Performance Evaluation of Industrial Firms Using DEA and DECORATE Ensemble Method

Hassan Najadat[1], Ibrahim Al-Daher[2], and Khaled Alkhatib[1]
[1]Computer Information Systems Department, Jordan University of Science and Technology, Jordan
[2]Computer Science Department, Jordan University of Science and Technology, Jordan

Abstract: *This study introduces an approach of combining Data Envelopment Analysis (DEA) and ensemble Methods in order to classify and predict the efficiency of Decision Making Units (DMU). The approach includes applying DEA in the first stage to compute the efficiency score for each DMU, then a variables' ranker was utilized to extract the most important variables that affect the DMU's performance, then J48 was adopted to build a classifier whose outcomes will be enhanced by Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE) Ensemble method. To examine the approach, this study utilizes a dataset from firms' financial statements that are listed on Amman Stock Exchange. The dataset was preprocessed and turned out to include 53 industrial firms for the years 2012 to 2015.The dataset includes 11 input variables and 11 output ratios. The examination of financial variables and ratios play a vital role in the financial analysis practice. This paper shows that financial variable and ratio averages are points of reference to evaluate and measure firms' future financial performance as well as that of other similar firms in the same sector. In addition, the results of this work are for comparative analyses of the financial performance of the industrial sector.*

Keywords: *Data Envelopment Analysis, Decision Trees, Ensemble Methods, Financial Variables, Financial Ratios.*

Received October 12, 2019; accepted April 8, 2020
https://doi.org/10.34028/iajit/17/5/8

1. Introduction

It is widely believed in the field of finance that future financial performance of firms can be determined, or at least forecasted, relying on their performance in the previous years. The examination of financial variables and ratios, which are normally collected from firms' financial statements and reports play a vital role in the financial analysis practice. Financial variables and ratios collected over different years can be useful as they provide information on the firms' sales, profitability, cash, and liquidity. More specifically, financial variable and ratio averages represent a point of reference to evaluate and measure firms' future financial performance as well as that of other similar companies in the same sector. Moreover, financial variables and ratios are used to predict the success or failure of a particular firm. In addition, they are used for comparative analyses of the financial performance of different sectors such as the industrial sector.

Despite the clear importance of financial predictions, this process is not an easy task, and it has formed a challenge for many researchers and practitioners. One reason for that is the variation in the empirical results. The variation in predicting results of firms' performance is attributed to the use of different financial variables and Ratios. Even using a single financial variable or ratio to predict the firms' performance, it may yield some variation due to the different potential directions and capabilities of the predicting process.

The prediction of firms' performance requires processing datasets that contain a group of firms as well as their financial variables and ratios over a particular number of previous years. This necessitates the use of analytical/statistical computer algorithms, softwares and techniques. One of these techniques is the Data Envelopment Analysis (DEA), which is a mathematical programing model that has several applications in various domains. For example, it has been employed in IT risk assessment [15]. DEA is also widely used in measuring the performance of Decision Making Units (DMUs) especially when the comparisons become difficult because of the presence of multiple inputs and outputs for the organizational units during the measuring relative performance task for these units [12]. It is also used for measuring the efficiency based on multiple inputs and multiple outputs for different DMUs. These DMUs could be departments, firms, institutions or any comparable units in any sector (e.g., hospitals, universities, banks and firms). The simple efficiency equation represents the ratio of the outputs of the desirable DMU to the input for that DMU, as illustrated in Equation (1).

$$Efficiency = \frac{Outputs}{Inputs} \tag{1}$$

Data mining techniques are also used in analyzing datasets in the field of finance. The ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions. Ensemble methods are also widely used for improving the performance of classification and prediction models [9, 16] such as Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE).

The process of predicting the performance of firms has received much research in the last few decades. For example, many researchers used the DEA and financial ratios for predicting the firms' performance [8, 17]. Other researchers used the data mining techniques and financial ratios for predicting the firms' performance, such as [7] who employed decision tree analysis to assess the Turkish firms' performance in the Istanbul Stock Exchange. In another study, Zibanezhad and Foroghi [22] developed a model for predicting bankruptcy of the firms in Tehran Stock Exchange during the period 1996 to 2009, depending on financial ratios extracted from financial statements. Similarly, Chen and Du applied neural networks and data mining techniques on financial ratios, non financial ratios and factor analysis in the financial statement to extract adaptable variables with the ultimate goal of improving the accuracy of the financial distress prediction [6]. Moreover, [17] proposed an ensemble (bagging and boosting methods) with neural network for improving the performance of traditional neural networks on bankruptcy prediction tasks.

Aksoy and Yildiz [1] evaluated the performance of Turkish firms listed in corporate governance index using some financial variables and corporate governance. The firms' performance was evaluated by conducting the DEA. Yang and Dimitrov utilized Support Vector Machines (SVMs) and DEA to provide a new method for predicting corporate failure of nonmanufacturing firms; they revealed that using only DEA scores provides better predictions of corporate failure than using the original raw data for the provided dataset [20].

Due to the importance of measuring firms' performance in general, a study is needed to examine the performance of Jordanian firms in the industrial sector. This study provides an indication of the firms' performance for decision makers. This study presents a Hybrid Approach of Data Envelopment Analysis and Ensemble Methods (HDEA-EM) which apply the DEA for extracting the efficiency of the Jordanian firms in the industrial sector. Then the DECORATE ensemble method is employed to build a predictive model for classifying new Jordanian firms across the industrial sector either as an efficient or inefficient firm. This predictive model will be also very useful for predicting the future performance of the firm itself. In the process of predicting firms' performance, inputs and outputs should be determined to make comparisons upon which the efficiency of these firms is identified and evaluated. The present study explains how firms can improve their performance based on the best-identified variables and ratios to produce the most efficiency.

The rest of this paper is organized as follows: section 2 presents the related work, section 3 shows the study methodology, section 4 presents the results and discussion of the study, section 5 provides the conclusion, while the future work is in the last section.

2. Related Work

Evaluating firm's performance is a central factor in business because it provides managers and stock investors with the necessary information about the financial conditions and managerial performance of the firms. Accordingly, this issue has been tackled using different theoretical and practical frameworks. For example, many researchers used the DEA and financial ratios for evaluating the firms' performance. Other researchers used the data mining techniques and financial ratios for predicting the firms' performance. This section highlights the major findings of some relevant studies that employ the DEA and the data mining technique.

For example, Doumpos and Cohen [10] introduce a new modeling approach for evaluating the efficiency of Greek local governments. They used accounting data, which is publically available from financial statements. In order to obtain efficiency estimates, they employed the DEA. They implemented this approach to a panel dataset of Greek municipalities over the period 2002-2009. The data analysis has revealed that state subsidies play an adverse role on the efficiency of municipalities. They also explored the optimal real location of the municipalities' inputs and outputs with the goal of providing the central government with policy recommendations that can be implemented in the context of budget reduction.

Similarly, Aksoy and Yildiz [1] employed the DEA, which is a Multicriteria Decision Making (MCDM) method, in their study to assess the performance of Turkish firms that are listed in the corporate governance index depending on some financial variables and corporate governance. The study covered 31 nonfinancial firms listed in Borsa Istanbul (BIST). The firms' financial statements in 2015 formed the source of the data of the study. The analysis was divided into two steps. Firstly, total assets and total equity were used as inputs, and sales and profit were used as outputs to build the DEA model. Secondly, the same DEA model was repeated by adding corporate governance rating as a new input only. The analysis revealed that the firms' efficiency increased significantly in almost every aspect, which shows that using corporate governance rating is required to get a proper firm performance and financial values.

Likewise, in their examination of the changes that affected the productive efficiency of Indian commercial banks in 1992 (after financial sector reforms), Paul and Das employed DEA [18]. The efficiency of a firm is measured in terms of its relative performance: the efficiency of a firm relative to the efficiencies of firms in a sample. The aim of implementing DEA was to specify the banks that are on the output frontier in light of the different inputs at their disposal. Indian Commercial banks were categorized into three groups:

a. Public sector banks.
b. Private commercial banks.
c. Foreign banks.

The changes of efficiency in these three groups of banks were compared taking into account the Input and Output dataset on the relevant parameters for the period 2000 2010. This study was only limited to the Constant Returns-to-Scale (CRS) assumption DMUs. Applying output-oriented CRS model on the Indian banking sector using the DEA methodology to evaluate banks' performance forms a substantial contribution to the field.

Reviewing these studies shows that DEA is effective in measuring firms' performance using the available financial variables and ratios that are obtained from financial statements over a particular period. Moreover, it can be concluded that DEA can be used either as an alternative of the conventional input-output analysis and the simple ratio analysis or as a complement to these types of analysis to assess the performance of different firms. Finally, DEA can be employed in a wide range of analytical studies in the field of finance.

In another study, in an attempt to assess the Turkish firms' performance in the Istanbul Stock Exchange, Delen *et al*. [7], employed decision tree analysis. For the sake of validating the dimensions of financial ratios, Exploratory Factor Analysis (EFA) was used. Subsequently, and in order to identify the potential relationships between the firm performance and financial ratios, they use predictive modelling methods. To examine the influence of the financial ratios on firm performance, they used four decision tree algorithms including Quaternion Estimation (QUEST), Classification and Regression Trees (C&RT), Chi Square Automatic Interaction Detector (CHAID), and Classification 5.0 (C5.0). Following the development of prediction models, they measured the importance of independent variables performing information fusion-based sensitivity analyses. The best prediction accuracy was gained using CHAID and C5.0 while earnings before tax-to-equity ratio and net profit margin were found to be the two most important variables according to the sensitivity analysis results.

The data mining approach is also adopted for the prediction of stock prices in financial stock markets. AlDarmaki *et al*. [2], for example, analyzed the closing process in Dubai Financial Stock Market (DFM) using two data mining techniques: supervised classification algorithms and unsupervised regression algorithms. They built a model for predicting the closing stock prices for the companies listed in the DFM and for identifying the most influential factors on stock prices. The findings revealed that using the classification algorithm was succeeded in predicting the closing price with accuracy higher than 92%. The regression algorithm predicts the stock prices with a correlation coefficient equal to 0.8889. Therefore, since this model can assist investors in predicting the closing pricing, it helps them better plan their future investments.

In another study, Zibanezhad *et al*. [22] developed a model for predicting bankruptcy of the firms in Tehran Stock Exchange during the period 1996 to 2009, depending on financial ratios or predictor variables extracted from financial statements. For the sake of mining financial variables, they used the Clementine software and the classification and regression tree method. The model showed efficiency to predict bankruptcy with a94.5% accuracy. The followings were shown to be the most important variables for predicting bankruptcy: Earnings Before Interest and Taxes (EBIT) to Interest, EBIT to total assets, cash flow to long-term debt, total debt to total assets, net income to total assets, current assets inventory-advance prepayment to short-term debt and current assets to sales.

Finally, it is possible to combine DEA and data mining techniques in assessing DMU performance. For example, Yang held a fair comparison between DMUs in Canadian banking market dataset and used DEA together with k-means clustering to measure the performance of DMUs [20]. As an initial step, k means clustering divided the bank branches DMUs into several clusters. DEA was applied to each cluster.

3. Methodology

This section introduces the proposed approach applied on the dataset. The main components of our approach, HDEA-EM, include the DEA models, the variable importance ranker, J48 classifier, and the DECORATE ensemble method.

3.1. The Dataset

The dataset for this study was collected from Amman Stock Exchange website by using the latest published companies' guide for year 2016. The guide contains financial ratios and variables of the industrial companies in Amman Stock Exchange [3].

The companies listed under the industrial sector are categorized into groups including chemical, electrical, engineering and construction, food and beverages, glass and ceramic, mining and extraction, paper and cardboard, pharmaceutical and medical, printing and packaging, textiles, leathers and clothings, and to baccoand cigarettes.

The original dataset contains 70 companies under the eleven categories of the industrial companies as shown in Table 1. There was a single file for each company's financial data for the years 2012, 2013, 2014 and 2015. A manual merging process was applied in order to aggregate the industrial sector's companies into one file for each year to be processed. The result was ended up having four files for years 2012 to 2015.

Some of the companies under investigation were found to have no registered financial values in one or more of the three years (2012, 2013, or 2014) because they were not listed in the Amman Stock Exchange at that/those year(s).

Table 1. The categories of the companies in the industrial sector.

Category	No. of companies
Chemical industries	9
Electrical industries	4
Engineering and construction	9
Food and beverages	12
Glass and ceramic	1
Mining and extraction	16
Paper and cardboard	3
Pharmaceutical and medical	7
Printing and packaging	1
textiles, leathers and clothings	6
Tobacco and cigarettes	2

Other companies were found to have no registered values in one or more years because of their closure or merging to other companies. All companies that have missing financial value in one or more years were excluded because DEA models do not accept null or zero values. Consequently, a manual process was applied to filter the companies. After this filtering process, the dataset turned out to include 53 companies.

There were 73 financial variables and 21 financial ratios for each company in the original dataset. Accordingly, another manual filtering process was applied to exclude the financial variables and ratios that have zero or null value(s). The variables that have a fixed value for all companies were also excluded due to its statistical insignificance. As a result, the dataset ended up containing 11 financial variables as inputs and 11 financial ratios as outputs for each company, as shown in Tables 2 and 3. We uploaded the dataset at the Mendeley Data website to be publically accessed https://data.mendeley.com/datasets/wpc276x84n/2.

The following rules are used to make sure that the DEA model is discriminatory:

a. Boussofiane *et al*. [4], provide that the number of DMUs should be the multiple of the number of inputs and outputs.
b. Golany and Roll [13] provide that the number of DMUs should be twice of the number of inputs and outputs at least.
c. Bowlin [5] provides that the number of DMUs should be three times of the number of inputs and outputs at least.
d. Dyson *et al*. [11], provide that the number of DMUs should be the product of number of inputs and outputs multiplied by 2.
e. In any DEA model, one of these rules of thumbs should be considered [19]. As mentioned earlier, the number of inputs and outputs were reduced to 11 each.

Table 2. Financial variables.

ID	Financial Variable
I1	Financial variable
I2	Total Current Assets
I3	Total Fixed Assets
I4	Total Assets
I5	Total Current Liabilities
I6	Total Liabilities
I7	Total Shareholders' Equity
I8	Operating Revenues
I9	Gross Profit
I10	Income Before Interest & Tax
I11	Net Income

Table 3. Financial ratios.

ID	Financial Variable
I1	Financial variable
I2	Total Current Assets
I3	Total Fixed Assets
I4	Total Assets
I5	Total Current Liabilities
I6	Total Liabilities
I7	Total Shareholders' Equity
I8	Operating Revenues
I9	Gross Profit
I10	Income Before Interest & Tax
I11	Net Income

The number of DMUs were also reduced to 53 firms. These numbers satisfy the second rule of thumb.

However, it remains to be said that the present study has the following limitations:

- Data Envelopment Analysis models do not deal with null, zero, and negative values.
- Selecting inputs and outputs for the DEA models as well as balancing between the number of Decision Making Unites and the number of the selected inputs and outputs are time consuming and tedious.
- The Dataset was unbalanced in magnitudes.

3.2. The HDEA-EM Approach

As mentioned earlier, the HDEA-EM approach is a hybrid one whereby both DEA and DECORATE ensemble method are adopted to evaluate firms' performance in Jordanian industrial sector. The general framework of the approach is described in Figure 1.

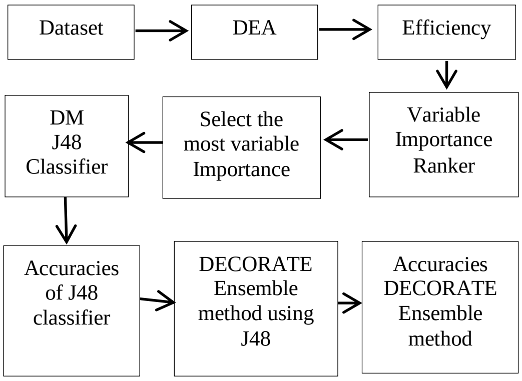

Figure 1. Framework flow chart.

3.2.1. DEA Models

For applying DEA models on the preprocessed dataset, the Open Source Data Envelopment Analysis (OSDEA) software was used to run DEA models. Every DEA model requires a dataset that contains DMUs, inputs, and outputs in order to give each DMU in the dataset an efficiency score compared to other DMUs in the same dataset. The efficiency score will be between 0 and 1. The DMU that gets 1 as efficiency score is considered efficient DMU, otherwise, it is considered inefficient DMU.

Several experiments were carried out in this stage. More specifically, four DEA models were applied for each year in the dataset. These four DEA models resulted from combining each DEA orientation (i.e., input and output) with each DEA return to scale type (i.e., constant and variable). Table 4 shows the four combinations that resulted from combining the orientations and the return to scale types of DEA.

The result of this stage is a class value for each DMU in the dataset regardless of the applied DEA models. Only the results of the output orientation models in both scales for the years 2012 to 2015 were considered. This study focuses on maximizing the output ratios while maintaining the same inputs. In other words, the study is aimed to gain the best output ratios by using the same operated variables from a financial perspective. In addition, the results of Output-Oriented Constant Returns-to-Scale (OO-CRS) and Output-Oriented Variable Returns-to-Scale (OO-VRS) for firms over the four years were aggregated together.

3.2.2. The Variable Importance Ranker

The purpose of assessing variables includes a) to find out the most important independent financial variable(s) and/or ratio(s) that affect(s) firms' efficiency, and b) to find out the least important independent financial variable(s) and/or ratio(s). The information gain attribute evaluation ranks variables and shows their importance based on their information gain values with respect to the classes. The ranker is implemented in the WEKA software.

Table 4. DEA orientations and return to scale types.

Name	Orientation and return to scale type
IO-CRS	Input-oriented Constant Returns to Scale
IO-VRS	Input-oriented Variable Returns to Scale
OO-CRS	Output-oriented Constant Returns to Scale
OO-VRS	Output-oriented Variable Returns to Scale

3.2.3. J48 Decision Tree Classifier

Constructing a model of classes from a set of tuples that have class labels is the classification process. Decision tree algorithm is used to find out the way of how the attributes-vector behaves for a number of instances. The classes for the new generated tuples are being found based on the training instances [21]. The rules for the prediction of the target variable are generated by the algorithm, J48, which is an extension of ID3. The J48 added the following features: missing values accounting, pruning decision trees, continuous attribute value ranges, and rules derivation, etc.

The third step of the approach is to apply the J48 decision tree classifier by utilizing the output of the DEA stage as a trained dataset. To enhance the J48 accuracy, the DECORATE ensemble method was used.

3.2.4. DECORATE Ensemble Method

DECORATE stands for Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples. DECORATE builds diverse ensembles of classifiers by using especially constructed artificial training examples.

DECORATE generates the ensemble iteratively as follows: the classifier is learned firstly and then it is added to the ensemble. The first ensemble is initialized as containing the classifier, which is trained on a given training data. In each iteration, classifiers are trained on auto-generated artificial data, which is combined with the original training data. The class labels for the artificial instances are selected to differ maximally from the current ensemble's predictions. Diversity data is the labeled artificially generated training set. The new classifier is trained on the data set of the training data and the diversity data, so forcing it to be different from the current ensemble. Therefore, adding this classifier to the ensemble will increase its diversity. A new classifier is rejected in case that if it is added to the existing ensemble, the accuracy will be decreased.

4. Results and Discussion

Figure 2 shows all financial variables and ratios ranked from the most important variable/ratio to the least one, which is the result of applying the Information Gain Attribute Evaluation. In addition, Figure 2 represents the most three important variables and/or ratios for industrial firms, which are Operation Revenues, Total Fixed Assets, and Total Assets.

The twenty-two independent financial variables and ratios, which are discussed in Tables 2 and 3, were used for predicting the firms' performance. The efficiency column (class variable) is handled with efficient or inefficient values. These values indicate the performance of the industrial firms. Table 5 presents the confusion matrix of the binary (efficient, inefficient) model. The matrix has the following four probabilities for any instance in the dataset:

1. True Positive: when the instance (firm) is efficient and classified as efficient by the classifier.
2. True Negative: when the instance (firm) is inefficient and classified as inefficient by the classifier.
3. False Negative: when the instance (firm) is inefficient but classified as efficient by the classifier.

Table 5. Firms' Performance confusion matrix.

	Observed Efficient	**Observed Inefficient**
Predicted (efficient)	True Positive	False Positive
Predicted (inefficient)	False Negative	True Negative

As for the accuracy measures of performance, the overall accuracy, precision, recall and F-Measure values were calculated in order to compare J48 classifier with the DECORATE ensemble method, as shown in the following equations:

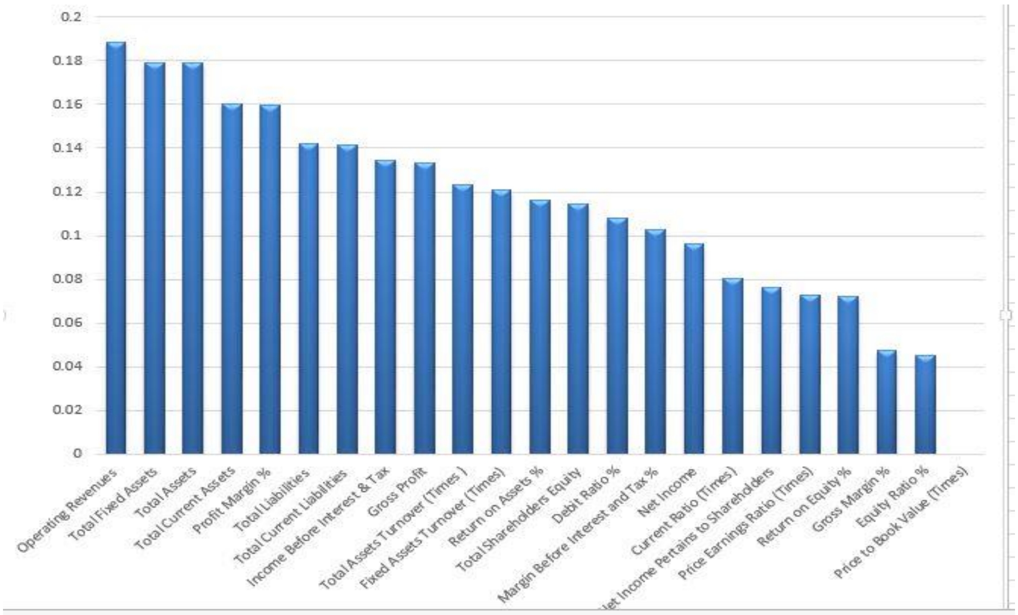

Figure 2. Industrial firms' financial variables and ratios importance by their InfoGain AttribEval values.

1. Overall Accuracy is defined as the percentage of the number of cases that were correctly classified to the overall number of cases, as Equation (2) illustrates:

$$Accurcy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

2. Precision refers to the ratio of the number of cases that were correctly classified as positive (efficient) to the sum of the number of cases that were correctly and incorrectly classified as positive, as shown in Equation (3):

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3. Recall or Sensitivity is defined as the ratio of the number of cases that were correctly classified as positive (efficient) to the sum of the number of cases that were correctly classified as positive and incorrectly classified as negative, as shown in Equation (4):

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4. F-Measure considers both recall and precision, as showed in Equation (5):

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Han *et al*. [14] provide details of Equations (2), (3), (4), and (5). Table 6 presents the comparison between J48 and the DECORATE ensemble accuracy results for the firms. As shown Table 6, the J48 shows overall accuracy 79.9%, precision 81.9%, recall 80%, and f measure 80.4%. Also, it shows the accuracy enhancement after applying the DECORATE ensemble for the J48. The DECORATE shows overall accuracy 81.6%, precision 82.1%, recall 81.6%, and F-measure with 81.8%. Comparing the DECORATE ensemble method with the J48 Classifier, it is clear that the DECORATE ensemble enhances the J48 decision tree accuracies as follows: 1.7% in the overall accuracy, 0.2% in the precision, 1.6% in the recall, 1.4% in the F-measure.

Table 6. Comparison between J48 and DECORATE.

Classifier	Overall Accuracy	Precision	(Sensitivity/ Recall)	F-Measure
J48	79.9%	81.9%	80%	80.4%
Decorate(J48)	81.6%	82.1%	81.6%	81.8%
Enhancement	1.7%	0.2%	1.6%	1.4%

Table 7 illustrates the confusion matrix of J48 classifier and DECORATE ensemble method for predicting industrial firms' performance.

Table 7. Confusion matrix of J48 and DECORATE.

	Observed Efficient	Observed Inefficient	Observed Efficient	Observed inefficient
	J48		DECORATE	
Predicted (efficient)	226	26	243	33
Predicted (inefficient)	59	110	45	103

Reviewing Table 7 reveals that DECORATE increased the True Positive instances (firms) from 226 to 243, increased the False Positive instances (firms) from 26 to 33, and decreased the False Negative instances (firms) from 59 to 45. However, DECORATE decreased the number of True Negative instances (firms) from 110 to 103.

5. Conclusions

In this study, the HDEA-EM hybrid approach was developed for measuring Jordanian firms' performance in the industrial sector. The work required processing a dataset that contains firms, financial variables, and ratios over previous years. The dataset was extracted from Amman Stock Exchange for the industrial sector. The efficiency variable was extracted for these firms (as DMUs in the approach) based on their financial variables as multiple inputs and financial ratios as multiple outputs by using the DEA as a first stage. Then, as a second stage, the information gain evaluation was adopted to identify the most important financial variables and/or ratios with respect to firms' performance in the industrial sectors. Then, in the third stage, the data mining J48 decision tree classifier is also used to get the accuracies of classifying the industrial firms. As a final stage, the DECORATE ensemble method was used in order to enhance the accuracy of the J48 decision tree.

The study reveals that the result of applying information gain evaluation as a second stage showed that operating revenues, total fixed assets and total asset are the most three influential variables on the industrial firms' performance. The study also shows that DECORATE enhanced the J48 classifier accuracies as follows: 1.7% in the overall accuracy, 0.2% in the precision, 1.6% in the recall, 1.4% in the F-measure.

6. Future Work

As for future work, it is recommended that the study can be expanded by including the financial sector as well as the upcoming year(s) that will be listed by Amman stock Exchange for financial, industrial and services sectors. Moreover, it is recommended that the study can be improved by adopting different methods such as the use of different data mining techniques (e.g., clustering) in order to categories firms with relative values under new groups or clusters.

References

[1] Aksoy E. and Yildiz A., *Global Business Strategies in Crisis*, Spring Link, 2017.

[2] AlDarmaki N., Mohamed E., AlMansouri N., Ahmed I., and Zaki N., "Prediction of the Closing Price in the Dubai Financial Market: A Data Mining Approach," *in Proceedings of 3rd MEC International Conference on Big Data and Smart City*, Muscat, pp. 1-7, 2016.

[3] Amman Stock Exchange, https://www.ase.com.jo/en, Last Visited, 2016.

[4] Boussofiane A., Dyson R., and Thanassoulis E., "Applied Data Envelopment Analysis," *European Journal of Operational Research*, vol. 52, no. 1, pp. 1-5, 1991.

[5] Bowlin W., "Measuring Performance: An Introduction to Data Envelopment Analysis," *The Journal of Cost Analysis*, vol. 15, no. 2, pp. 3-27, 1998.

[6] Chen W. and Du Y., "Using Neural Networks and Data Mining Techniques for the Financial Distress Prediction Model," *Expert Systems with Applications*, vol. 36, no. 2, pp. 4075-4086, 2009.

[7] Delen D., Kuzey C., and Uyar A., "Measuring Firm Performance Using Financial Ratios: A Decision Tree Approach," *Expert Systems with Applications*, vol. 40, no. 10, pp. 3970-3983, 2013.

[8] Devinney T., Yip G., and Johnson G., "Using Frontier Analysis to Evaluate Company Performance," *British Journal of Management*, vol. 21, no. 4, pp. 921-938, 2010.

[9] Dietterich T., "Ensemble Methods in Machine Learning," *in Proceedings of International Workshop on Multiple Classifier Systems*, Cagliari, pp. 1-15, 2000.

[10] Doumpos M. and Cohen S., "Applying Data Envelopment Analysis on Accounting Data to Assess and Optimize the Efficiency of Greek Local Governments," *Omega*, vol. 46, pp. 74-85, 2014.

[11] Dyson R., Allen R., Camanho A., Podinovski V., Sarrico C., and Shale E., "Pitfalls and Protocols in DEA," *European Journal of Operational Research*, vol. 132, no. 2, pp. 245-59, 2001.

[12] Emrouznejad A. and Charles V., Big Data for the Greater Good, *Springer*, 2019.

[13] Golany B. and Roll Y., "An Application Procedure for DEA," *Omega*, vol. 17, no. 3, pp. 237-250, 1989.

[14] Han J., Kamber M., and Pei J., *Data Mining: Concepts and Techniques*, Elsevier, 2011.

[15] Hatefi M. and Fasanghari M., "A Dea-Based Approach for Information Technology Risk Assessment through Risk Information Technology Framework," *The International Arab Journal of Information Technology*, vol. 13, no. 1, pp. 51-58, 2016.

[16] Kim M. and Kang D., "Ensemble with Neural Networks for Bankruptcy Prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3373-3379, 2010.

[17] Kumar M., Charles V., and Mishra C., "Evaluating the Performance of Indian Banking Sector Using DEA During Post-Reform and Global Financial Crisis," *Journal of Business Economics and Management*, vol. 17, no. 1, pp. 156-172, 2016.

[18] Paul J. and Das K., "Efficiency of Commercial Banks in India: A Non-Parametric Study Using Data Envelopment Analysis," *Asian Journal of Research in Banking and Finance*, vol. 5, no. 6, pp. 37-48, 2015.

[19] Sarkis J., Modeling Data Irregularities and *Structural Complexities in Data Envelopment Analysis*, Spring Link, 2007.

[20] Yang X. and Dimitrov S., "Data Envelopment Analysis May Obfuscate Corporate Financial Data: Using Support Vector Machine and Data Envelopment Analysis to Predict Corporate Failure for Nonmanufacturing Firms," *Information Systems and Operational Research*, vol. 55, no. 4, pp. 295-311, 2017.

[21] Zaki M. and Meira Jr W., *Data Mining and Analysis Fundamental Concepts and Algorithms*, Cambridge University Press, 2014.

[22] Zibanezhad E., Foroghi D., and Monadjemi A., "Applying Decision Tree To Predict Bankruptcy," *in Proceedings of IEEE International Conference on Computer Science and Automation Engineering*, Shanghai, pp. 165-169, 2011.

Hassan Najadat is an associate professor of Computer Science. He works in the department of Computer Science and in the department of Computer Information Systems at Jordan University of Science and Technology, Irbid, Jordan. He earned his Ph.D. in Computer Science from North Dakota State University, ND, USA in 2005. His research interests include data mining, artificial intelligence, data science, and data envelopment analysis

Ibrahim Al-Daher received his BSc degree in Computer Information Systems from Jordan University of science and Technology, Irbid, Jordan in 2013 and the MSc degree in Computer Science from the same university in 2018. He has worked as a Research and Teaching Assistant in Computer information Systems department, Jordan University of Science and Technology, Irbid, Jordan in 2014. His Research interests include data mining, machine learning, and data envelopment analysis.

Khalid Alkhatib is an associate professor of accounting and finance at the department of Computer Information Systems at Jordan University of Science and Technology, Irbid, Jordan. He received his Ph.D. in accounting and finance, postgraduate diploma in social science research methods from Cardiff University/Wales, master's degree in management and bachelor's degree in business administration from the UK. His fields of interests are accounting information systems, technology and banking.