A data-driven approach to predicting the most valuable player in a game

Francisco P. Romero¹ | Catalina Lozano-Murcia¹ | Julio A. Lopez-Gomez¹ | Eusebio Angulo Sanchez-Herrera² | Eduardo Sanchez-Lopez¹

¹Department of Information Technologies and Systems, University of Castilla La Mancha, Ciudad Real, Spain
²Department of Mathematics, University of Castilla La Mancha, Ciudad Real, Spain

Abstract

The identification of outstanding behaviors is a matter of essential importance in sports analytics. However, analyzing how human experts select each match’s most valuable player (MVP) according to objective and subjective factors is a great challenge. This article proposes a data-driven approach for sports team performance based on the weighted aggregation of statistical indicators. The proposal is divided into two approaches: The first conducts a principal component analysis to examine the relationship between each game’s statistical indicators. The other addresses a meta-heuristic analysis to weight the attributes and choose the MVPs optimally. Finally, we apply the proposed approach to the 2018 European Men’s Handball Championship and take the “Player of the Match” of each game as an example to illustrate its usefulness and efficacy. We perform multiple analyses, including a comparison with a fuzzy multi-criteria decision-making method that show that the data-driven approach can predict the “Player of the Match” in most matches. It also allows us to estimate and quantify the expert evaluations, which are often difficult to obtain in a disaggregated form.

KEYWORDS

handball, meta-heuristics, most valuable player, optimization, principal component analysis, sports analytics

1 | INTRODUCTION

In sports, a most valuable player (MVP) award is an honor typically bestowed upon an individual as the best performing player (or players) in a match, in a tournament or on a specific team. Other terms, like “player of the year,” are used for a season-spanning award and “player of the match” for individual games. The notion of a player’s value is also relevant to other tasks in which professional teams engage, for example, drafting, paying, and trading players.

The concept of MVP is fundamentally vague. However, in the context of professional sports, such a vague category is valuable to the extent that it encourages active discussion of the different indicators of excellence found within a specific sport and the weight that should be assigned to each of them.

A panel of sportswriters and Broadcasters commonly decides MVP in a competition. There are no official criteria for being voted the MVP of a match. The criteria are whatever the voter decides. Many factors come into play, and these factors change from expert to expert. As a result, many observers feel they do not select the most appropriate players. Although the voting procedure has its drawbacks, it serves as a consistent metric for how the public perceives player performance.
and value across the game. There are several angles that might be taken when discussing the merits of MVPs. Statistics provides one of these angles.

The most important problem to deal with is how to model expert judgment using a data-driven approach. Judgment is defined as, “the ability to make considered decisions or come to sensible conclusions.” We consider that judgment implies using information and knowledge to produce an assessment or evaluation. However, judgments are subjective, and subjectivity is a topic that has generated heated debate in the data science community. Expert choices are subjective in the sense that every expert builds knowledge and judgment into their framework of understanding, which is not likely to be identical to that of anyone else. Nevertheless, expert judgment is needed for accurate statistical and scientific analyses. Moreover, to analyze how human experts judge this situation, due to information overload and other factors, is a significant challenge.

Some machine learning algorithms have been proposed to solve the MVP selection problem. However, there are generally problems with these algorithms, including complex parameter setting, unfaithful statistical modeling, and the limited ability of subjectivity management. For example, Reference 4 describes a mathematical model to predict the winners of an MVP in baseball. This model quantifies the value of several statistics, much as a sportswriter would when determining whom they should select as MVP. They then created a scale from 0 to 10 to determine how likely a specific player was to win the award.

The purpose of this article is to develop a data-driven approach for determining the MVP of a handball match. We believe that this approach lends itself to a more appropriate representation of the effect the candidate for the award has on team wins. We would also like to determine how often the experts select the player who best meets the criteria, and to discuss which indicators could be considered relevant. Our results provide significant insights into the features that distinguish an outstanding handball performance in a short tournament from those of the other players. In consequence, the main contributions of this article lie in the fact that we address a number of complicated matters not previously considered in the literature: (1) Definition of the problem of choosing the MVP in a data-driven manner; (2) comparison of different strategies (principal component analysis (PCA), meta-heuristics, fuzzy multi-criteria decision-making method ) for weighting performance criteria; and (3) comparison with the selection of the MVP in a tournament.

This article is organized as follows: Section 2 sets out some related approaches to solving the MVP problem. In Section 3, we explain the different alternatives for establishing the weights of the evaluation criteria: PCA and meta-heuristics. Section 4 presents our case study for assessing the performance of handball players at the 2018 European Men’s Handball Championship. Conclusions and future work are outlined in Section 5.

2 RELATED WORK

Some models have been proposed to address the MVP question in several contexts; the problem is deeply analyzed in References 1 and 2. The same author suggests that the problem remains unsolved because the most suitable model must focus on player-related value, and player value is indeterminate by nature.

The application of PCA in the selection of key performance indicators in sports is not new. For example, in Reference 7, it is shown how PCA can be used to extract a few abstract key performance indicators from a large set of performance indicators by using a local combination of them. In this case, our purpose is not to build a new set of indicators from the pre-existing set, but to give the appropriate weights to a limited number of them.

Multiple criteria decision-making (MCDM) methods have been shown to be very useful for the challenge of MVP selection in some contexts. For example, Reference 8 uses the analytic hierarchy process (AHP) to determine the weight of each decision criterion, and the technique for order preference by similarity to ideal solution (TOPSIS) is applied to the weighting of decision-makers (DMs) and the ranking of alternatives. Another approach based on the Fuzzy analytic hierarchy process (FAHP) and TOPSIS, for the selection of candidates eligible to become basketball players, is described in Reference 9. Reference 10 proposes a two-phase solution approach for the selection of soccer players. The key performance indicators for each player role are prioritized using AHP; after that, an integer linear programming model is applied to compute the weights of these attributes. On the other hand, some approaches use fuzzy logic-based techniques to handle the subjectivity related to player selection; for example, Reference 11 presents weighted schema based aggregation of expert judgments using fuzzy logic techniques. Expert-based methods for determining objective weights in multiple-criteria problems have many drawbacks in terms of reliability (different conclusions when examining the same data) and bias. The use of mathematical and optimization models allows the weight vector corresponding to the specific
problem to be found, and the given alternative, found by weighted multi-criteria evaluation, is then selected as Player of the Match.

A method for applying the MVP approach to solve problems is presented in Reference 12. This algorithm is based on the dynamic whereby the players, grouped by teams, fight together for a joint goal (championship), but also for a personal goal such as being the tournament’s MVP. This algorithm does not intend to find the MVP in a given match but uses the philosophy within the MVP selection problem to solve the optimization problem.

3 | PROPOSAL

An ideal automatic MVP selector for use in practice is required to be simple and easy to use and interpret. Exploiting the additive effects of performance indicators, the problem presents as a weighted sum model problems (WSM). This is one of the best known and simplest multi-criteria decision-making methods for evaluating several alternatives in many decision criteria.

3.1 | Statement of problem

In this section, the problem statement is introduced, and a mathematical model suitable for resolution purposes formulated. Considering Figure 1, we have a sequence of time events (matches) that follow each other in parallel or in sequence. A data matrix is then generated for each time event. This data matrix contains the evaluation of different elements from a set of behavior indicators. Only one of these elements (players) is highlighted by an expert or a panel of experts. The problem is to find out why these elements are selected and to create a data-driven method to detect them in new time events.

A set of features (performance indicators) characterizes each item. These features have different effects on the highlighting process: positive, negative, and neutral. Neutral does not mean no contribution to the result; it only means that we do not know previously if the feature’s contribution is positive or negative with regard to the selection. They are also some relations between them, and there can be additive effects/ratios. The objective is to identify those features that allow the specific element (player) to be highlighted within the event (match) and a prediction model to be built that, given a new fact, can choose the highlighted item.

This kind of problem arises in several fields, such as cybersecurity, fraud detection, or sports analytics. As previously stated, this article provides a data-driven approach to identifying the most valuable element for each time event, that is,
the MVP, for each match. We assume that only one player is highlighted by the expert, in a subjective manner but based on a set of quantitative attributes.

### 3.2 Mathematical preliminaries

Given a dataset $D$ of $s$ time events $t_1, t_2, \ldots, t_s$, where each time event is a multidimensional vector of $m$ elements, $p_1, p_2, \ldots, p_m$, each element has $n$ attributes. Each attribute makes a positive, negative, or neutral contribution to the behavior. Thus, a sign vector is associated with the set of attributes: $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ where $\sigma_i \in \{+, -, 0\}$. For each time event, only one element is highlighted as being the most outstanding. We would like to find a model to identify this element according to the attribute values.

The problem consists of selecting the MVP from among $m$ players and $n$ performance criteria. Then, $w_j$ denotes the relative weight of importance of the criterion $C_j$ and $c_{ij}$ is the performance value of player $p_i$ when evaluated in terms of criterion $C_j$. The total (i.e., when all the criteria are considered simultaneously) score of player $P_i$, denoted as $P_i^{\text{WSM-score}}$, is defined as shown in Equation (1):

$$P_i^{\text{score}} = \sum_{j=1}^{n} w_j c_{ij} \quad \text{for} \quad i = 1, 2, 3, \ldots, m.$$  (1)

Because the weights of the criteria can influence the outcome of the decision-making process, it is essential to pay particular attention to objectivity factors in attribute weights. Determining attribute weights is a problem that frequently arises in multi-criteria decision-making (MCDM) problems. After the performance indicators and categories are identified, the next step is to determine the relative weight associated with each indicator in its category. This study uses two approaches to designate each criterion’s weight: PCA and meta-heuristics optimization.

### 3.3 Principal component analysis

Our aim is to produce a score composed of multiple input variables (performance indicators) as decision criteria. Our data-driven model consists of aggregating the attributes of the events (performance indicators) according to PCA loadings. PCA is a statistical technique that transforms a dataset defined by possibly correlated variables into a set of uncorrelated variables, called principal components (PC). Each of the PC is a linear combination of the original variables, ordered according to the amount of variance, as described in Reference 14. Therefore, each PC could be seen as a weighted sum of all variables after the input variables are collected and scaled. In Equation (2), the first PC (and the remainder) was the sum of all input variables, denoted by $c_i$, weighted by PC-specific loadings, denoted by $L_i$.

$$PC_i = \sum_{i=1}^{N} L_i * c_i$$  (2)

### 3.4 Meta-heuristics

The process of choosing the best player for each match can also be addressed as an optimization problem. In this way, given the evaluations provided by experts, in which they identify the players in the match, we can approximate the position of the MVP in the ranking of the players involved. To do this, we define an optimization problem to optimize the weights of the indicator set to obtain a ranking where the best player occupies the first position.

Meta-heuristic algorithms were used to solve the above-mentioned problem. Meta-heuristics are algorithms that are designed to solve a broad range of optimization problems, guiding the search process, and trying to explore the whole search space (see References 15 and 16). In comparison to heuristics, they are at a higher level of abstraction, incorporating and controlling different heuristics and adding mechanisms to escape from local optima and reach the global optimum. Specifically, particle swarm optimization (PSO) has been used to solve these problems (see Reference 17). It is a state-of-the-art population-based meta-heuristic inspired by bird flocking. PSO starts by initializing a population of
TABLE 1  Selected criteria

| Code | Indicator | Description               | Pos./Neg. | Code | Indicator | Description          | Pos./Neg. |
|------|-----------|---------------------------|-----------|------|-----------|----------------------|-----------|
| N1   | 7mMiss    | 7-meter Shots Missed      | -         | P1   | 7mGoals   | 7-meter Goals        | +         |
| N2   | 6mCMiss   | 6-meter Shots Missed      | -         | P2   | 6mCGoals  | 6-meter Goals        | +         |
| N3   | WingMiss  | Wing Shots Missed         | -         | P3   | WingGoals | Wing Goals           | +         |
| N4   | BTMiss    | Breakthrough Missed       | -         | P4   | BTGoals   | Breakthrough Goals   | +         |
| N5   | FBMiss    | Fast Break Missed         | -         | P5   | FBGoals   | Fast Break Goal      | +         |
| N6   | FTOMiss   | Fast Throw off Missed     | -         | P6   | FTOGoals  | Fast Throw off Goal  | +         |
| N7   | 9mMiss    | 9-meter Shots Missed      | -         | P7   | 9mGoals   | 9-meter Goals        | +         |
| N8   | Excl      |                           | -         | N9   | Prov7     | Provoke 7m           | +         |
| A1   | Time      | Time Played               | =         |      |           |                      |           |

particles. These are the initial solutions or initial weights for each indicator from the indicator set which will be optimized during execution of PSO. Each solution has attributes of position and speed, which are initialized first. Then the particles start to move, and update their positions and speeds according to the positions that the particle has traveled through and the best position in the swarm. These rules guarantee an excellent trade-off between exploration and exploitation and ensure the algorithm’s convergence (see References 18 and 19). The specific expressions for updating each particle’s position and speed depend on the particular variant or PSO implementation used.

4  | EXPERIMENTS

The proposed data-driven approach was applied to data collected from the Official Statistics of Handball Euro 2018.* This section contains explanations about the dataset and the performance metrics used. These experiments compare the PCA, the meta-heuristic approach, and the application of a fuzzy method to weight criteria based on expert subjectivity previously used in the literature for this purpose.11

4.1  | Datasets

We describe a performance evaluation based on a realistic scenario using the selection of “Most Valuable Player” in the 2018 European Men’s Handball Championship. The data from multiple PDFs was pre-processed using data fusion algorithms. We generate one data sample for each match with statistical information and MVP selection. Thus, in this real-world experiment, we analyzed about 1500 records distributed over 47 matches (events). We discard some due to data quality problems, leaving some 1187 records. Each event usually has about 31 or 32 records, and the experts evaluate only one of them as the MVP. Attributes (Table 1) are divided into those which contribute positively to highlight the behavior, and those whose contribution is negative.

Moreover, some features are neutral a priori, and others are characteristics created by adding or subtracting from the rest. This knowledge is relevant to solving the problem. The analysis is limited to outfield players, and does not include goalkeepers. Although some goalkeepers are considered for the MVP award, it is not easy to compare performances with different attributes to be evaluated.

4.2  | Performance metrics

The evaluation provided in this article is not only a quantitative verification, it is also qualitative. The objective is that the proposed methods identify the MVP of each match and do so by defining useful weights for handball. Therefore, we need to answer three research questions:

*http://activities.eurohandball.com/analyses
• **RQ1** - *Is it possible to identify the MVP of each match?* We will use as metrics the percentage of MVPs correctly identified by the method for this purpose. (Accuracy).

• **RQ2** - *What is the rank of the MVP in a match according to the method?* With this aim, we will use the Mean Reciprocal Rank (MRR) according to the rank assigned to the MVP of the match (explained above).

• **RQ3** - *Can the weights obtained be interpreted from a handball expert’s point of view?* For this purpose, an expert analyzes the obtained weights to identify those weights that are correct for understanding the game (Interpretability).

The mean reciprocal rank (MRR) tries to answer the question *What is the rank of the MVP in a match?*. More formally, MRR calculates the reciprocal of the rank at which the first relevant element was retrieved (see Reference 20). Reciprocal Rank is 1 if the MVP was retrieved at rank 1; if not, it is 0.5 if an MVP was retrieved at rank 2, and so on. Thus, given a dataset $D$ where best is the player ranked first in each match, the aim is to obtain a set of weights for the indicators from the indicator set, to build a ranking where best ranks first. Formally, it is computed as the inverse of the best player’s position in the ranking built according to Equation (3).

$$MRR = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\text{rank}_{\text{best}}}$$

where $n$ is the number of times the ranking is built or the number of repetitions of the experiment and $\text{rank}_{\text{best}}$ is the position occupied in the ranking by the best player. The MRR function is bounded between 0 and 1. Therefore, when the experiment is formulated as an optimization problem, the aim is the following: given a dataset $D$, the best player best, and a fixed number of experimental repetitions $r$, the aim is to obtain a set of weights which maximizes MRR. This means the best player will rank first in each repetition’s ranking built by the optimization algorithm. This problem is formulated through Equation (4).

$$\text{Maximise } MRR(D, \text{best}, r)$$

### 4.3 Results

In this section, we describe the experiments that were carried out to analyze the results provided by the previously described approaches. The meta-heuristic experiment was programmed using the MATLAB programming language. The different optimization techniques were applied using the PSO MATLAB Library. On the other hand, the PCA experiment was programmed in R using the stats package and the fuzzy method implemented in Python using the scikit-fuzzy library. Both experiments were run on a computer server with an AMD FX-8370 Eight-Core-Processor operating up to 4.00 GHz.

#### 4.3.1 Principal component analysis

Considering PCA as a technique to assign weights to performance indicators, the procedures to aggregate these indicators are PCA loadings and a PCA-based method to generate the players’ score statistically (Figure 2). Figure 3 illustrates the weights obtained using this strategy. All of the indicators have unequal weights. This contrasts with the usual practice of assigning equal weights to indicators. Using equal weights in most indices is not the best method for obtaining the score in this study. Even so, if the indicators are highly correlated, summing them with equal weights maximizes the variance of the scores. As a result, there were six PCs with eigenvalues of 1.0 or greater. These accounted for 74.8% of the variance in the data (see Figure 2 where the X-axis represents the components and the Y-axis represents the proportion of variance).

The results obtained are reasonable when evaluating the player’s performance, although they show some inconsistencies, the most remarkable are: (1) 6-meter goals have very little value (0.04) compared to the rest. (2) Red cards minimally penalize the valuation. These two types of actions may have these weights because 6-meter goals have a low percentage of failures and because Red Cards are very few in high competition. Therefore, these two parameters do not differentiate between players. The other parameters have reasonable weights, subtracting those that hurt the team’s result, and vice
versa if they bring the team closer to winning the match. In conclusion, the PCA approach has low noise sensitivity and gives advantageous weights for evaluating player performance. However, the results predicted are not satisfactory; only about 65% of MVPs are correctly identified.

4.3.2 Meta-heuristics

The problem addressed in this study consists of maximizing the so-called reciprocal-rank metric. Hence, the aim is to obtain a set of weights to rank the MVP first in the ranking. Figure 4 illustrates the results obtained using this technique. The results are encouraging, and it seems that the approach might work. There are 28 matches (76% accuracy) in which the MVP is found (reciprocal Rank value = 1). On the other hand, there are three matches where the reciprocal rank value is 0.5 and five matches where the reciprocal rank is 0.33. The main concern with these results is the difficulty in interpreting the weights obtained from the handball perspective. The resulting weights from the evaluation of handball performance and the difficulty of the actions considered show several inconsistencies: (1) The 6m shots add 1.67 if they result in a goal and subtract 0.84 if they miss, when the % of efficiency should be high for this type of shot. (2) Throws from the Wings subtract very little if they fail (0.11) and add a lot (1.66) if they are goals. In this way, with very low percentages, some Wing shots would have positive scoring. (3) In 9-meter shots, something similar happens; the penalty for missing should be higher (is only 0.18) since a 9-meter goal adds 1.79.
4.4 Results and discussion

In this section, we show the results obtained in the analysis of the Players of the Match of the 2018 European Men’s Handball Championship. We compare the PCA, the Meta-heuristic approach, and the application of a fuzzy method to weight criteria based on the subjectivity of experts. The resulting weights obtained using this fuzzy approach are shown in Figure 5.

The best interpretation technique is the fuzzy method (Figure 6 - gray). This is reasonable because it is based on a method to weight criteria based on experts opinions. However, the fuzzy method is not able to represent the quantitative part of the phenomena and therefore the results according to accuracy and MRR are the worst. The PCA (Figure 6 - blue) offers results that are better in interpretability, and slightly worse in accuracy and MRR, than the Meta-heuristic approach (Figure 6 - orange). The optimization approach has a better ability to rank and to identify the MVP; however, the lack of interpretability of the weights obtained reduces the possibility of its application in a real-world context. The results are so low ([65–75]% ) that they lead to concern about how the data was gathered and whether the results are meaningful. Nevertheless, as Reference 5 said, the MVP selections matched the objective criterion just under half the time, and the sportswriters selected one of the top three performers nearly 70% of the time. Therefore, our results contribute to understanding the process of selecting the “Player of the Match” in handball tournaments.
There are several reasons why the sport-writers’ picks for this award disagree with those computed by the data-driven method described in this article:

• Reputation: A player’s reputation has some effect on the voting for that award. Reputation is not considered in our analysis.

• Soft skills: It is possible that hard to quantify skills which were not included in our analysis, such as team leadership, could have a noticeable effect.

• Similarity: Difficulty in distinguishing two similarly qualified players who have different strengths and weaknesses, that is, two players could be dramatically different in terms of skills they offer their team, and it is hard to compare their values using head-to-head statistics.

5 | CONCLUSIONS

The selection of the “Player of the Match” or the “Most Valuable Player” is a very complicated procedure that focuses on examining several different criteria. This article’s main contribution is to show how to use intelligent techniques, such as PCA and meta-heuristics, to compute a score for each player and then use this score to pick the best player for each match. Based on the 2018 European Men’s Handball Championship, the evaluation scenario is designed to compare several intelligent techniques in a real tournament. This experiment shows both techniques, PCA and meta-heuristic based methods, performing acceptably in terms of identifying and ranking the best handball players. It implies that we can obtain the criteria weights to be used to identify and rank the best players by using these methods. This approach also allows for estimating and quantifying the expert evaluations, which are often difficult to obtain in a disaggregated form.

The solution of the “MVP problem” returns interesting results: on the one hand, it is possible to analyze the subjectivity introduced by the experts, allowing them to determine which indicators had more weight in their judgment. On the other hand, experts will have feedback to assess their criteria and improve them, since they usually evaluate the players according to qualitative criteria and opinions. In future research, it would be interesting to consider (i) how the players’ reputation affects MVP selection, (ii) what type of soft skills might be most useful to be an MVP, and (iii) what physical and technical factors affect the outstanding performance of an MVP.

ORCID
Francisco P. Romero © https://orcid.org/0000-0002-6993-2434

REFERENCES
1. Kershnar S. Solving the most valuable player problem. J Soc Philos. 2008;02(39):141-159.
2. Kershnar S, Feit N. The most valuable player. J Philos Sport. 2001;28(2):193-206.
3. Brownstein NC, Louis TA, O’Hagan A, Pendergast J. The role of expert judgment in statistical inference and evidence-based decision-making. Am Stat. 2019;73(Supp 1):56-68.
4. Sparks RL, Abrahamson DL. A mathematical model to predict award winners. *Math Horizons*. 2005;12(4):5-13.

5. Fritz K, Bukiet B. Objective method for determining the most valuable player in major league Baseball. *Int J Perform Anal Sport*. 2010;10(2):152-169.

6. Kershnar S. The most-valuable-player problem remains unsolved. *J Philos Sport*. 2011;38(2):167-174.

7. O’Donoghue P. Principal components analysis in the selection of key performance indicators in sport. *Int J Perform Anal Sport*. 2008;8(3):145-155. https://doi.org/10.1080/24748668.2008.11868456.

8. Nikjo B, Rezaeian J, Javadian N. Decision making in best player selection: an integrated approach with AHP and extended TOPSIS methods based on WeFA framework in MAGDM problems. *Int J Res Ind Eng*. 2015;4(1):1-14.

9. Ball S, Korukoğlu S. Development of a fuzzy decision support framework for complex multi-attribute decision problems: a case study for the selection of skilled basketball players. *Expert Syst*. 2014;31(1):56-69.

10. Romero FP, Angulo E, Serrano-Guerrero J, Olivas JA. A mathematical model using AHP priorities for soccer player selection: a case study. *South African J Ind Eng*. 2016;27(2):190-205.

11. Bouchekara H. Most valuable player algorithm: a novel optimization algorithm inspired from sport. *Oper Res*. 2020;20(1):139-195.

12. Chao YS, Wu CJ. Principal component-based weighted indices and a framework to evaluate indices: results from the medical expenditure panel survey 1996 to 2011. *PLoS One*. 2017;12(9):e0183997.

13. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. Springer Series in Statistics New York. Vol 1. New York, NY: Springer; 2001.

14. Talbi EG. A taxonomy of hybrid metaheuristics. *J Heuristics*. 2002;8:541-564.

15. Boussaïd I, Lepagnot J, Siarry P. A survey on optimization metaheuristics. *Inf Sci*. 2013;7(237):82-117.

16. Shi Y, Eberhart RC. A modified particle swarm optimizer. Paper presented at: Proceedings of the IEEE International Conference on Evolutionary Computation; 1998:69-73; IEEE Computer Society, Washington, DC.

17. Jiang S, Ji Z, Shen Y. A novel hybrid particle swarm optimization and gravitational search algorithm for solving economic emission load dispatch problems with various practical constraints. *Int J Electr Power Energy Syst*. 2014;55:628-644.

18. Samma H, Lim CP, Saleh JM. A new reinforcement learning-based memetic particle swarm optimizer. *Appl Soft Comput*. 2016;43:276-297.

19. Craswell N. Mean reciprocal rank. In: Liu L, Özsu MT, eds. *Encyclopedia of Database Systems*. Mean Reciprocal Rank. New York, NY: Springer; 2018:2217-2217.

20. Warner J, Sexauer J. Scikit fuzzy, twmeggs, AMS. A. Unnikrishnan, G. Castelão, F. Batista, TG Badger and H. Mishra, JD Warner/scikitfuzzy: Scikit-Fuzzy 0.3 1; 2017.

**AUTHOR BIOGRAPHIES**

Francisco P. Romero was born in Ciudad Real, Spain in 1978. He received the B.S. and M.S. degrees in Computer Science from the University of Castilla La Mancha, in 2001 and the Ph.D. degree in Advanced Computer Science Technologies from University of Castilla La Mancha in 2008. From 2001 to 2008, he was a Project Manager with the Indra Software Labs. Since 2005, he has been an Assistant Professor with the Information Technologies and Systems Department of the University of Castilla La Mancha. He is the author of more than 50 conference communications and 20 journal articles. His research interests include machine learning application in complex environments, information retrieval, clinical decision support systems. He is a member of the Editorial Board of Information Processing and Management Journal. Dr. Romero is now the Head of Information Technologies and Systems Department of the University of Castilla La Mancha.

Catalina Lozano-Murcia was born in Bogotá, Colombia in 1990. She received the degree in Economics from the Universidad Nacional de Colombia, in 2012, then she obtained the MSc in Actuarial and Financial Science from Universidad Carlos III de Madrid in 2015, and MS in Big Data and Data Science from Universidad Internacional de Valencia in 2018. Now she is a PhD student in Advanced Computer Technologies from Universidad de Castilla La Mancha. From 2012 to 2013, she was a research economist in CIDEI, was an internship in Agroseguro in Spain since 2014 to 2015, and from 2016 to nowadays she has worked as actuary in Deloitte Colombia. She is an Associate Professor in the Ms Actuarial from Escuela Colombiana de Ingenieros. Her research interests include actuarial sciences, data sciences apply in risk modelling and financial models.
Julio Alberto Lopez-Gomez was born in Bolaños de Calatrava, Ciudad Real, Spain in 1993. He received the B.S. degree in computer engineering with mention in computer science from the University of Castilla la Mancha, Spain, in 2015, M.S. degree in data science and computer engineering with mention in data science and intelligent technologies from the University of Granada, Spain, in 2016 and the Ph.D. degree in advanced information technologies from University of Castilla la Mancha, Spain, in 2019. From 2015-2020, he has been a researcher in different regional and national projects publicly funded. Since 2017, he is also Assistant Professor with the Information Technologies and Systems, University of Castilla la Mancha. He is the author of some twenty articles published in national and international conferences as well as high impact journals. His research interests include optimization using metaheuristic algorithms, artificial intelligence, machine learning, neural networks and fuzzy logic.

Eusebio Angulo Sánchez-Herrera was born in Pozuelo de Calatrava, Ciudad Real, Spain, in 1983. He received the B.S. degrees in computer engineering from the University of Castilla-La Mancha in 2007, M.S. and the Ph.D. in Advanced Information Technology from the University of Castilla-La Mancha in 2009 and 2012 respectively. From 2005 to 2011, he was a Research Assistant of Models and Algorithms for Transportation Systems (MAT) research group. Since 2011, he has been an Professor with the Mathematics Department, Castilla-La Mancha University, in the University College of Computer Science. He is the author of 14 journal article indexed in the JCR ranking, has involved in 6 regional and national projects and has participated in more than 40 national and international research conferences. His research interests include heuristic algorithms, operations research, statistics and big data applying to sports analytics and transport.

Eduardo Sanchez-Lopez was born in Ciudad Real. He received the B.S. degrees in computer engineering from the University of Castilla-La Mancha in 2018. Currently, he is Data Engineer at Wenea Mobile Energy, and previously he worked at Telefonica in the Quality Assurance Team and as a Data Scientist Intern in Prometeus Global Solutions. His research interests include the application of the most recent deep learning and machine learning techniques to solve real world problems.

How to cite this article: Romero FP, Lozano-Murcia C, Lopez-Gomez JA, Angulo Sanchez-Herrera E, Sanchez-Lopez E. A data-driven approach to predicting the most valuable player in a game. Comp and Math Methods. 2021;3:e1155. https://doi.org/10.1002/cmm4.1155