Yugoslav Journal of Operations Research
24 (2014), Number 3, 347-358
DOI: 10.2298/YJOR140430030R

# TWO-PHASED DEA-MLA APPROACH FOR PREDICTING EFFICIENCY OF NBA PLAYERS

Sandro RADOVANOVIĆ
*University of Belgrade,*
*Faculty of Organizational Sciences, Serbia*
*sandro.radovanovic@gmail.com*
Milan RADOJIČIĆ
*University of Belgrade,*
*Faculty of Organizational Sciences, Serbia*
*radojicic.milan@hotmail.com*
Gordana SAVIĆ
*University of Belgrade,*
*Faculty of Organizational Sciences, Serbia*
*gordana.savic@fon.bg.ac.rs*

Received: April 2014 / Accepted: August 2014

**Abstract**: In sports, a calculation of efficiency is considered to be one of the most challenging tasks. In this paper, DEA is used to evaluate an efficiency of the NBA players, based on multiple inputs and multiple outputs. The efficiency is evaluated for 26 NBA players at the guard position based on existing data. However, if we want to generate the efficiency for a new player, we would have to re-conduct the DEA analysis. Therefore, to predict the efficiency of a new player, machine learning algorithms are applied. The DEA results are incorporated as an input for the learning algorithms, defining thereby an efficiency frontier function form with high reliability. In this paper, linear regression, neural network, and support vector machines are used to predict an efficiency frontier. The results have shown that neural networks can predict the efficiency with an error less than 1%, and the linear regression with an error less than 2%.

**Keywords**: Data envelopment analysis; Efficiency analysis; Predictive analytics; Machine learning.

**MSC:** 90B50, 90C29.

# 1. INTRODUCTION

Contemporary sports have a significant impact on the world economy, and therefore increasing attention is paid to the analysis of sports teams and athletes. Consequently, it has become necessary to determine their impact, not only on the field, but also in the economy and society as a whole. In the field of sports analytics, clubs from the United States of America (like the Boston Red Sox) and clubs from Europe (like AC Milan) are making the fastest progress (Schumaker et al., 2010).

Sports analytics is considered primarily as a statistical analysis (t-test, $\chi 2$ test, ANOVA, descriptive statistics, etc.), analysis of efficiency, and more recently, a sports data mining. In most cases, events on the field, such as number of shots on goal, number of passes in 90 minutes of football game, or number of homeruns in baseball are being analysed in order to improve team results and to identify opponents' weaknesses. However, with the growth in popularity of sports and the funds invested in it, sports analytics require more complex analysis. As stated in Schumaker et al., (2010), the forerunner of data analysis in sports is Anatoly Zelentsov, who created a computer application which performed tests of mental stability, durability, memory, reaction time and coordination in Dynamo Kiev football club during the mid 70's. The application was used to determine whether young players were able to play for the first team. The results were surprisingly successful and allowed the Dynamo Kiev to win UEFA Cup Winners' Cup in 1975 and 1986. The efficiency analysis began with the work of Scully (1974) on baseball, and work of Zaketal (1979) on basketball. After achieving success in quantifying the relationship between sports-related inputs and sports success, the aforementioned authors' efficiency analysis found its application not only in basketball (Lee & Worthington, 2012; Hill & Jolly, 2012; Moreno & Lozano, 2014), but also in many other sports, such as football (Ribeiro & Lima, 2012; Fernandez et al., 2012), baseball (Jane, 2012; Regan, 2012), or chess (Jeremic & Radojicic, 2010).

Inspired by these and other works, this paper has a purpose to provide a comprehensive assessment of the efficiency of the National Basketball Association's (NBA) players. Fortunately, research in sports analytics and sports economics has recently embraced the statistical and mathematical methods for the assessment of sports efficiency. Furthermore, many statistical algorithms (linear regression and least median square regression) and machine learning algorithms (neural networks and support vector machines) are used to predict an efficiency frontier. This approach allows us to predict the relative efficiency of a new player based on the DEA efficiency indexes. In other words, this approach allows the prediction of a relative efficiency frontier using machine learning algorithms. The main idea is to ease a calculation of the relative efficiency for a new player without conducting a new DEA model.

The remainder of the paper is structured as follows. Section 2 explains the methodology, and Section 3 presents the finding and analysis. Section 3 is divided into two subsections, whereby the first one contains the ranking of NBA players through the DEA, and the second one presents the predicting and testing of the efficiency frontier. Section 4 concludes the paper.

## 2. METHODOLOGY

The NBA players at the guard position who had notable results (in terms of points, assists, and other basketball measures) during the season 2011/12 were considered for the efficiency analysis in this paper. In the first phase, selected players are considered as decision-making units (DMUs), and the comparative analysis of their efficiency is performed by data envelopment analysis. In the second phase, the DEA results are used as a basis for predicting the efficiency of a new player via frontier form learned by a machine learning algorithm.

### Phase 1 – Data envelopment analyses

Data envelopment analysis (DEA), introduced by (Charnes et al., 1978), is a popular non-parametric method for the relative efficiency evaluation. It allows the performance measurement of a decision making unit, compared to achievement of other units in the observing set (NBA players), which operate in similar circumstances and produce the same outputs using the same inputs (homogeneity property). During the last 30 years, DEA has been used for performance evaluation in different areas, from the non-profit sector, including evaluation in education, power plants, and hospitals (Savic et al., 2012; Sueyoshi & Mika, 2013) to the profit sector such as evaluation of banks, hotels and casinos (Tsang & Chen, 2012; Savic et al., 2013; Jayaraman et al., 2013). In order to make difference between the efficient NBA players, and to allow their ranking, super-efficiency measuring model proposed by (Andersen & Petersen, 1993) is used. Let us suppose that $DMU_j$ ($j$ = 1,…, $n$) uses inputs $x_{ij}$ ($i$= 1,…,$m$) to produce outputs outputs $y_{ij}$ (r = 1,…,s). Input-oriented weighted version of Andersen-Petersen's super-efficiency DEA model is as follows:

$$
\begin{aligned}
&(max)h_k = \sum_{r=1}^{s} \mu_r y_{rk} \\
&s.t. \\
&\sum_{i=1}^{m} v_i x_{ij} = 1 \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \; j = 1, \dots, n, \; j \neq k \\
&\mu_r \geq \varepsilon, \quad r = 1, \dots, s \\
&v_i \geq \varepsilon, \; i = 1, \dots, m
\end{aligned}
\tag{1}
$$

The optimal values of efficiency scores $h_k$ are obtained by solving the linear model $n$ times (once for each DMU in order to compare it with other DMUs). The efficiency score $h_k$ is greater or equal to 1 for all efficient units, and smaller than 1 for the inefficient units. This way, the ranking of units is enabled according to their efficiency (Ray, 2004).

The NBA players' database consists of eight indicators, two of which are considered to be input factors, and the six others as output factors. The input indicators for all players covered by analysis are gross salary and minutes on the court. The outputs used in analysis are number of points, number of assists, number of rebounds, number of steals, number of turnovers, and number of blocked shots made by a player during the regular season 2011/12. All data can be found on (National Basketball Association, 2013; ESPN, 2013). Based on the correlation matrix shown in Table 1, it can be concluded that the property of isotonicity is satisfied.

**Table 1.** Correlation matrix (Spearman's rho)

| | Points | Assists | Rebounds | Steals | Turnovers | Blocks |
|---|---|---|---|---|---|---|
| Salary | 0.540 | 0.006 | 0.361 | 0.092 | 0.062 | 0.100 |
| Minutes | 0.672 | 0.348 | 0.759 | 0.752 | 0.600 | 0.418 |

The input-oriented Andersen-Petersen's variable return to scale DEA is the most suitable model for a given problem because increasing an input does not result in the proportional increase of the output. The input-oriented model is used because a decision maker can only influence inputs, i.e. the sports-team managers can consider to reduce the gross salary for the next year, or to limit the players' playing time, while they cannot affect the number of points scored. According to (Lovell & Rouse, 2003), one of the most important factors, especially when variable return to scale model is supposed, is the control of the weight restrictions. A measurement of the super efficiency in the variable return to scale model can cause unnatural solutions, or even the case in which the model does not have a solution. In order to avoid this problem, the assurance regions of type I and II were introduced.

$$0{,}2 \leq \frac{\text{Salary}}{\text{Points}} \leq 2 \tag{2}$$

$$0{,}3 \leq \frac{\text{Minutes}}{\text{Salary}} \leq 1 \tag{3}$$

$$(\text{Points+Assists+Turnovers})\text{-}(\text{Rebounds+Steals+Blocks}) = 0 \tag{4}$$

$$0{,}4 \leq \frac{\text{Steals}}{\text{Turnovers}} \leq 2 \tag{5}$$

A modified super-efficiency DEA model (1)-(5) is used for the efficiency evaluation of NBA players considered as DMUs.

**Phase II – Predictive analytics**

Predictive analytics presents a variety of techniques, such as modelling, machine learning, and data mining, used to analyse present and historical facts in order to make predictions about future events. Predictive analytics deals with extracting information from data and using them to predict trends and behaviour patterns. It can refer to predictive, descriptive, or decision modelling. Therefore, the DEA can be considered as a descriptive but not predictive model since it evaluates a comparative efficiency within an observed set. New analyses should be performed in order to evaluate efficiency of a new DMU.

Regression and machine learning techniques can be used as predictive analytics techniques. Machine learning was originally employed to develop techniques for enabling computers to learn, but nowadays it includes a number of advanced methods for regression and classification (Mitchel, 1997). In certain applications, it is sufficient to directly predict the dependent variable without focusing on the underlying relationships between variables. In other cases, the underlying relationships can be very complex with unknown mathematical form of the dependencies. In such cases, machine learning techniques emulate human cognition and learn from training examples to predict future events. Different techniques can be used for machine learning. This research has used the models of linear regression, least median square regression, isotonic regression, 5-nearest neighbours, Gaussian process with the radial-basis function (RBF) kernel, support vector machines for the regression with a dot kernel, and neural network, through the RapidMiner software (Mierswa et. al., 2006).

In this study, special attention is given to support vector regression (SVR) (Drucker et al., 1996), widely applied in the field of regression and approximation. The objective is to learn an unknown function based on a training set of N input-output pairs in a black box modelling approach. DEA also evaluates an efficiency frontier considering input-output process as a black box. However, the approximation performance of SVR relies on the training data and a kernel function. The kernel is called an admissible support-vector kernel (SV kernel) if the Mercer's condition is satisfied (Zhang et al., 2010). The Mercer's condition is one of the most popular methods to validate whether a prospective kernel is a positive definite function, since any SV kernel should be capable of corresponding to a dot product in a high-dimensional feature space. The kernel function is regarded as a useful trick, which benefits the computation of dot products in the feature space using simple function defined on the pairs of input patterns. All SVR algorithms aim at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data, based on the structural risk minimization principle. The main idea is to train a model which minimizes a general risk function.

For the purpose of our study, we used the DEA efficiency results trying to predict efficiency on a new DMU by training data set and by predicting the shape of the efficiency frontier. A hybrid approach, which combines the DEA method, rough set and support vector machines, was used in (Yeh et. al., 2010) for predicting a business failure. DEA and machine learning are also used for clustering, and to determine a stepwise path for improving efficiency of each inefficient system integration project (Hong et al., 1999).

## 3. FINDINGS AND ANALYSIS

In order to obtain the DEA results, software EMS 1.3 was used (Scheel, 2000) for academic purposes. For the need of the analysis, we choose 26 players and the data taken from the official statistics of National Basketball Association 2013 (ESPN, 2013).

### 3.1. NBA players ranking

Based on the results of the efficiency analysis, ten out of 24 players have been efficient (Table 2). Player with the highest efficiency score is John Wall, with score of 115.30%. On the second position is Russell Westbrook with 114.26%. O.J. Mayo, James Harden, and Mo Williams have been close to the efficiency frontier, while Jason Terry, Kobe Bryant, and Joe Johnson have been extremely inefficient.

**Table 2.** Players ranking

| Rank | DMU | Score | Rank | DMU | Score |
|---|---|---|---|---|---|
| 1 | *John Wall* | *115.30%* | 14 | Tyreke Evans | 94.23% |
| 2 | *Russell Westbrook* | *114.26%* | 15 | Danny Granger | 91.11% |
| 3 | *Dwayne Wade* | *108.95%* | 16 | Andre Iguodala | 90.45% |
| 4 | *Derrick Rose* | *107.23%* | 17 | Carmelo Anthony | 88.24% |
| 5 | *Rajon Rondo* | *104.72%* | 18 | Tony Parker | 84.57% |
| 6 | *Ray Allen* | *104.24%* | 19 | Rudy Gay | 83.99% |
| 7 | *Kyrie Irving* | *102.23%* | 20 | Ben Gordon | 83.33% |
| 8 | *Kevin Durant* | *101.78%* | 21 | Monta Ellis | 82.68% |
| 9 | *LeBron James* | *101.55%* | 22 | Paul Pierce | 82.21% |
| 10 | *Chris Paul* | *100.47%* | 23 | Deron Williams | 80.81% |
| 11 | O.J. Mayo | 99.49% | 24 | Jason Terry | 78.06% |
| 12 | James Harden | 98.52% | 25 | Kobe Bryant | 74.59% |
| 13 | Mo Williams | 97.18% | 26 | Joe Johnson | 71.73% |

For each inefficient unit, DEA identifies the corresponding set of efficient units, making a reference group for that inefficient unit. This group consists of units that are optimal, with their optimal weights (Ray, 2004).

Authors (Radovanovic et al, 2013) have shown that the DEA results are correlated with the results of other official ranking methods when salary is included. They have also concluded that DEA can be an appropriate method to measure the NBA player's efficiency because it includes an additional dimension, a salary, which is very important in economic sense. Now, it can be concluded that DEA can change the approach to the efficiency of players.

### 3.2. Predicting the efficiency frontier

When a new player is added to the efficiency analysis, the whole computation of the mathematical model has to be repeated. If there are too much DMUs, or if we want to compare whether a player is efficient in a different group of DMUs, it could be time consuming. Therefore, the efficiency frontier can be predicted through regression models, and the efficiency index for a new player can be derived without re-conducting the mathematical model. In this research, several regression algorithms were tested: default model (which always return an average and is used as a benchmark model), linear regression, least-median square regression, isotonic regression, 5-nearest neighbours, Gaussian process with the RBF kernel, support vector machines for the regression with a dot kernel, and neural network.

- Linear regression used the M5 Prime feature selection technique, which selected the attribute with the smallest standardized coefficient, removed it and performed the regression analysis. When the result improved the Akaike information criterion (AIC), the attribute was dropped. This process was repeated until there were no more attributes remaining to be dropped.
- Gaussian process used the radial-basis function to generate the maximum of 200 kernels.
- Support vector machine for the regression used dot kernel and had the complexity parameter with the zero value.
- Neural network used a learning rate of 0.3 and a momentum of 0.2 with 8 neurons in the hidden layer.

The experiment was conducted on a super-efficiency model, and in order to achieve better results, new features were constructed. These were of a simple efficiency, and could be presented as the ratio of the sum of outputs to the sum of inputs, expressed as an output per dollar and an input per minute. Each feature could be calculated based on a single player, and therefore did not require any further action from the user. Training and testing of models were performed by the RapidMiner software (Mierswa et al., 2006) through the leave-one-out validation, meaning that the training of the model was performed on 25 players and tested on one player (since 26 players were used in this research). The process was repeated 26 times, so each player was used exactly once in a test data set. Finally, an average performance for a player was reported as a result.

A mean absolute error was used as a performance measure, and the results related to performances of different learning algorithms were given in Table 3. For each algorithm performance: mean absolute error and standard deviation are given. If an average efficiency had been used to predict a player's efficiency, an average error of prediction would have been 10.4%. The best performance was obtained with a neural network where mean absolute error was 0.007, meaning that the error of prediction was less than one percent. For further analyses, we will show the results of the linear regression as a second best option, and the support vector machine which uses the similar propositions as DEA.

**Table 3.** Performance of learning algorithms

| ***Algorithm*** | ***Mean absolute error*** |
|---|---|
| ***Default model*** | ***0.104 +/- 0.067*** |
| ***Linear regression*** | ***0.019 +/- 0.013*** |
| ***Least median square regression*** | ***0.028 +/- 0.039*** |
| ***Isotonic regression*** | ***0.070 +/- 0.043*** |
| ***5-NN*** | ***0.062 +/- 0.050*** |
| ***Gaussian process*** | ***0.054 +/- 0.033*** |
| ***RBF Network*** | ***0.075 +/- 0.048*** |
| ***SVR*** | ***0.027 +/- 0.042*** |
| ***Neural network*** | ***0.007 +/- 0.006*** |

The efficiency index (score) and predictions obtained by three selected algorithms are shown in the following figures.

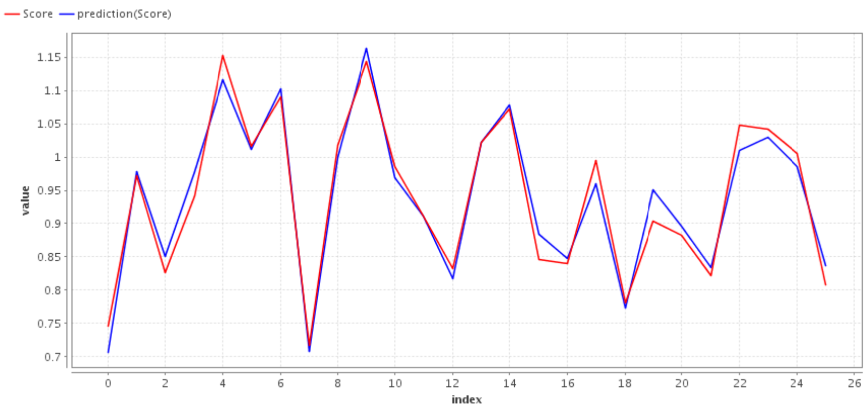


**Figure 1**. Efficiency frontiers – linear regression

The advantage of a linear regression over other more complex algorithms is the ability to generate a human-understandable model with a weight vector over attribute space. The generated linear regression model for the given problem is presented below:

Efficiency = -0.153 * Salary – 0.568 * Minutes + 0.308 * Points + 0.376 * Assists + 0.354 * Rebounds + 0.367 * Steals + 0.172 * Blocks + 0.434 * 1/Turnovers - 0.464 * Points per minute + 0.311 * Points per dollar + 0.634

With this equation for a calculation of player efficiency, a decision maker can see which attributes influence efficiency more and which do not influence it at all. Since the linear regression initially used the M5 Prime feature-selection techniques in order to eliminate highly collinear attributes, the simple efficiency was not selected as an important attribute.

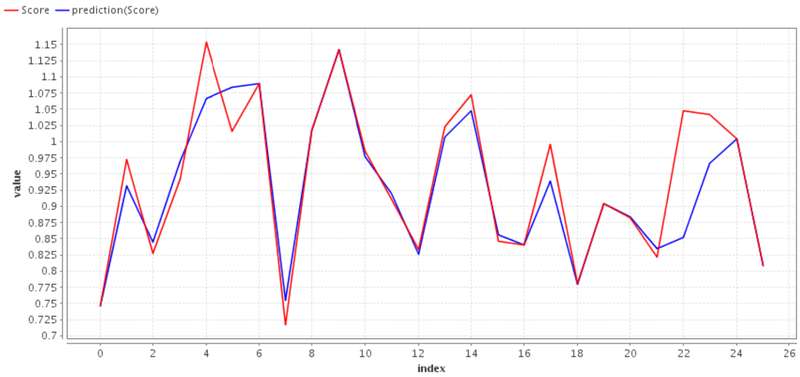


**Figure 2**. Efficiency frontiers – support vector machine

As illustrated in Figure 2, the support vector machine algorithm does not generate as successful results as the linear regression algorithm (there is a larger difference between actual scores and the prediction). Support vector machines, as well as linear regression, generate weights from which the attribute's importance can be obtained (Table 4).

**Table 4.** Support vector machine weights

| Salary | Minutes | Points | Assists | Rebounds | Steals | Blocks | 1/Turnovers | Stat Per Minute | Stat Per Dollar | Simple Efficiency |
|---|---|---|---|---|---|---|---|---|---|---|
| -0.034 | -0.098 | 0.073 | 0.041 | 0.027 | 0.019 | 0.057 | 0.067 | -0.021 | -0.011 | 0.086 |

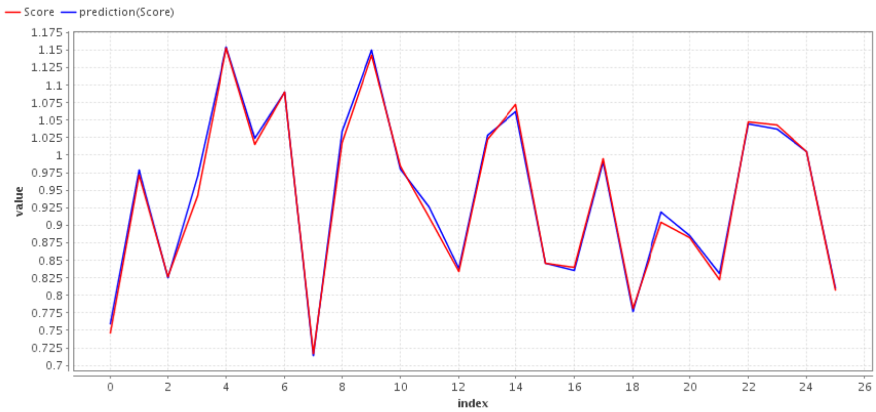


**Figure 3.** Efficiency frontiers – neural networks

The mean absolute error has shown that neural networks define the best fitting efficiency frontier on the entire data set, Figure 3.

For the purpose of further evaluation, we repeated analyses by splitting the data set, whereby 70% of data was used as a training data set, and the rest of 30% was used for testing purposes. It is worth to notice that the local random seed with the value of 1992 was used in order to get repeatable results. Validation results are shown in Table 5.

**Table 5.** Performance of learning algorithms

| ***Algorithm*** | ***Mean absolute error (on 30%)*** |
|---|---|
| ***Linear regression*** | ***0.044 +/- 0.053*** |
| ***SVR*** | ***0.079 +/- 0.065*** |
| ***Neural network*** | ***0.055 +/- 0.059*** |

The best results on the smaller data set were obtained by liner regression (whereby the mean absolute error was 4.4%), meaning that this method showed the highest level of robustness. Eventually, we can make a conclusion that the DEA results can be used as a basis for “learning“ the frontier, and for predicting efficiency through the machine learning algorithms, such as linear regression, neural networks, and support vector machines for regression.

## 4. CONCLUSION

In the first phase, this paper employed the DEA to evaluate the efficiency of NBA players during the regular season 2011/2012. In the second phase, machine learning techniques were used to predict an efficiency frontier. Using the data obtained from NBA.com, and the available sources of information on NBA salaries, we had an opportunity to examine the efficiency in a non-traditional way. We were able to perform the ranking through the Andersen-Petersen's model. Thus, we included salary in a calculation of efficiency, which was crucial for making important decisions, such as hiring, play positions, salaries.

Afterwards, we used machine learning algorithms, such as linear regression, support vector machines, and neural networks to predict the efficiency of a new DMU (player). Thus, we tried to overcome the weaknesses of the DEA. Namely, the DEA model is appropriate to estimate the relative efficiency of a DMU, but in order to evaluate the efficiency of a new DMU, we need to develop and solve the new DEA model. In this paper, through the example of 26 NBA players, we have shown that the DEA efficiency indexes can be used for “learning” of models through various machine learning algorithms. The results obtained by a neural network are highly reliable, with the expected absolute mean error approximately equal to 0.7%. We also have shown that the results of the linear regression algorithm fit better for the smaller data set. Expected absolute error for testing on 30% of data set was the lowest (4.4%). This paper has used machine learning (regression) algorithms as suitable method for the efficiency frontier prediction. The machine learning (regression) results have been claimed as “more accurate estimation than the value determined by OLS and NN for determining the “true” or optimal DEA frontier“ in doctoral dissertation by Poiter (2010). In other papers, such as Jiang et. al. (2013), and Kao et. al. (2013), the DEA and SVM combination has been used to improve classification, especially in handling the small data set. The results presented in this paper indicate that the two-phase DEA-MLA approach provides a "user friendly" efficiency evaluation, and a framework for prediction.

As a part of our future work, we plan to perform other types of relative-efficiency evaluation, such as distance based analysis (DBA). Furthermore, the efficiency frontier for this efficiency can be, like in this paper, predicted through the machine-learning algorithms. We also plan to compare the results gathered with the DEA and DBA. One way to improve this is to consider player efficiency over time.

Additionally, we plan to evaluate the efficiency on a team level, like in the paper (Aizemberg et al., 2014), where the cross efficiency of a team was evaluated over several seasons. By adding the machine learning algorithms, we can predict the efficiency of a team or player in the following years.

## REFERENCES

[1] Aizemberg, L., Roboredo, M. C., Ramos, T. G., de Mello, J. C. C. S., Meza, L. A., Alves, A. M., "Measuring the NBA Teams' Cross-Efficiency by DEA Game", *American Journal of Operations Research*, 4 (03) (2014) 101. DOI=http://dx.doi.org/10.4236/ajor.2014.43010.

[2] Andersen, P., Petersen, N. C., "A Procedure for Ranking Efficient Units in Data Envelopment Analysis", *Management Science*, 39 (10) (1993) 1261-1264.

[3] Charnes, A., "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2 (6) (1978) 429-444.

[4] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V., "Support Vector Regression Machines", *Advances in Neural Information Processing Systems*, 9 (1997) 155-161.

[5] ESPN, "ESPN", 2013, Retrieved from http://espn.go.com/nba/salaries.

[6] Fernandez, R. C., Gomez Nunez, T., Garrido, R. S., "Analysis of the Efficiency of Spanish Soccer League Players (2009/10) Using the Metafrontier Approach*", Estudios de economíaaplicada*, 30(2) (2012) 565-578.

[7] Hill, J., Jolly, N., "Salary Distribution and Collective Bargaining Agreements: A Case Study of the NBA*", Industrial Relations: A Journal of Economy and Society*, 51 (2) (2012) 342-363.

[8] Hong K.H., Ha S.H., Shin C.K., Park S.C., "Evaluating the Efficiency of System Integration Projects Using Data Envelopment Analyses (DEA) and Machine Learning", *Expert Systems with Application*, 16 (1999) 283-296.

[9] Jane, W.J., "Overpayment and Reservation Salary in the Nippon Professional Baseball League: A Stochastic Frontier Analysis", *Journal of Sports Economics*, (2012) 579-598.

[10] Jayaraman, A. R., Srinivasan, M. R., Jeremic, V., "Empirical Analysis of Banks in India using DBA and DEA" *Management*, 18 (69) (2013) 25-35.

[11] Jeremic, V., Radojicic, Z., "A New Approach in the Evaluation of Team Chess Championships Rankings", *Journal of Quantitative Analysis in Sports*, 6 (3) (2010) 1-11.

[12] Jiang, B., Chen, W., Zhang, H., Pan, W., "Supplier's Efficiency and Performance Evaluation using DEA-SVM Approach", *Journal of Software, North America*, 8 (1) (2013) 25-30.

[13] Kao H-Y, Chang T-K, Chang Y-C, "Classification of Hospital Web Security Efficiency Using Data Envelopment Analysis and Support Vector Machine," *Mathematical Problems in Engineering*, (2013) Article ID 542314, 8 pages, 2013. doi:10.1155/2013/542314

[14] Lee, B. L., Worthington, A. C., "A Note on the 'Linsanity' of Measuring the Relative Efficiency of National Basketball Association Guards", *Applied Economics*, 45 (29) (2013) 4193-4202.

[15] Lovell, C., Rouse, A., "Equivalent standard DEA models to provide super-efficiency scores", *Journal of the Operational Research Society*, 54 (2003) 101-108.

[16] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T., “Yale: Rapid Prototyping for Complex Data Mining Tasks”, in: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, (2006) 935-940, ACM.
[17] Mitchell T., “*Machine Learning*”, McGraw Hill, 1997.
[18] Moreno, P., Lozano, S., “A Network DEA Assessment of Team Efficiency in the NBA”, *Annals of Operations Research*, 214 (1) (2014) 99-124.
[19] National Basketball Association, National Basketball Association, 2013, Retrieved from http://www.nba.com/.
[20] Poiter K., *True Efficient Frontiers in Production Performance Assessment Using Machine Learning and Data Envelopment Analysis*, Phd Thesis, (2010) Retrievedform http://scholarlyrepository.miami.edu/cgi/viewcontent.cgi?article=1934&context=oa_dissertations
[21] Radovanović, S., Radojičić, M., Jeremić, V., Savić, G., “A Novel Approach in Evaluating Efficiency of Basketball Players”, *Management*, 67 (2013) 37-45.
[22] Ray, S., “Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research”, United Kingdom, Cambridge, Press Syndicate of the University of Cambridge, 2004.
[23] Regan, C. S., “The Price of Efficiency: Examining the Effects of Payroll Efficiency on Major League Baseball Attendance”, *Applied Economics Letters*, 19(11) 2012 1007-1015.
[24] Ribeiro, A., and Lima, F., “Portuguese Football League Efficiency and Players' Wages”, *Applied Economics Letters*, 19 (6) (2012) 599-602.
[25] Savic, G., Makajic-Nikolic, D., Suknovic, M., “AHP-DEA Measure for Study Program Selection”, SymOrg 2012, Zlatibor, University of Belgrade, Faculty of Organizational Sciences, 1217-1223.
[26] Savic, G., Radosavljevic, M., Ilijevski, D., “DEA Window Analysis Approach for Measuring the Efficiency of Serbian Banks Based on Panel Data” *Management,* 65 (2013) 5-14.
[27] Scheel, H., “EMS: Efficiency Measurement System User’s Manual”, Operations Research and Witshaftinsformetik, University of Dortmund, Germany, 2000.
[28] Schumaker, R., Solieman, O., Chen, H., “Sports Data Mining”, New York, Springer, 2010.
[29] Scully, G. W., “Pay and Performance in Major League Baseball*”, The American Economic Review*, 64 (6) (1974) 915-930.
[30] Sueyoshi, T., Mika, G., “Returns to Scale vs. Damages to Scale in Data Envelopment Analysis: An Impact of U.S. Clean Air Act on Coal-Fired Power Plants”, *Omega*, 41 (2) (2013) 164-175.
[31] Tsang, S.S., Chen, Y.F., “Facilitating Benchmarking with Strategic Grouping and Data Envelopment Analysis: The Case of International Tourist Hotels in Taiwan”, *Asia Pacific Journal of Tourism Research*, 18 (5) (2012) 1-16.
[32] Yeh C.C., Dhi D.J., Hsu M.F., “A Hybrid Approach of DEA, Rough Set and Support Vector Machine for Business Failure Prediction”, *Expert Systems with Application*, 37 (2010) 1535-1541.
[33] Zak, T., Huang, C., Siegfried, J., “Production Efficiency: The Case of Professional Basketball”, *The Journal of Business*, 52 (3) (1979) 379-392.
[34] Zhang W, Tang S.Y., Zhu Y.F., Wang W.P., “Comparative Studies of Support Vector Regression between Reproducing Kernel and Gaussian Kernel”, World Academy of Science, Engineering and Technology, 65 (2010) 933-941.