textstyle\sum_{r=1}^{s} u_r\, y_{rj} - \sum_{i=1}^{m} v_i\, x_{ij}$$

$$\lambda_j \geq 0 \qquad ,j = 1,\dots,n \qquad j \neq 0$$

$$v_i \;\; u_r \geq £ \qquad i = 1,2,\dots,m;\ r = 1,2,\dots,s$$

where £ > 0 is a non-Archimedean construct, and θ_0 is the CCR efficiency of DMU_0derived from the CCR model. The benevolent formulation for cross-efficiency evaluation is achieved by putting max in the objective function of Model (19) in place of min. See Model (20) for details.

$$Max \sum_{r=1}^{s} u_r \,(\sum_{\substack{j=1\\ j\neq 0}}^{n} y_{rj})$$

$$St. \tag{21}$$

$$\textstyle\sum_{i=1}^{m} v_i\, (\sum_{\substack{j=1\\ j\neq 0}}^{n} x_{ij}) = 1$$

$$\textstyle\sum_{r=1}^{s} u_r\, y_{ro} - \theta_0 \sum_{i=1}^{m} v_i\, x_{io} = 0$$

$$\textstyle\sum_{r=1}^{s} u_r\, y_{rj} - \sum_{i=1}^{m} v_i\, x_{ij}$$

$$\lambda_j \geq 0 \qquad ,j = 1,\dots,n \qquad j \neq 0$$

$$v_i \;\; u_r \geq £ \qquad i = 1,2,\dots,m;\ r = 1,2,\dots,s$$

If a cross-efficiency matrix is given, it must be decided how to aggregate cross-efficiencies to one score for each DMU in the next step. Other associated studies can be addressed through several similar methods [35-45].

D. Evaluation in clustering:

Clustering is a foremost duty of explorative data mining, and a public procedure for numerical data analysis utilized in several areas, containing machine learning, pattern recognition, and bioinformatics.

In this study, 70 percent of the data were designated as training data sets, and 30 percent of the data were selected as experimental data sets. To randomly select the experimental data, the Excel software has been used. Finally, to compare and to find the superior algorithms, eight designated clustering algorithms in WEKA software are widely discussed below [46]:

1) Farthest First

This algorithm is a Modified of K-means that seats each cluster center in sequence at the point farthest from the exiting cluster center lying inside the data range. It is appropriate for large-scale data sets. Farthest first is a heuristic created process of clustering. It also selects centroid and allocates the items in clusters. This algorithm provides fast clustering in most of the cases since less relocation and modification is required.

2) Expectation Maximization (EM)

This algorithm is an iterative technique for the detection of maximum possibility in statistical models, and undetected hidden variables determine it. It provides a valuable result for the actual world data set. The EM iteration substitutes among acting an expectation (E) step, which produces a purpose intended for the expectation of the log-possibility assessed utilizing the existing evaluation for the factors, and a maximization (M) step, which calculates factors maximizing the expected log-possibility establish on the E step. These factor-estimates are then utilized to conclude the delivery of the hidden variables in the next E step.

IV. Discussion

A. Discussion in cross-efficiency in CCR-BCC model for 2014, CCR-BCC OVER 2014-2019 and BCC-CCR over 2014-2019:

Just as an example, we consider the first period of the CCR- BCC model as a cross-efficiency evaluation. As we discussed above, the diagonal members show the CCR-BCC efficiency scores. Table II represents a cross-efficiency ranking in the first period (2014) for 12 banks.

It can be concluded from Table II:

- DMU_{10} has the first and the highest average efficiency score in the CCR-BCC model for the first period
- DMU_6 has the second average efficiency score in the CCR-BCC model for the first period
- DMU_3 has the third average efficiency score in the CCR-BCC model for the first period

The data covers in this study are six years from 2014 to 2019 for 12 banks in the two developing countries. The number of DMUs is N or 12, and the period is T or 6.

It can be concluded from Table II:

- DMU_{10} has the first and the highest average efficiency score in the CCR-BCC model for the first period
- DMU_6 has the second average efficiency score in the CCR-BCC model for the first period
- DMU_3 has the third average efficiency score in the CCR-BCC model for the first period

TABLE II: Cross-efficiency ranking over 2014 for 12 banks in CCR-BCC

Efficiency score	**100**	**99.95**	**99.17**	**91.90**	**45.88**	**99.96**	**57.62**	**34.13**	**99.97**	**62.31**	**73.33**	**87.08**
Avg.	93.63	75.06	45.97	64.05	34.26	75.61	36.76	25.02	88.75	49.41	36.67	48.94
Ranking	1	4	8	5	11	3	9	12	2	6	10	7
DMUs	10	11	12	1	2	3	4	5	6	7	8	9
10	**100**	0.27	0.24	3.74	0.9	3.13	1.22	0.19	1.88	1.88	0.19	1.06
11	27.13	**99.95**	18.72	30.41	16.02	34.28	40.37	21.42	100	21.96	23.32	38.72
12	100	100	**99.17**	44.86	34.92	90.25	6.62	29.83	63.27	54.68	65.80	12.85
1	100	89.21	40.13	**91.90**	45.90	75.12	57.64	30.17	100	57.88	29.88	87.1
2	100	89.21	40.13	91.92	**45.88**	75.12	57.64	30.17	100	57.88	29.88	87.1
3	100	25.2	25.98	64.42	29.82	**99.96**	35.39	14.32	100	54.33	23.89	32.38
4	100	89.21	40.13	91.92	45.90	75.12	**57.62**	30.17	100	57.88	29.88	87.1
5	100	100	76.49	68.37	41.75	100	30.04	**34.13**	100	62.33	61.13	44.35
6	100	18.53	13.28	80.36	31.31	79.34	57.28	11.82	**99.97**	50.78	11.79	51.48
7	100	100	76.49	68.37	41.75	100	30.04	34.15	100	**62.31**	61.13	44.35
8	72.44	100	80.77	40.59	31.02	100	9.56	33.75	100	53.08	**73.33**	13.7
9	100	89.21	40.13	91.92	45.90	75.12	57.64	30.17	100	57.88	29.88	**87.08**

The average cross-efficiency-CCR-BCC for all banks over 2014-2019 is given in Table III.

TABLE III: Comparing Cross-efficiency-CCR-BCC model for all banks over 2014-2019

Banks	Efficiency Score **(2014)**	Ranking **(2014)**	Efficiency Score **(2015)**	Ranking **(2015)**	Efficiency Score **(2016)**	Ranking **(2016)**	Efficiency Score **(2017)**	Ranking **(2017)**	Efficiency Score **(2018)**	Ranking **(2018)**	Efficiency Score **(2019)**	Ranking **(2019)**	**Avg.**
1	91.90%	5	61.63%	10	27.24%	12	64.58%	8	44.89%	10	87.96%	7	62.93%
2	45.88%	11	50.47%	12	35.55%	11	45.11%	11	23.31%	12	39.48%	10	39.96%
3	99.96%	3	59.69%	11	45.81%	9	99.96%	3	55.09%	8	99.95%	4	76.74%
4	57.62%	10	99.95%	4	99.94%	5	59.92%	10	64.93%	7	32.47%	11	69.14%
5	34.13%	12	85.95%	7	99.95%	4	33.76%	12	50.90%	9	49.76%	9	59.08%
6	99.97%	2	81.19%	9	99.96%	3	99.97%	2	99.97%	2	99.94%	5	96.82%
7	62.31%	9	99.97%	2	71.80%	6	62.39%	9	99.96%	3	99.96%	3	82.73%
8	73.33%	8	99.96%	3	49.22%	7	72.65%	7	40.62%	11	99.97%	2	72.62%
9	87.08%	7	85.34%	8	49.16%	8	85.42%	6	99.95%	4	27.57%	12	72.41%
10	100%	1	99.94%	5	45.30%	10	100%	1	99.94%	5	100%	1	90.84%
11	99.95%	4	100%	1	100%	1	99.95%	4	100%	1	75.72%	8	95.92%
12	99.17%	6	96.47%	6	99.97%	2	99.94%	5	78.14%	6	99.94%	6	95.60%
Avg.	79.27%		85.05%		68.66%		76.97%		71.43%		76.04%		

The average cross-efficiency-BCC-CCR for all banks over 2014-2019 is given in Table IV.

TABLE IV: COMPARING CROSS-EFFICIENCY-BCC-CCR MODEL FOR ALL BANKS OVER 2014-2019

Banks	Efficiency Score **(2014)**	Ranking **(2014)**	Efficiency Score **(2015)**	Ranking **(2015)**	Efficiency Score **(2016)**	Ranking **(2016)**	Efficiency Score **(2017)**	Ranking **(2017)**	Efficiency Score **(2018)**	Ranking **(2018)**	Efficiency Score **(2019)**	Ranking **(2019)**	**Avg.**
1	91.91%	5	61.64%	10	27.25%	12	64.59%	8	44.9%	10	87.97%	7	62.94%
2	45.89%	11	50.48%	12	35.56%	11	45.12%	11	23.32%	12	39.49%	10	39.97%
3	99.97%	3	59.70%	11	45.82%	9	99.97%	3	55.10%	8	99.96%	4	76.75%
4	57.63%	10	99.96%	4	99.95%	5	59.93%	10	64.94%	7	32.48%	11	69.15%
5	34.14%	12	85.96%	7	99.96%	4	33.77%	12	50.91%	9	49.77%	9	59.09%
6	99.98%	2	81.20%	9	99.97%	3	99.98%	2	99.98%	2	99.95%	5	96.83%
7	62.32%	9	99.98%	2	71.81%	6	62.40%	9	99.97%	3	99.97%	3	82.74%
8	73.34%	8	99.97%	3	49.23%	7	72.66%	7	40.63%	11	99.98%	2	72.63%
9	87.09%	7	85.35%	8	49.17%	8	85.43%	6	99.96%	4	27.58%	12	72.42%
10	100%	1	99.95%	5	45.31%	10	100%	1	99.95%	5	100%	1	90.85%
11	99.96%	4	100%	1	100%	1	99.96%	4	100%	1	75.73%	8	95.93%
12	99.18%	6	96.48%	6	99.98%	2	99.95%	5	78.15%	6	99.94%	6	95.61%
Avg.	79.28%		85.06%		68.67%		76.98%		71.44%		76.05%		

It can be concluded from Table III, and IV:

- BCC-CCR and CCR-BCC models have the same ranking for all DMUs
- BCC-CCR model has the first average efficiency score over 5-years period for 12 DMUs
- CCR-BCC model has the second average efficiency score over 5-years period for 12 DMUs

According to the evaluation of cross-efficiency in two suggested models in Table III and IV for 12 banks with quantified input and output principles:

Based on the **CCR-BCC** model in Table III:

- The 6th bank has the 1st or the highest average efficiency score of 96.82.
- The 11th and 12th banks are in the 2nd and 3rd places with efficiency scores of 95.92 and 95.60, respectively.
- The 2nd bank has the 12th and the lowest efficiency score of 39.96.
- The 7th and 9th banks are in the 11th and 12th places with efficiency scores of 62.93 and 59.08, respectively.

Based on the **BCC-CCR** model in Table IV:

- The 6th bank has the 1st or the highest average efficiency score of 96.83.
- The 11th and 12th banks are in the 2nd and 3rd places with efficiency scores of 95.93 and 95.61, respectively.
- The 2nd bank has the 12th and the lowest efficiency score of 39.97.
- The 22nd and 30th banks are in the 28th and 29th places with productivity scores of 62.94 and 59.09, respectively.

BCC-CCR and CCR-BCC models are in the 1st, and 2nd places, respectively. Finally, the following relation is applicable for all DMUs in all cross-efficiency and all years:

$$\text{BCC-CCR} > \text{CCR-BCC} \tag{21}$$

B. Discussion in the clustering

After applying DEA models in the DEA-SOLVER at the first step, we apply all the efficient and inefficient DMUs in WEKA software. Finally, the accuracy and average accuracy are presented in Table V.

TABLE V: ACCURACY COMPARISON CONTAINED BY CLUSTERING ALGORITHMS (ALL NUMBERS ARE IN PERCENT)

Algorithms	**CCR-BCC**	**BCC-CCR**
Farther First	74.8182	77.5455
Expectation Maximization	65.4831	69.6431
Average	70.1506	73.5943

It can be concluded from Table V, from CCR-BCC to BCC-CCR model:

- The maximum of accuracy within two assessment approaches is improved.
- The average accuracy within two algorithms, is augmented.
- The accuracy of two algorithms is increased.

And finally, after applying BCC-CCR and CCR-BCC models at the first step, in the extraction of hidden rules of the second clustering step, BCC-CCR was the superior model.

V. CONCLUSION

In this study, we describe how banks operate in the presence of similar banks. Therefore, those banks which have a higher score can improve their efficiency. The more taking available information, the higher accurate and accessible data will be available. Each bank needs a productivity measurement to know its current status. So, efficient banks are the best reference for increasing the efficiency of inefficient banks. The BCC-CCR model has a more positive impacts on efficiency and accuracy scores in DEA optimization approach and machine learning clustering method compare with CCR-BCC model. The proposed approach, geometric average, results, and predictions derived from the period and efficiencies in cross-efficiencies can help the practitioner to compare the efficacy of uncertain cases and instruct accordingly. In the future, applying window analysis and Malmquist Productivity Index (MPI) and comparing final productivities and efficiencies result with cross-efficiency will be valuable. Since the window analysis method is based on a moving average, it is useful for finding per efficiency trends over time. Meanwhile, using fuzzy and random data for cross-efficiency will be interesting as a final comparison. So, the results and predictions can be helpful for managers of these banks and other managers who benefit from this approach to achieve a higher relative efficiency score. Besides, managers can compare the efficiency of the current year with other similar companies over the past years.

REFERENCES

[1] Z Svitalkova, Comparison and evaluation of bank efficiency in selected countries in EU. *Procedia Economics and Finance*, 12(1), 644–653, 2014.

[2] P Wanke, M. A. K Azad and C Barros. Predicting efficiency in Malaysian Islamic banks: A two-stage TOPSIS and neural networks approach. *Research in International Business and Finance*, 36(January (1)), 485–498, 2016.

[3] A Charnes, W Cooper, E Rhodes, Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(November (6)), 429–444, 1978.

[4] A Charnes, W. W., Cooper, A. Y Lewin and L.M. Seiford, Data Envelopment Analysis: Theory, methodology, and applications. Springer, 1994.

[5] I. E. Tsolas, and V Charles, incorporating risk into bank efficiency: A satisficing DEA approach to assess the Greek banking crisis. *Expert Systems with Applications*, 42(May (7)), 3491–3500, 2015.

[6] P Schure, R Wagenvoort, and D OB¨rien, The efficiency and the conduct of Europe on banks: Developments after 1992. *Review of Financial Economics*, 13(January (4)), 371–396, 2004.

[7] I. Řepková. Efficiency of the Czech banking sector employing the DEA window analysis approach. *Procedia Economics and Finance*, 12(1), 587–596, 2014.

[8] T. T. Lin, C.C Lee, and T.F. Chiu. Application of DEA in analyzing a bank operating performance. *Expert Systems with Applications*, 36(July (5)), 8883–8891, 2009.

[9] Y. Luo, G. Bi, L. Liang, Input/output indicator selection for DEA efficiency evaluation: An empirical study of Chinese commercial banks. *Expert Systems with Applications*, 39 (January (1)), 1118–1123, 2012.

[10] S.T. Liu Measuring and categorizing technical efficiency and productivity change of commercial banks in Taiwan. *Expert Systems with Applications*, 37(April (4)), 2783–2789, 2010.

[11] A. Sokic, Cost efficiency of the banking industry and unilateral euroization: A stochastic frontier approach in Serbia and Montenegro. *Economic Systems*, 39(September (3)), 541–551, 2015.

[12] F. Kamarudin, F Sufian, and A.M Nassir, Does country governance foster revenue efficiency of Islamic and conventional banks in GCC countries? *EuroMed Journal of Business*, 11(2), 181–211, 2016.

[13] F. Kamarudin, F Sufian, F.W. Loong, and N. A. M. Anwar, Assessing the Domestic and Foreign Islamic Banks Efficiency: Insights from Selected Southeast Asian countries. Future Business Journal, 3(1), 33–46, 2017.

[14] A. N. Berger, and D. B. Humphrey. Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, 1, 1, 1997.

[15] M. Mirmozaffari, "Eco-Efficiency Evaluation in Two-Stage Network Structure: Case Study: Cement Companies," *Iranian Journal of Optimization* (IJO), Dec. 16, 2018.

[16] M. Mirmozaffari, and A. Alinezhad, "Ranking of Heart Hospitals Using cross-efficiency and two-stage DEA," *7th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, pp. 217-222, 2017.

[17] A. Alinezhad, and M. Mirmozaffari, "Malmquist Productivity Index Using Two-stage DEA Model in Heart Hospital, *Iranian Journal of Optimization*. Volume 10, Issue 2, 2018.

[18] M. Mirmozaffari, "Presenting an expert system for early diagnosis of gastrointestinal diseases," *International Journal of Gastroenterology Sciences*, Vol 1; Issue 1; Page 21-27, 2020.

[19] M. Mirmozaffari, "Developing an Expert System for Diagnosing Liver Diseases," *EJERS*, vol. 4, no. 3, pp. 1-5, Mar. 2019.

[20] M. Mirmozaffari, "Presenting a Medical Expert System for Diagnosis and Treatment of Nephrolithiasis," *EJMED*. May; 1:1, 2019.

[21] A. Aranizadeh, M. Kazemi, H. Berahmandpour, and M. Mirmozaffari, "MULTIMOORA Decision Making Algorithm for Expansion of HVDC and EHVAC in Developing Countries (A Case Study)," *Iranian Journal of Optimization*, 2020.

[22] A. Aranizadeh, I. Niazazari, and M. Mirmozaffari, "A Novel Optimal Distributed Generation Planning in Distribution Network using Cuckoo Optimization Algorithm," *European Journal of Electrical Engineering and Computer Science* 3 (3), 2019.

[23] A. Azadeh, A. Boskabadi, and S. Pashapour. "A unique support vector regression for improved modelling and forecasting of short-term gasoline consumption in railway systems," *International Journal of Services and Operations Management*, *21*(2), 217-237, 2015.

[24] A. Boskabadi. "Using support vector regression (SVR) for weekly oil consumption prediction in railway transportation industry," no. December 1-12, 2011.

[25] L. Drake, M.J. Hall, and R. Simper, Bank modelling methodologies: A comparative nonparametric analysis of efficiency in the Japanese banking sector. *Journal of International Financial Markets, Institutions and Money*, 19(February (1)), 1–15, 2009.

[26] G.J. Benston, Branch banking and economies of scale. *The Journal of Finance*, 20(May (2)), 312–331, 1965.

[27] C. W. Sealey, and J. T. Lindley. Inputs, outputs and a theory of production and cost at depository financial institutions. *The Journal of Finance*, 32(September (4)), 1251–1266, 1977.

[28] E. Zimková. Technical efficiency and super-efficiency of the banking sector in Slovakia. Procedia Economics and Finance, 12(1), 780–787,2014.

[29] G.A Assaf, C.P. Barros, C. P., and R Matousek, Technical efficiency in Saudi banks. *Expert Systems with Applications*, 38 (May (5)), 5781–5786, 2011.

[30] A. Yilmaz, and N. Güneş. Efficiency comparison of participation and conventional banking sectors in Turkey between 2007–2013. *Procedia - Social and Behavioral Sciences*, 195(July (1)), 383–392, 2015.

[31] A.I. Ali, and L.M. Seiford, Computational accuracy and infinitesimals in data envelopment analysis. *INFOR* 31 (4), 290–297, 1993.

[32] S. Mehrabian, G.R. Jahanshahloo, M.R. Alirezaee, G.R. Amin. An assurance interval for the non-Archimedean epsilon in DEA models. *Operations Research* 48 (2), :344–347, 2000.

[33] T.R. Sexton, R.H. Silkman, A.J. and Hogan. Data envelopment analysis: critique and extensions. In: Silkman, R.H. (Ed.), Measuring Efficiency: An Assessment of Data Envelopment Analysis. Jossey-Bass, Francisco, CA,1986.

[34] J. Doyle, and R. Green. Efficiency and cross-efficiency in DEA: derivations, meanings and uses. Journal of the Operations Research Society 45 (5), 567–578,1994.

[35] A. Boskabadi and A. Azadeh "A fuzzy model for a distribution network problem in a multi-product supply chain system," *5th national & 3rd international LOGESTICS & SUPPLY CHAIN CONFERENCE*, 75-85, 2012.

[36] H. Kamalzadeh, S.N. Sobhan, A. Boskabadi, M. Hatami and A. Gharehyakheh, "Modeling and Prediction of Iran's Steel Consumption Based on Economic Activity Using Support Vector Machines," *arXiv preprint arXiv:1912.02373*, 2019.

[37] S.S..Fazeli, S. Venkatachalam, R.B. Chinnam, and A. "Murat. Two-Stage Stochastic Choice Modeling Approach for Electric Vehicle Charging Station Network Design in Urban Communities," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[38] M. Mirmozaffari, G. Azeem, A. Boskabadi, A. Aranizadeh, A. Vaishnav, and J. John, "A Novel Improved Data Envelopment Analysis Model Based on SBM and FDH Models", *EJECE*, vol. 4, no. 3, May 2020.

[39] S. Amin-Nejad, T.A. Gashteroodkhani, and Basharkhah, K., "A Comparison of MVDR and LCMV Beamformers' Floating Point Implementations on FPGAs," *Wireless Personal Communications*, vol. 98, no. 2, pp.1913-1929, 2018.

[40] S. Amin-Nejad, K. Basharkhah, and T.A. Gashteroodkhani "Floating Point versus Fixed point Tradeoffs in FPGA Implementations of QR Decomposition Algorithm," European Journal of Electrical and Computer Engineering, vol. 3, no. 5, 2019.

[41] O.A. Gashteroodkhani, M. Majidi, M. Etezadi-Amoli "A Fuzzy-based Control Scheme for Recapturing Waste Energy in Water Pressure Reducing Valves" *IEEE Power and Energy Society General Meeting* (PESGM), pp. 1-5, Portland, OR, Aug 2018.

[42] M. Mirmozaffari, A. Alinezhad, and A. Gilanpour, Data Mining Apriori Algorithm for Heart Disease Prediction. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), pp.20-23, 2017.

[43] M. Mirmozaffari, A. Alinezhad, and A. Gilanpour, Data Mining Classification Algorithms for Heart Disease Prediction. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), pp.11-15, 2017.

[44] M. Mirmozaffari, A. Alinezhad, and A. Gilanpour, Heart disease prediction with data mining clustering algorithms. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), pp.16-19, 2017.

[45] M. Mirmozaffari, and A. Alinezhad, Window analysis using two-stage DEA in heart hospitals, October 2017.

[46] N. Sharma, A. Bajpai, and R. Litoriya, Comparison the various clustering algorithms of WEKA tools 2, 73–80, 2012.