Information Systems 93 (2020) 101562

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

Sports analytics — Evaluation of basketball players and team performance

Vangelis Sarlis, Christos Tjortjis *

The Data Mining & Analytics Research Group, School of Science and Technology, International Hellenic University, 14th km Thessaloniki - Moudania, 57001 Thermi, Greece

ARTICLE INFO

Article history:
Received 20 February 2020
Received in revised form 19 May 2020
Accepted 20 May 2020
Available online 23 May 2020
Recommended by D. Shasha

Keywords:
Sports Data Mining (DM)
Sports analytics
Data analysis
Business intelligence
Statistics
Machine learning

ABSTRACT

Given the recent trend in Data Science (DS) and Sports Analytics, an opportunity has arisen for utilizing Machine Learning (ML) and Data Mining (DM) techniques in sports. This paper reviews background and advanced basketball metrics used in National Basketball Association (NBA) and Euroleague games. The purpose of this paper is to benchmark existing performance analytics used in the literature for evaluating teams and players. Basketball is a sport that requires full set enumeration of parameters in order to understand the game in depth and analyze the strategy and decisions by minimizing unpredictability. This research provides valuable information for team and player performance basketball analytics to be used for better understanding of the game. Furthermore, these analytics can be used for team composition, athlete career improvement and assessing how this could be materialized for future predictions. Hence, critical analysis of these metrics are valuable tools for domain experts and decision makers to understand the strengths and weaknesses in the game, to better evaluate opponent teams, to see how to optimize performance indicators, to use them for team and player forecasting and finally to make better choices for team composition.

1. Introduction

Players' performance prediction by using current and past data has gained attention, particularly in basketball [1,2]. Sports analytics and forecasting through these data is a rapid growing field with many methods that can be implemented from a different perspective for each situation [3]. In a team, and specifically for the technical staff and coaches, the knowledge of advantages and disadvantages for each player can give an added value in roster composition, in new transfers, in changing the rhythm during a match and other vital qualitative and quantitative factors [4]. The aforesaid Performance Analytics are critically valuable for a team, in order to minimize budget costs, maximize team value and improve the processes in all layers and segments of the flow [5].

In addition, many teams and countries invest large amounts of money to train athletes who can win competitions, Olympic Games etc. Over the last few years, basketball analytics began to have bigger traction and attempt to analyze the game in more depth by finding advanced analytics to optimize team and player performance [6]. New technological findings can give the opportunity to collect more data and requiring new methods of analysis to be performed. Therefore, the new analysis methods could exploit and generate these added values for defining the basketball players behavior and help the technical staff and coaches in better decision making [7].

Generally, sports data are irregular and sparse. They are sparse because the majority of the players do not have long careers, and do not remain in the same league and/or team for many years. The data are not regular because the career of each player belongs to different chronological periods [8]. The big variety of sports data such as shots attempted, fouls committed, the defense metrics during the game and the kilometers they run, and many other parameters of a game can be tracked with the use of SportVU cameras. However, it is significantly difficult to distinguish the dominant performance analytics of each team/player in comparison with the opponents' performance as well. There are outlier factors, such as the psychological or physical condition of each player/team that can be analyzed and give an extra important information for decision making [9]. These are recognized electronic devices named Electronic Performance and Tracking Systems (EPTS) that can measure all these additional data through the gyroscopes, magnetometers and accelerometers sensors that provide opportunities to explore all these significant aspects in more depth [10].

This research attempts to gather all the proper analytics used in sports as state-of-the-art performance indicators through sports data in decision making for basketball games, teams and

* Corresponding author.
E-mail addresses: e.sarlis@ihu.edu.gr (V. Sarlis), c.tjortjis@ihu.edu.gr (C. Tjortjis).

https://doi.org/10.1016/j.is.2020.101562
0306-4379/

players. Data mining is looking for unknown structure and data analytics [11,12]. Hence, this could help decision-making and predict uncertain data [13]. In conclusion, sports analytics could be very helpful to educate the next generation of players, technical staff and managers. The teams can take advantage of them for future prediction of roster composition, optimization of tactics and avoiding unexpected circumstances [14,15].

The structure of this paper follows:

2. ***Background***: this section illustrates some important concepts and terminologies of sports analytics and the literature.
3. ***Methodology — Research Design:*** Problem definition, important remarks, existing sports analytics and applied research algorithms.
4. ***Findings:*** Providing the comparison of existing advanced metrics for teams and players performance.
5. ***Discussion, Case Study & Forecasting Scenario:*** In this section we present and discuss results and observations. Additionally, a clear comparison over the existing and historical basketball analytics is made. Furthermore, a case study is provided in order to explain in more systematic way basketball analytics presented in previous sections. In the forecasting scenario section, we introduce two prediction formulas for the MVP and Defender of the year in basketball.
6. ***Conclusion & Future Work:*** The intention to evaluate related performance analytics that they or their expansion can be applied to different domains of sports analytics. Based on them, directions for future work are proposed.

Appendix: Categorization of important basketball analytics based on different factors of basketball advanced statistics.

2. Background

2.1. Terminology

This section focuses on key terms required to understand this paper.

Sports Analytics

The term "Sports Analytics" also referred as "Statistics in Sports" in the bibliography comprises the segment of data collection and management, predictive modeling and computational methods in order to find valuable information for sport related decision making [16].

Alternatively, Sports Analytics is a scientific field that deals with the collection and analysis of past and current sports data [17]. This collection combines and applies methods that can give an added value to a player or a team. Through this gathering and investigation these metrics can give a qualitative analysis to owners, players, coaches and team staff in order to help them to predict future situations or make suitable decisions.

Sports data can be either qualitative or quantitative and come from different sources such as boxscores, videos, demographical, medical and scouting reports. Data collection should be standardized, integrated and analyzed through different basketball analytics in order to enable decision makers to take critical decisions [18,19].

One of the most significant topics in sports analytics is the identification of performance analytics in teams and players. By analyzing them there is direct impact on teams composition, on players evaluation and decision making of subject matter experts, coaches, and technical staff [20].

Recently, there is a highly increasing trend in sports analytics, recognized as one of the hottest topics of analytics in general. There are many web pages and blog articles, but also a scarcity in credible peer reviewed research articles. The challenge of Sports Analytics is that domain experts need to combine the scientific research experience with sports professional career (as player or coach) and understand how to critically analyze important sophisticated basketball analytics [21].

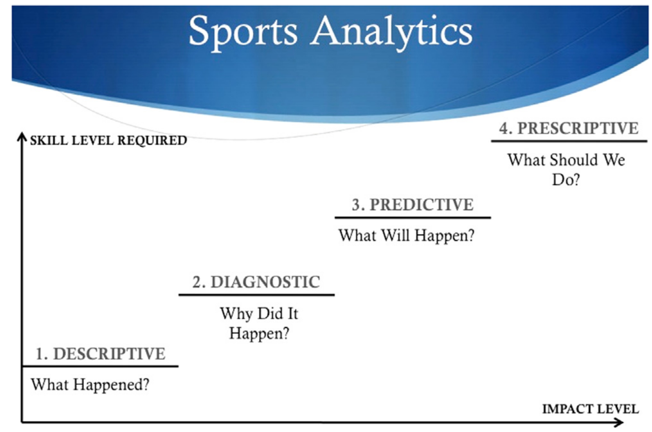

Fig. 1. Sports Analytics Skills vs. Impact correlation.

The combination and application of appropriate tactics can give an added value in a player or a team in order to implement this critical initiative to make the right decision in the right time. Through the proper data gathering and data analysis, basketball analytics can provide qualitative analysis to team owners, players, coaches, and technical staff to help them predict future situations and make the right decisions to improve their performance [22].

The required skillset and the influence in each sport is not only to describe and provide the data but to diagnose, forecast and make critical reasoning.

Fig. 1 illustrates the correlation between the influence of sports analytics based on the required skills. There is a recent trend in sports analytics to move from descriptive (sports reports) to prescriptive analytics with the purpose to understand the game in more depth. By using data modeling techniques and optimizations the technical team and subject matters can notice valuable insights and recommendations [23]. Sports analytics is also increasingly adopted by the business ecosystem in a plethora of companies, sports segments, teams, technical staff and athletes [24].

Data Mining and Sports

Data Mining is the discovery of patterns or rules from large amounts of data, as well as the process of searching for valuable information in data [11]. Therefore, it is the engagement of one or more techniques into automated analysis and knowledge data extraction. Furthermore, it is the process of data analysis in the examined datasets with purpose to solve the described problems. Sports teams use data mining methodologies either for interpretation or segmentation purposes that will finally help them in decision making. Assembling DM techniques and important information can boost a team and give a competitive advantage.

The classification of individual players or teams based on their performance can show different perceptions or ways of play. Once the preferences in each position have been decided, gameplay is set, the managers and coaching staff finally can understand, drill down and analyze insights to choose the best option for each situation. Through these advanced basketball analytics, this sophisticated approach can be personalized for each team/player preference or performance [25]. With the implementation of this concepts we can automate such procedures for optimized classification, segmentation and forecasting.

There are some sophisticated algorithms that can be used in sports analytics listed below:

(1) **Random Forest** are classification and regression models [26,27] and [28].
(2) **Adaboost** is a classification algorithm that can be extended for regression [29–32] and [33]
(3) **Multilayer Perceptron (MLP)** is a network of perceptrons, which are neurons with multiple inputs and one output [11,32] and [29].
(4) **Radial Basis Function (RBF) Networks** is a class of functions with the value increases\decreases based on distance of central point [34–36] and [37].
(5) **Association Rule based models** are algorithms that use techniques to extract relationships from hidden items through different datasets. They discussed in [13,29,34,38–40] and [41].

2.2. Literature review

DM and ML algorithms used in sports

We evaluate here multiple modeling approaches combined, in order to provide the optimum performance rating accuracy for players and technical staff. Past research reported in the literature, have used classification methods, clustering or both. The state-of-the-art algorithms used in sports analytics based on the bibliography include the following:

(1) **Neural Networks** can be used for both classification and prediction purposes (deep learning, dropout) as stated in [25,26,30,34,42] and [29].
(2) **Decision Trees** are predictive models as described in [6,9, 23,31,37,43–45] and [46].
(3) **Bayesian Networks** are probabilistic classifiers (Markov Blanket) as explained in [16,30,32,38,47] and [33].
(4) **Support Vector Machines** (SVM) are classifier and regression analysis algorithms as referred in [32,34,37,48,49] and [30].
(5) **Linear and Logistic Regression** are models for team and player forecasting [8,16,50–52] and [9].
(6) **Unsupervised learning** through clustering is an algorithm for partitioning where the center of each cluster is displayed by the mean value of the objects [20,25,53–58] and [59].

3. Methodology – Research design

Research questions/hypothesis.

1. How does performance evaluation in Basketball Players & Teams happen? (RQ1)
2. How can these ratings, techniques and methodologies be optimized? (RQ2)
3. How to understand the impact of basketball performance analytics and identify the correlations between them? (RQ3)
4. How can we identify the dominant attributes for prediction of Most Valuable Player (MVP) and Defender of the year? (RQ4)

The answers to these questions are critical for the technical staff and coaches in order to take decisions and determine the performance and future player career trajectories with more accuracy [25].

Overall aim. Based on comprehensive analysis, the main aim of this paper is to evaluate the existing performance analytics used in Europe and NBA (in USA) basketball. Therefore, all sophisticated performance game-related analytics allow to distinguish defense, offense, overall, miscellaneous and performance ratings that exist in bibliography are reviewed. In addition, a comparison matrix is provided for these basketball analytics, trends and patterns that may have been overlooked in current bibliography.

The aforesaid analytics could give an added value to a team and can be treated as a competitive advantage [11]. In general, sports include two important variables. The first one is luck and the second is skill. Luck is something random that you cannot predict. The luck percentage differs from sport, league competition and country. For example in NBA, luck is around 35% which is really a lot [60]. Therefore, the objective of this paper is to compare performance basketball analytics with the purpose to increase the understanding of important insights and minimize the possibility of uncertain current or future events [61]. Hence, it is crucial to understand, analyze and forecast the aforementioned statistics to enable meaningful analytics and statements. Another aim of this research is to quantify player performance attributes in order to increase forecasting accuracy [61].

However, due to the complexity of sports and the huge unstructured retrieved data there is lack of specificity and context that through the help of analytics and proper analyses can exploit in more depth valuable information [3]. Performance forecasting for team and players is a common practice used in the sports industry and by betting companies by gathering data from different perspectives, related to training, matches, injury, psychological etc., and used for short/mid/long term predictions [62].

Specific objectives. It is crucial for sports teams to be able to understand the team/player performance and then in the next step to make the proper decisions [63].

An objective of this research, is to review basketball performance analytics used worldwide [8]. The paper also tries to analyze and compare on Euroleague and NBA basketball leagues in order to find useful insights in micro-level during a game and how they can use this information for critical statements [64]. For sports teams generally (especially for managers and coaches), the team roster selection criteria, it is very significant to have an insight and high-level estimation of how players selected for the upcoming season roster will perform [3]. Forecasting models applied to NBA basketball analytics with the purpose of identifying major player performance attributes to predict the future MVP and Defender of the year. In referred basketball analytics there is a clear analysis of existing algorithms used till now and we aim to verify whether these terms can be optimized. One of the objectives of this work was to predict the NBA MVP and the Defender of the year based on the existing basketball statistics (RQ4).

4. Findings

Basketball is a sport that presents a lot of uncertainty [7]. Although, there are plenty of past data, there are no advanced tools that can forecast players' performance. Some useful metrics provided already for each basketball match are the following (RQ1):

- **Plus/Minus (+/−or PM)**: measures the impact of a player in a game (quality and contribution). It is the calculation or difference of points that a team scores versus the opponents scored points [50]. The problem with that metric is that it does not count the matchup between the players.

- **Adjusted Plus Minus (Adj +/−or APM)**: is the player statistic for rating. It calculates the influence of a player with his presence or not in the team line-up [65]. In NBA is the APM is one of the dominate evaluation indicators [66].
- **Real Plus Minus (Real +/− or RPM or RAPM):** Included the Real Plus Minus wins (RPM Wins) and the number of the possession's metrics. The RPM is the net value of ORPM (Offensive Real Plus Minus) and DRPM (Defensive Real Plus Minus) for the estimated on-court influence on team performance.
- **PIPM (Player Impact Plus Minus):** is another version of plus-minus metric that adjusts the box-score value with the luck adjusted plus minus data [67].
- **Player Impact Estimate (PIE):** calculates the overall player's contribution against the total stats in games that they played [68,69].
- **CARMELO:** focus on win forecasting topic based on players statistics and ELO ratings. This model takes into account personal stats of wins and losses [50].
- **Expected Possession Value (EPV):** evaluate and quantify values that makes a player to decide during the game [70–72].
- **Wins Above Replacement (WAR):** reflects a combination of a player's projected playing time and his projected productivity while on the court. It is computed using the BPM variable [64]. It is the same with WORP.
- **Performance Index Rating (PIR)** is used in European Basketball leagues giving a view of player's total performance [24].
- **Game Score (GmSc)** gives attention to any statistic detail of player's box score [8].
- **Net Rating (NetRtg)** is used in NBA for counting a team's point differential per 100 possessions [50].
- **Pythagorean Win Percentage** is an estimation that shows a team's win percentage based on their points for and against [4,73].
- **Player Efficiency Rating (PER)** is a per-minute rating. PER sums up all a positive action of players, deducts the negative events, and returns a per-minute rating of a player's performance [74].
- **Value over Replacement Player (VORP)** is a box score estimate of the points per 100 team possessions that a player contributed above a replacement-level (−2.0) player, translated to an average team and prorated to an 82-game season. Multiplied by 2.70 it converts to wins over replacement (WORP). VORP is a positive\negative real-value analytic [64].
- **Win Shares (WS):** It is an estimation of the number of wins that each player contributed to his team's win overall on the season [32].
- **Tendex:** Statistical model to determine player efficiency of basketball players. Considered the first rating formula by using linear weights [75].

Nevertheless, most of the metrics share similarities in measurement that intend to give a total perspective on a player's statistical performance [38]. A great player will achieve high performance in most of these metrics. On the contrary, conventional players will show low values in all of them [25].

The aim of this research is to satisfy the increasing demand for new techniques and provide significant insights and advanced analytics for teams, technical staff and players. Players' performance prediction depends on many variables such as psychology, injury risk [9], bad shots in the starting minutes of the match, opponents match-ups which can give important impact in future performance, but are very difficult to define and quantify [4]. Nevertheless, technical staff and data analysts need to evaluate these metrics, monitor and track the performance and finally take important decisions for the future acquisitions or selections in a team [76].

The team line-up evaluation and final choice is critical during the game, but also before the game when structuring the strategy of the team against the opponents. LinNet is a calibrated network embedding model for line-up evaluation [65]. The quantification of each player rating in team roster and the profile building based on player style of the current season and the number possessions can drive the construction of the line-up [77].

To conclude, this work could also be combined in the future with data driven sensors methodologies in each athlete and SportVU camera data that were introduced in recent years [78]. Therefore, an integrated solution can be developed in the short future that will include advanced metrics with proper visualizations, heat maps, player tendencies [57]. Therefore, the complete solution will quantify not only game statistics, but also important behavioral metrics that depend on physical conditions as well. This information is important in order to optimize performance and can lead a team to win more games [14].

The rating KPIs (Table 1) are quite important in data analysis but can be enforced with the analysis of unexplored moves or decisions before them. This gap is important in order to understand the conceptual of players contribution to final result [71].

Based on previous researches four factors of basketball analytics with different weights assigned (eFG%, FTr, REB and TOV) show that the evaluation of them give a big impact on team and player performance [6,9,66].

5. Discussion, case study & forecasting scenario

5.1. Discussion

Tables 1–6 in the Appendix illustrate important basketball analytics that dominate the game. A clear segmentation was conducted in this paper with the objective to categorize player or team characteristics in basketball and guideline subject matters.

In Table 1 there are the crucial performance basketball analytics. USG% counts the percentage of player engagement during the time played. NetRtg determines the amount of scored points by a player per 100 possession minus the opponent same formula against the player. Win Shares (WS) is a five-part formula that examines the offensive play in a very precisely way but does not explain the defensive play in all important criteria. Regarding shooting efficiency there are effective field goal percentage (eFG%) and true shooting (TS%). Both of them are excellent basketball analytics that can better explain the shooting ability. eFG% takes has a weighted formula by adding three points while TS% takes into account all shooting categories into the counted formula. REB% calculates the percentage of rebounds that a player takes when he is on court.

The defensive criteria (Table 2) include steals and blocks as basic metrics. DefRtg is an advanced analytic that shows the difference of on court vs. off court in the defense performance of player or team. Deflections and Def Loose Balls Recovered analytics are important hustle metrics that influence the defensive game a lot. In addition, the influence on opponent shots is crucial in order to adjust the proper tactics. The aforementioned are known as "really big plays". In fact, they are the actions that could inspire or ignite a team as an extra boost to change the momentum and final result of the game. STL% and BLK% shows the ability of a player in steals or blocks in each team possession respectively. In addition, a successful steal or block does not let the opponent to score, but at the same time gives the opportunity for a fast break offense. BLK% considers field goal attempts (FGA) in comparison with the attempted shots in order to explain the blocking abilities.

In offense (Table 3), the basic analytics are points, rebounds, assists and the shooting percentages in each distance. The Usage (USG) rate calculates the possessions finished by a player, but there are cases of players with high usage rate while they were assisted on most of their field goals or they had on their own field goal. Additionally, a good screen to the teammate can go into an easy basket (Screen Assists metric and PTS). An offensive loose ball recovery or an offensive rebound are important because they can give an extra team possession. AST% is important because it determines the pace and volume of the game with an adjustment of the time played. Points Per Possession (PPP) explains the scoring efficiency while a player has the ball. The number of three-point field goal (3P) attempts have increased dramatically the last decade due to better defense, tactics and athletic abilities of the players. 3P capability can be recognized as a game changer because that offensive skill can be trained and evaluated in big men of a team as well. By having a benchmark between FGA/Poss, OR/Poss and TOV/Poss helps to target for high scores in offensive rebounds, increase FGA and offset turnovers. PTS/Poss has an elasticity point (as we are saying in economics) where there is a critical point between tempo and scoring in order to achieve good results on average scored points at each time that a player touches the ball during the offense.

In Table 4 there are overall categorized basketball analytics. TOV% is a rate metric that focuses on the percentage that a player makes mistakes over time while is on the court. Assists/Turnover ratio (AST/TOV) can better measure the efficiency since correlates the offensive and defensive criteria according to their possessions and it is more representative than to compare the average of assists and/or turnovers. AST/Poss and AST/FGM are important metrics because it shows how well performed each possession and turns into basket. STL/DP and OR/Poss should have higher values than TOV/Poss for better performance results. Teams that have high tempo might have more TOV/G and less TOV/Poss compared with other team.

In Table 5 illustrated the comparison matrix of advanced basketball analytics. The term of NetRtg refers to the offensive and defensive rating (OffRtg and DefRtg), it is a metric that usually the team or players with the higher value are the winners but if a sports analyst uses only this criterion the evaluation of performances is not robust.

Table 6 indicates a comprehensive research with a classification matrix in Sports analytics bibliography that used DM algorithms and techniques for each purpose. Our literature verifies that Sports Analytics is an emerging scientific field that penetrates even more in Sports industry by using DS, ML and DM techniques to optimize performances and forecast in more accuracy.

5.2. Case study

In basketball and for sports in general there are many important analytics that through them we can make important decisions during or after the game as lessons learned for improvement. In order to understand in more depth, the basketball analytics referred in Tables 1–5 we provide a comprehensive analysis of different case studies of Top5 NBA basketball players in the 2018–19 season. According to the most notable analytics these players are Giannis Antetokounmpo (MVP of the year), James Harden (top scorer of the year), Paul George (key player in many categories), Stephen Curry (most efficient shooter) and Rudy Gobert (Defense player of the year) [79]. This research shows the most remarkable achievements of each player of the season with the purpose of benchmarking them across the most significant basketball analytics.

- ***Giannis Antetokounmpo***: he had high scores in the majority of basketball analytics and lead in WS/48 (0.292), PIPM (7.8), PIE (21.8), EFF (35.3), PER (30.9), PACE (105.27) AST Ratio (19), PFD (7.7). Based on high performance in these categories took the award of MVP for the year 2018–19.
- ***James Harden***: he had high skills in scoring and performance analytics such as PTS (36.1), AST (7.5), AST% (39.4), td3 (7), WS (15.2), BPM (11.7), GmSc (16.9), PRA% (64.36), FP (58.7), VOPR (9.9) and USG% (39.6). He was the real competitor for the MVP title.
- ***Paul George***: he was valuable key player for his team by taking lead in NBA league in performance analytics categories such as Deflections (3.8), Loose Balls Recovered (2.1), STL%, RPM (7.63) and WINS (19.9).
- ***Stephen Curry:*** he was the offensive critical point for his team with his effective shooting capabilities. He was notable in analytics categories like AST/TO (1.88), Wins Added (18.8), NetRtg (13.7), TOV% (11.6), eFG% (60.4), TS% (64.1) and PIPM (7.4).
- ***Rudy Gobert:*** he was the defensive player of the year and showed a big hustle for every opponent. He led in REB% (19.4), Screen Assists (6), Screen Assists PTS (13.8), dd2 (66), BLK (2.3), TOV (1.6), DWS (5.7), DBPM (5.1), DRPM (4.4), eFG% (66.9) and TS% (68.2).

According to aforesaid analysis a normalized radial chart (Chart 1) of these top5 players presents the marks of these performance basketball analytics. In addition, Table 1 indicates the ranges of average values in each metric in order to benchmark high–low values.

To conclude, the player's nominees for the season 2018–2019 validated either in terms of overall basketball analytics or in specific categories based on analyzed remarks.

5.3. Forecasting scenario

During recent years the research community and betting companies focused on teams win forecasting instead of concentrating on player impact to the game, as well as the identification of appropriate attributes that are the most important for forecasting purposes.

We illustrate here a forecasting scenario for three basketball seasons 2017 up to 2020 (2017–18, 2018–19 and 2019–20) in NBA basketball competition. The data were retrieved from various sources [79,85,86] and aggregated into a single dataset. After that we performed data cleansing in order to get data ready for analysis and forecasting. For that reason, each season (82 basketball games = Q1–Q4) was split into four groups. The first group represents the first quarter of the season (∼20 games = Q1), the second group relates to basketball analytics for half the season (∼40 games = Q1–Q2), the third group relates to statistics from about 60 games (Q1–Q3). Based on these analytics, our analysis provided predictions for the MVP and Top Defender nominee. Hence, 20 NBA players were selected, on condition of participating to at least 30 games per season and at least 15 min per game playing time on average for the whole season. An additional condition was that they achieved nominations in different statistical categories for these 2 seasons, as can been seen in Table 7. This table represents the awards for MVP, Best Defender, Top in Assists, Points and Rebounds and other important nominees. The final season (2019–20) is not finished yet, thus this research will put forward a prediction for the MVP and Defender of the year. In order to predict these two awards, we introduced and validated two below formulas (the data analyzed and normalized in the scale of 100).

For the 1st introduced formula, box score statistics and important rating basketball analytics selected as variables as an

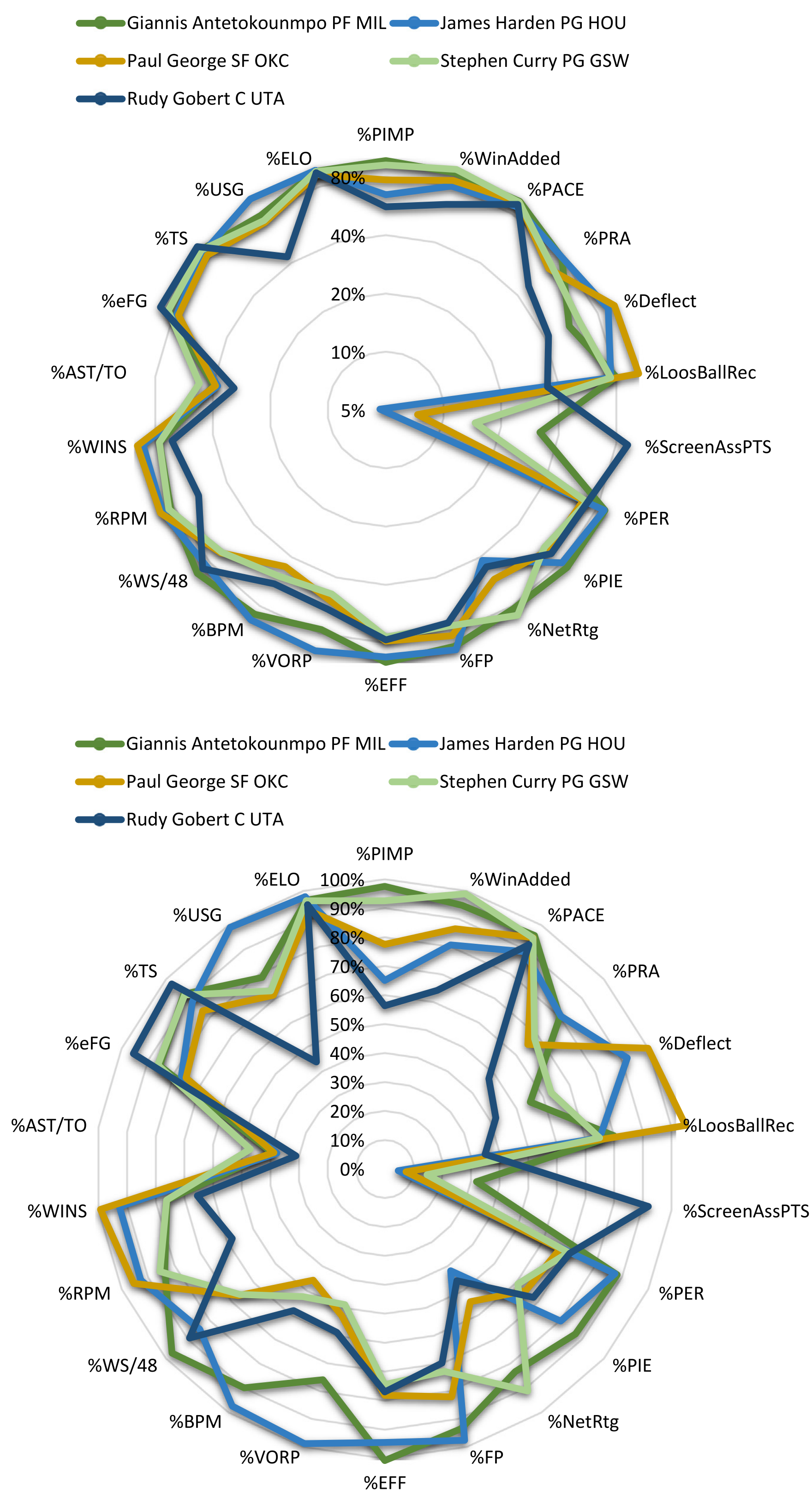

Chart 1. Radial Charts on percentage values and logarithmic normalization.

Aggregated Performance Indicator (API) with following formula:

$$\begin{aligned}\boldsymbol{API} &= [RPM(+/-) + \%PER + \%PIE + \%4Factors + \%NETRTG \\ &+ \%EFF + \%PIR + \%Tendex + \%BPM + \%PIPM + \%GmSc + \%FP \\ &+ \%WS/48 + \%TeamELO + \%EFG\% + \%TS\% + \%VORP + \%WinsRPM \\ &+ \%WAR + \%EWA + \%Deflections + \%PACE + \%USG\% + \%AST/TO \\ &+ \%ScreenAssistsPTS + \%PRA + \%REB\% + \%LooseBallsRecovered \\ &+ \%PPP + \%ASTRatio]/30\end{aligned}$$

Table 1
Advanced ***Rating KPIs***.

Glossary	Description	Metric Type	Formula	Explanation
eFG%	Effective Field Goal Percentage (Values of avg ranges from 0 to 70)	Player, Team	(FG + (0.5 * 3P FG)) \ FGA	One of the recognized Four Factors. Measures field goal percentage adjusting for made 3-point field goals being 1.5 times more valuable than made 2-point field goals. (EFF FG%)
+/−(PM)	Plus Minus (Values of avg ranges up to 12)	Player	TPOC − OPOC Team Points on Court vs Opponent Points on Court	The point differential when a player or team is on the floor. A box score estimates of the points per 100 possessions a player contributed above a league-average player, translated to an average team.
Adj. +/−(APM)	Adjusted Plus Minus (Values of avg ranges up 12)	Player	TPOC48 − TPOC48 Points on Court per 48 min versus Points off Court per 48 min	The prediction is the difference in efficiency of the home team against the opponent of the points per 100 possessions
AR	Assist Ratio (Values of avg ranges from 0 to 35)	Player, Team	(Assists × 100) divided by [(FGA + (FTA × 0.44) + Assists + Turnovers]	The percentage of a player's possessions that ends in an assist.
EFF	Efficiency (Values of avg ranges from 0 to 36)	Player, Team	(PTS + REB + AST + STL + BLK − Missed FG − Missed FT - TO)/GP	Composition of efficiency statistic regarding offensive and defensive contribution
EWA	Estimated Wins Added (Values of avg ranges from 0 to 31)	Player, Team	Value Added divided by 30	This calculation it gives the estimated number of wins a player adds to a team's season total above what a 'replacement player' would produce.
FP	Fantasy Points (Values of avg ranges from 0 to 62)	Player, Team	(1 * PTS) + (1.2 * TRB) + (1.5 * AST) + (3 * STL) + (3 * BLK)−(1 * TO)	The number of fantasy points a player accumulates
GmSc	Game Score (Values of avg ranges from 0 to 21)	Player, Team	PTS + (0.4 × FG) – (0.7 × FGA) – (0.4 × (FTA – FT)) + (0.7 × OREB) + (0.3 × DREB) + STL + (0.7 × AST) + (0.7 × BLK) – (0.4 × PF) – TOV	It is intended to give a "total perspective" on a player's statistical performance in a basketball game, taking into account every statistic listed on a player's box score.
NetRtg	Net Rating (Values of avg ranges up to 17)	Player, Team	OFFRTG − DEFRTG = (100*PTS/(Team FGA + Team TOV + (0.44*Team FTA) – Team OREB)) – (100*Opp PTS/(Opponent FGA + Opponent TOV + (0.44* Opponent FTA) – Opponent OREB))	Net Rating (NetRtg): calculates a team's point differential per 100 possessions. On player level this statistic is the team's point differential per 100 possessions while he is on court.
PER	Player Efficiency Rating (Values of avg ranges from 0 to 33)	Player, Team	• Step1: uPER calculation uPER = (1/M)*(TP+(2/3)*A+(2-factor*(TA/TFG)*FG)+FT*0.5*(1+1−(TA/TFG) + (2/3)*(TA/TFG))−VOP*T−VOP*DRB%*(FGA−FG) −VOP*0.44*(0.44+(0.56*DRB%))*(FTA−FT)+VOP*(1−DRB%)*(TRB−OREB)+VOP*DRB%*OREB+ VOP*S+VOP*DRB%*B−PF*((LFT/LPF)−0.44*(LFTA/LPF)*VOP)) • Step2: Factor and VOP calculation Factor = (2/3)-(0.5*(LA/LFG))/(2*(LFG/LFT)) VOP = LP/(LFGA-LOREB+LT+0.44*LFTA) • Step3: Pace and league adjustment for PER PER = (uPER*(LPace/TPace))*(15/LuPER)	PER calculates all positive and negative accomplishments in per minute rating of player and team performance. Player Efficiency Rating is the overall rating of a player's per-minute statistical production. The league average is 15.00 every season. - pace adjustment = lg_Pace/team_Pace - estimated pace adjustment = 2 * lg_PPG/(team_PPG + opp_PPG) - aPER = (pace adjustment) * uPER
PIE	Player Impact Estimate (Values of avg ranges from 0 to 25)	Player, Team	(PTS + FGM + FTM−FGA−FTA + DREB + (.5 * OREB) + AST + STL + (.5 * BLK)−PF−TO)/(GmPTS + GmFGM + GmFTM−GmFGA−GmFTA + GmDREB + (.5 * GmOREB) + GmAST + GmSTL + (.5 * GmBLK)−GmPF−GmTO)	PIE measures a player's overall statistical contribution against the total statistics in games they play in. PIE yields result which are comparable to other advanced statistics (e.g. PER) using a simple formula.
PIR	Performance Index Rating (Values of avg ranges from 0 to 40)	Player, Team	(PTS + REB + AST + STL + BLK + PFD) – (Missed FG + Missed FT + TOV + BLKA + PF)	It is a metric primarily used in European leagues that attempts to calculate player or team performance.
Pythagorean Win Percentage	Pythagorean Win Percentage	Team	Winning Percentage = GP * (PTS)*16.5/ [PTS Scored)16.5 + (PTS Allowed)16.5)] • Daryl Morey exponential is set to 13.91 • John Hollinger exponential is set to 16.5	Pythagorean Win Percentage is an estimation that shows where a team win percentage based on their points for and points against
Real +/−(RPM)	Real Plus Minus (Values of avg ranges up to 12)	Player, Team	It is the player's or team's average influence of net points different per X (100) offensive and defensive possessions.	Player's estimated on-court impact on team performance, measured in net point differential per 100 offensive and defensive possessions. RPM takes into account teammates, opponents and additional factors

(*continued on next page*)

Table 1 (*continued*).

Glossary	Description	Metric Type	Formula	Explanation
PIPM	Player Impact Plus Minus (Values of avg ranges up to 9)	Player, Team	{[(ORtg + DRtg) + (AvgORtg + AvgDRtg)] ∗ (Min∧2)}/(G*Min) G*Min = 82*48 if its NBA	PIPM is a Plus-minus metric that adjusts the luck part with box-score data. These 3 components are: box-score prior, luck-adjusted on-off data, and luck-adjusted net rating.
Tendex	Tendex (Values of average ranges from 0 to 0.4)	Player, Team	• Standard Tendex Rating: (raw statistical formula) (PTS) + (REB) + (AST) + (STL) + (BLK)−(Missed FG)−0.5 * (Missed FT)−(TOV)−(PF)/(MP)/(Game Pace). • Modified Tendex Rating: (weighted average statistical formula) (PTS) + (REB) + 1.25 ∗ (AST) + 1.25 ∗ (STL) + (BLK) − 1.25 ∗ (TOV) − (Missed FG) − (Missed FT/2) − (PF/2)/(*MP*) /(Game Pace).	It is one of the first formula for performance calculation. Tendex is the summary of positive and negative efforts
USG%	Usage Rate (Values of avg ranges from 0 to 40)	Player	{[FGA + (FT Attempts × 0.44) + (AST × 0.33) + TOV] × Total MIN × League Pace} divided by (MIN × Team Pace)	The number of possessions a player is finishing per game.
VA	Value Added	Player, Team	[MIN * (PER−PRL)]/67 PRL = 11.5 for power forwards, 11.0 for point guards, 10.6 for centers, 10.5 for shooting guards and small forwards	The estimated number of points a player adds to a team's season total above what a 'replacement player' (for instance, the 12th man on the roster) would produce.
VORP	Value over Replacement Player (Values of avg ranges from 0 to 10)	Player	[BPM – (−2.0)] * (% of MP) * (GP/82)	A box score estimates of the points per 100 TEAM possessions that a player contributed above a replacement-level. (RL is −2.0 of an avg team of 82 game season)
Wins Added	Wins Added	Player, Team	{[Avg(onORtg, onDRtg) + (PIPM/2)∧13.91]Avg(onORtg, onDRtg) + (PIPM/2)∧13.91 + [Avg(onORtg, onDRtg) − (PIPM/2)∧13.91] − RPL} * [MP/Min]	It is the combination of on court Rtg, the PIMP above a replacement level (RPL)
WS	Win Shares (Values of avg ranges from 0 to 0.3)	Player, Team	(PP-0.92 * LPPP * (FGA+0.44 *FTA+TO)) /(0.32*LPPG*(TP/LP)) + (MP/TMP * TDP * (1.08 * LPPP-DefRtg/100)/(0.32*LPPG*(TP/LP)	Win Shares is the sum of offensive and defensive Win Shares
WS/48 or 40	Win Shares (Per 48 or 40 min)	Player, Team	Win Shares divided with minutes played (league average is approximately .100)	An estimate of the number of wins contributed by a player per 48 min
PACE	Pace (avg values from 95 to 110)	Player, Team	MP * ((Tm Poss + Opp Poss)/(2 * (Tm MP/5))).	The number of possessions a team uses per game. Pace factor is an estimate of the number of possessions per Minutes Played (MP) by a team.
USG%	Usage Percentage (ranges of avg values from 5 to 40)	Player	(FGA + POSS Ending FTA + TO)/POSS	The % of tactic plays in a player used while on the floor
TS%	True Shooting Percentage (Values from 40 to 75)	Player	PTS/2 * [(FGA + (0.44 * FTA)	It captures the effect of shooting percentage if we accounted for free throws and 3-pointers. The factor 0.44 can be adjusted based of linear model of past seasons.
TOR	Turnover Ratio	Player, Team	(TOV × 100)/[(FGA + (FTA × 0.44) + AST + TOV]	The percentage of possessions that end in a turnover
POSS	Possessions	Player, Team	0.96 * (FGA) + (TO) + (0.44 * (FTA)−(OREB))	The number of possessions played by a player or team.
POSS\G	Possession per Game	Player, Team	Total FGA+ (0.475*FTA) + TOV−OREB	The number of possessions that played divided by the played games
REB% or REBr	Rebound Rate (Values of avg ranges from 0 to 25)	Player, Team	(100 × (REBs × Team MIN))/[Player MIN × (Team REBs + Opponent REBs)]	One of the recognized Four Factors. The % of missed shots that a player rebound. (Rebounding Percentage)
PRA/G	Points, Rebound and Assist (Values of avg ranges from 0 to 50)	Player, Team	Points + Rebounds + Assists/Game	The average of player with points, rebounds and assists per game
WAR or WARP	Wins Above Replacement (Values of avg ranges from 0 to 20)	Player	(Win%−RL) * (Min/48) Where: Win% = TmOffRat∧14/(TmOffRat∧14 + TmDefRat∧14)	The evaluation of player performance of him and four average players of his team compared with the opponent team with four average players and a replacement level player.

Figs. 2 and 3 show the forecast, according to API, for the MVP for the years 2017–18 and 2018–19, respectively, which can be verified and cross-referenced with Table 7. Fig. 4 illustrated the forecast for the MVP in year 2019–20. In Fig. 2 (Season 2017–18), James Harden in the whole season (Q1–Q4) had the best performance, in term of API formula and took the MVP award. For Fig. 3 (Season 2018–19), Giannis Antetokounmpo voted as MVP and verified his performance (76.5%) in API scale.

Table 2
Defensive criteria *- Advanced basketball statistics.*

Glossary	Description	Metric Type	Explanation
BLK%	Block Percentage or Block Rate	Player	The opponent % two-point field goal attempts blocked while he was on the floor. 100*(BLK*(TMP/5))/(MP*(OFGA-O3PA))
DREB%	Defensive Rebound Percentage	Player	The % of available defensive rebounds a player grabbed while he was on the floor.
PF%	Percent of Team's Personal Fouls	Player	The % of a team's personal fouls that a player has while on the court
STL%	Steal Percentage	Player, Team	The % of opponent possessions that end with a steal by the player on the floor. 100*(STL*(TMP/5))/(MP*OP)
BLK	Blocks	Player, Team	A block occurs when the defense player tips the ball, blocking their chance to score
Deflections	Deflections (Values of avg ranges from 0 to 5)	Player, Team	The number of the defense tackle the ball on a non-shot attempt
DREB	Defensive Rebounds	Player, Team	The number of rebounds a player or team has collected while they were on defense
STL	Steals	Player, Team	Number of times that takes the ball from a player on offense, causing a turnover
DefRtg	Defensive Rating	Player, Team	The number of points allowed per 100 possessions by a team. For a player, it is the number of points per 100 possessions that the team allows while that individual player is on the court. The formula is: 100*((Opp Points)/(Opp POSS)).
DBPM	Defensive Plus/Minus	Player, Team	A box score estimates of the defensive points per 100 possessions a player contributed above a league-average player, translated to an average team.
DEF EFF	Defensive Efficiency	Team	The number of points a team scores per 100 possessions. The formula is: (100*Opp Points/(Opponent FGA + Opponent TOV + (0.44* Opponent FTA) – Opponent OREB))
DPR	Defensive Player Rating	Player	The formula is: (Player spg+ Player bpg/team minutes played)-(times blown by*Pace of Players Era)* Total Average of Possessions +(Players DRTG*Team Pace)/Total number of years played
DRPM	Defensive Real Plus Minus	Player, Team	Player's estimated on-court impact on team defensive performance, measured in points allowed per 100 defensive possessions.
DWS	Defensive Win Shares	Player	The number of wins contributed by a player due to his defense.

PLAYER	Quarter	API
J. Harden	Q1	75.1%
J. Harden	Q1-Q3	73.1%
J. Harden	Q1-Q2	72.0%
J. Harden	Q1-Q4	71.9%
L. James	Q1	71.3%
L. James	Q1-Q2	70.5%
S. Curry	Q1-Q2	70.0%
L. James	Q1-Q4	68.2%
S. Curry	Q1	67.5%
S. Curry	Q1-Q3	67.5%
S. Curry	Q1-Q4	66.7%
L. James	Q1-Q3	66.5%

Fig. 2. API forecast for 2017–18.

PLAYER	Quarter	API
A. Davis	Q1	76.7%
G. Antetokounmpo	Q1	76.6%
G. Antetokounmpo	Q1-Q4	76.5%
G. Antetokounmpo	Q1-Q3	75.2%
A. Davis	Q1-Q2	74.8%
G. Antetokounmpo	Q1-Q2	73.2%
J. Harden	Q1-Q4	72.4%
A. Davis	Q1-Q3	72.1%
J. Harden	Q1-Q3	71.0%
A. Davis	Q1-Q4	69.7%
J. Harden	Q1-Q2	66.7%
J. Harden	Q1	64.4%

Fig. 3. API forecast for 2018–19.

During our analysis for Season 2019–2020 (which is in process right now) API formula predicts that Giannis Antetokounmpo has significant advantage for the MVP award against the second James Harden (77.8% versus 67.4% respectively) in Q1 up to Q3 of the regular season.

The 2nd introduced formula focused as defensive criterion for the selection of the Defensive Player of the Year, the basketball analytics variables selected implies the equation of **Defensive Performance Indicator (DPI):**

$$\boldsymbol{DPI} = BLK - BLKA + PFD - PF + STL + Deflections$$
$$+ LooseBallsRecovered - TOV + ScreenAssistsPTS + AST/TO$$

Figs. 5 and 6 verified the forecasts with the actual results of Table 7, while Fig. 7 shows the expected result for the 2019–20. In Fig. 5 Gobert and Davis shared the same highest score in DPI scale but Gobert voted as the Best Defender of the year 2017–18. For Fig. 6 (Season 2018–19), DPI formula verified that Gobert was the Best Defender. In the last Fig. 7 (2019–20 from Q1 up to Q3), DPI predicts that Gobert with 92.8% is the highest candidate for this award for the regular season.

Data scrapping was performed through Python packages. All data were retrieved from various NBA sports sources ([79,85] and [86]) and aggregated in an Excel file followed by data cleansing actions. In addition, a normalization process of the final data was performed with the purpose to use them in the suggested

Table 3
Offensive criteria *- Advanced basketball statistics.*

Glossary	Description	Metric Type	Explanation
3PA%	3 Point Field Goals Percentage	Player, Team	The % of 3-point field goals attempted while on the court
3PM%	3 Point Field Goals Percentage Made	Player, Team	The % 3-point field goals made while on the court
BLKA%	Percent Blocked Field Goal Attempts	Player, Team	The % own blocked field goal attempts while on the court
FGA%	Field Goal Percentage Attempted	Player, Team	The % field goals attempted while on the court
FGM%	Field Goal Percentage Made	Player, Team	The % made field goals while on the court
FTA%	Free Throw Percentage Attempted	Player, Team	The % made free throws has made while on the court
FTM%	Free Throw Percentage Made	Player, Team	The % made free throws while on the court
FTr	Free Throw Factor	Player, Team	One of Four Factor. How often it goes to line and how often they made it. FT/FGA
FTM/FTA%	Percent of Team's Free Throws Made	Player, Team	Team free throw attempts made per field goal attempt
OREB%	Offensive Rebound Percentage	Player	The % of available offensive rebounds a player took while is on the floor
PFD%	% of Team's Personal Fouls Drawn	Player, Team	The % of a team's personal fouls drawn by a player has while on the court
PTS%	% of Team's Points	Player, Team	The % of a team's points that a player has while on the court
PTS 2PT%	% of Points (2-Point Field Goals)	Player, Team	The % of points scored by a player or team from 2-point field goals
PTS 3PT%	% of Points (3-Point Field Goals)	Player, Team	The % of points scored by a player or team from 3-point field goals
PTS FBPS%	% of Points (Fast Break Points)	Player, Team	The % of scored points by a player or team from fast break opportunities
PTS FT%	Percent of Points (Free Throws)	Player, Team	The % of scored points by a player or team from free throws
2nd PTS	Second Chance Points	Player, Team	The % of isolation plays that shoots free throws of a shooting foul
3PA	3 Point Field Goals Attempted	Player, Team	The number of 3-point field goals that a player or team has attempted
3PM	3 Point Field Goals Made	Player, Team	The % of a team's 3-point field goals made while on the court
FBPS	Fast Break Points	Player, Team	The number of points scored by a player or team while on a fast break
FGA	Field Goals Attempted	Player, Team	The number of 2-point field goals attempted
FGA/ Poss	FGA/Possession	Player, Team	Calculated the shot attempts in each possession.
FGM	Field Goals Made	Player, Team	The number of 2-point field goals made
FTA	Free Throws Attempted	Player, Team	The number of free throws attempted
FTM	Free Throws Made	Player, Team	The number of free throws made
OR/P	Offensive Rebounds/Possession	Player, Team	Offensive Rebounds per completed possession
OREB	Offensive Rebounds	Player, Team	The number of rebounds gathered while they were on offense
PFD	Personal Fouls Drawn	Player, Team	The number of personal fouls that are drawn by a player or team
PITP	Points in the Paint	Player, Team	The number of points scored by a player or team in the paint
PTS/Poss	Points/Possession	Player, Team	The made points in each time touches the ball.
PTS	Points	Player, Team	The number of scored points.
PTS Off Tov	Points Off Turnovers	Player, Team	The number of points scored following an opponent's turnover.
OBPM	Offensive Plus/Minus	Player	A box score estimates of the offensive points per 100 possessions a player
OFF EFF	Offensive Efficiency	Team	The number of points a team scores per 100 possessions. The formula is: (100*Points/(Team FGA + Team TOV + (0.44*Team FTA)−Team OREB))
OffRtg	Offensive Rating	Player, Team	Measures a team's or player (on court) points scored per 100 possessions. (100*Points/(Team FGA + Team TOV + (0.44*Team FTA)−Team OREB))
ORPM	Offensive RPM	Player, Team	Player's on-court impact on team offensive performance in points scored per 100 offensive possessions
OWS	Offensive Win Shares	Player	The number of wins contributed by a player due to offense.
PPP	Points Per Possession (Values of avg ranges up to 1.8)	Player, Team	The number of points a player or team scores per possession PTS/(FGA+0.44*FTA+TOV)

formulas (API and DPI). Our code and the corresponding Excel file we used for data analysis can be found on GitHub at the following link: https://github.com/vsarlis/nbastats.

Some researches tried to correlate players' salaries with their performance in order to predict their "real" salaries with the use of PIE and the WinsRPM (Pythagorean Win Estimation) as basketball performance analytics and performed regression analysis [87]. The MVP award is a multivariate type of selection between players' performance advanced analytics and team's worth in the league [88]. The Total Performance Index (TPI) was

Table 4
Overall Performance criteria *- Advanced basketball statistics.*

Glossary	Description	Metric Type	Explanation
AST%	Assist Percentage	Player, Team	The % of teammate's FG that a player assists when is on the floor 100*AST/(((MP/(TMP/5))*TFG)-FG) AST = Assists, MP = Minutes Played, TMP = Team Minutes Played, TFG = Team Field Goals, FG = Field Goals
TOV%	Turnover Percentage	Player, Team	The number of turnovers committed per 100 possessions. One of the recognized Four Factors. 100*TOV/(FGA+0.44*FTA+TOV)
TRB%	Total Rebound Percentage	Player	100*(TRB*(TMP/5))/(MP*(TTRB+OTRB)). It is a weighted average of total rebounds a player took while he was on the floor.
AST/Poss	Assists per Possession	Player, Team	It shows how well passed the ball through the game.
AST	Assists	Player, Team	The number of assists that goes to a made basket
BLKA	Blocks against	Player, Team	The number of shots attempted and blocked by a defender
DD2	Double doubles	Player	The number of double-digit number total in two of the five statistical categories in a game that a player achieves
TD3	Triple doubles	Player	The number of double-digit number total in three of the five statistical categories in a game that a player achieves
FT/ Poss	FT/Possession	Player, Team	The free throw shot attempt per every possession.
FTA RATE	Free Throw Attempt	Player, Team	The number of free throws attempts in comparison to the number of field goal attempts
GP	Games Played	Player, Team	The number of games a team or player played
L	Losses	Player, Team	The number of games lost by a team or player
MPG/MIN	Minutes Played	Player	The number of minutes played by a team or player
PF	Personal Fouls	Player, Team	The number of personal fouls a player or team committed
PRL	Position Replacement Level	Player, Team	PRL = 11.5 for power forwards, 11.0 for point guards, 10.6 for centers, 10.5 for shooting guards and small forwards
REB	Rebounds	Player, Team	The number of total rebounds a team or player has collected on either offense or defense
STL/DP	Steals/Defensive Possession	Player, Team	How many steals your defense gets for every one of your opponent's offensive possessions.
TOV/Poss	Turnovers/Possession	Player, Team	How often a team or player made a turnover every time you touch the ball.
TOV	Turnovers	Player, Team	A turnover occurs when the player or team on offense loses the ball to the defense
W	Wins	Player, Team	The number of games won by a team or player
%WIN	Win Percentage	Player, Team	W/GP. The percentage of games played that a player or team has won
Loose Ball Rec	Loose Ball Recovered (Values of avg ranges from 0 to 12)	Player, Team	The defensive or offensive actions while trying to secure a loose ball
AST/TOV	Assist to Turnover Ratio (Values of avg ranges from 0 to 3)	Player, Team	The number of assists for a player or team compared to the number of turnovers they have committed

PLAYER	Quarter	API
G. Antetokounmpo	Q1-Q3	77.8%
G. Antetokounmpo	Q1-Q2	77.2%
G. Antetokounmpo	Q1	77.0%
J. Harden	Q1	72.2%
J. Harden	Q1-Q2	71.5%
L. Doncic	Q1	68.1%
J. Harden	Q1-Q3	67.4%
L. Doncic	Q1-Q2	65.5%
L. Doncic	Q1-Q3	61.8%

Fig. 4. API forecast for 2019–20.

PLAYER	Quarter	DPI
R. Gobert	Q1	87.8%
R. Gobert	Q1-Q3	87.3%
R. Gobert	Q1-Q4	86.8%
A. Davis	Q1-Q4	86.8%
A. Davis	Q1-Q3	85.8%
A. Davis	Q1	84.9%
R. Gobert	Q1-Q2	73.5%
A. Drummond	Q1-Q2	73.2%
A. Drummond	Q1-Q3	73.0%
A. Drummond	Q1-Q4	72.1%
A. Drummond	Q1	67.5%
A. Davis	Q1-Q2	66.7%

Fig. 5. DPI forecast 2017–18.

introduced as a proposed basketball performance metric [89] and compared with PIR analytics. TPI based on previous research shows better results than PIR (64.6% vs 32.7%) but focusing on box score quantitative metrics and not on qualitative.

Cumulative Individual Accolades (CIA) was suggested as a formula but also failed to predict with accuracy the MVP of the year 2017–28 and proposed J. Harden in the 2nd place [90]. The most accurate forecasting was performed with the use of Back Propagation (BP) Neural Networks based on trained data from NBA seasons 2010–11 up to 2017–18 with the correlation of PER basketball analytic [91]. In order to avoid overfitting and overtraining of the model they adopt L2 regularization method. The prediction of BP neural algorithm shows accurate results for

Table 5
Comparison matrix *for basketball performance analytics.*

Rating type	Advantages	Disadvantages	Type
+/−or BPM or PM	- Can be used either single game or season - Impact of player while he is on the court - Can be tied with VORP and USG% for better player performance estimation	- Do no show the solely impact of a player in the scoring - Poor handling of offense outliers - Poor handling of block shots - Not good defense rating - Overvalue players with high values of USG and REB	Linear regression model
Adj. +/−or ABPM or APM	- Shows the efficiency for both opponents and teammates on the court	- Do not show the specific player ability as an individual apart from team impact. Some coaches also select some player duos or trios frequently - High variance even with the use of regression with different roles, coaching tactics, teammates and matchups - The increment of data do not decrease the statistical significance- Do not have extra info for player tracking or play by play stats - Bloated variances due to a non-invertible distribution of players	Multiple regression model
Real Plus Minus (Real +/−or RPM or RAPM)	- Based on the development of APM and uses in more detail aging curves and Bayesian priors. - One of the most important basketball indicators till now - It is simple and understandable analytic	- Based only in scoring factor and impact of each player compared to a league average of players per 100 possessions.	Linear regression model with a weight played on the square of the coefficients
EFF	- Can be used either single game or season	- Focus on box score data of a game - Does not have any specific weight of a statistic category	Linear formula
ELO Rating	- Each team correlated with ranking based on expected wins - It is a good way to rate teams and use it for future projections	- ELO calculation focus only teams rating and cannot count players rating	Linear regression model with weights
FP	- It was started as a fantasy sports analytic for performance evaluation. - It is ideal for competition with long regular season because the statistics could be normalized	The weights that set sometimes can give advantage in some categories, can be disadvantage in other statistics.	Linear formula with specific weights
GmSc	- All statistics are weighted differently based on frequency that they occurred - Positive and negative coefficient according to the contribution	- Do not apply to seasonal statistics - Do not show the specific player ability as an individual apart from team impact - In shooting categories the player should have 57% for breakeven - Focuses on player efficiency	Linear formula with specific weights
NetRtg	- A normalized metric of defensive and offensive over 100 possessions can count the ability to count the pace of teams or players	- The estimation of number of possessions can give different results or forecasting. Sometimes there is a bias in the presented results, and this can drive into undesirable comparisons of players or teams. - Overestimation in possessions can drive into underestimations in ratings.	Linear formula with specific weights
PACE	Shows how controlled is a team or a player since a faster pace can give more possessions\opportunities through a game.	Do not give attention into multivariate factors for better player or team performance estimation	Linear formula with specific weights
PER (Player Efficiency Rating)	-Performance rating by calculating positive and negative accomplishments per minute -Accounting the team's pace comparing to league average	- Does not count any other parameter other than steals and blocks. - Rewarding the inefficient shooting. Two points field goal made worth 1.65 points and three points field goal made worth 2.65 while missed costs 0.72 points. Hence, the shooting value could be break even in 30% 2points shots and 21% for 3points. - A player who shots more with the aforementioned results can gain better PER - Non logical phenomenon of extra high performed PERs in extremely limited minutes	Multiple regression model
PIE	PIE is quite similar to PER logic by calculating the per minute offense production and defense categories.	- It captures many parameters but cannot explain in depth how performed well a player or a team. - There is arbitrage in PIR weights calculations	Linear formula with specific weights
PIR	PER representing the performance with focusing in per-minute contribution and pace adjustment. Hence, can easily compare normalized performances between team and players	- The logic is similar with PIE with having different weights under considerations - PER fails to identify the most accurate value due to the fact that the weights are arbitrarily calculated	Linear formula with specific weights
Pythagorean Win Percentage	- Estimation of Win Percentage based on Point for and Points Against	- Simple estimation without any weight factor or regression model	Linear formula

(*continued on next page*)

Table 5 (*continued*).

Rating type	Advantages	Disadvantages	Type
Tendex	A weighted basketball analytic that counts the performance. It gives extra value for positive aspects more in assists and steals and negative in missed shots or turnovers	As all weighted formula focuses in specific criteria of the game. This approach can boost some teams or players and underestimate some others.	Linear formula with specific weights
TS%	- Measuring the equivalent of FTA with FGA - An adjustment of factor 0.44 based on season statistics could give precisely results	- TS% is biased in terms of FTs and underestimates the number of scored points per possession - A proposed formula for precisely calculations could be: PTS/POSS with an FG/FTA	Linear formula with specific weights
eFG%	- Measuring the impact of shooting efficiency of three points added value	Measuring only the shooting performance of players without adding other important factors	Linear formula with specific weights
USG%	Interpret the player usage while he was on the floor.	A player that likes to pass more than to shoot does not mean that has lower impact in the game	Linear formula with specific weights
VORP	- Based on BPM as an enhanced version and convert through the calculation into the estimate of overall contribution	- Correlated with replacement player factor	Linear regression model
PIPM	It is a differentiation of Plus Minus metrics by measuring the influence of possessions	Extra or less possession through the games can impact game result	Multiple regression model
WS or WS/48	- The comprehensive evaluation in offensive play of single player - Better evaluation of a single player due to the division of the minutes played - It offers a model of marginal offense per marginal points per win as a contribution result for the victory - Based on expected Pythagorean Win Percentage rather than actual wins.	- Being part of a good team implies better score in WS - Not an overall good evaluation calculation - Better evaluation take players that have big amount of time on the court	Multiple regression model
EPV (Expected Possession Value)	- It is a forecasting analytic in a continuous manner that helps in decision making - It is a framework for basketball analytics that can overcome conventional box-score metrics	Focuses in micro-analytic for sports and related of how many possessions took a team	Forecasting methods based on regression models
Four Factors	- Shows the importance and in which way can each player or team acts - All these metrics associated with team success -Based on researches the accuracy level is 94% on average	- These basketball analytics do not capture the winning tendencies of players and teams. - Simple logic of major factors that impact the game. Score when is possible and take more possessions when it is not	Linear formula with specific weights
CARMELO	- Make projections based on past and current data based on RPM and BPM and for defense by adding the DRAYMOND metric - It is a blend of latest used analytics for team and player performance	- CARMELO methodology cannot be replicated and cannot validate the accurate results of BPM and WS. - Do not account factors as injuries, psychology and work ethics	Forecasting methods based on regression models
WAR (Wins Above Replacement)	- Players can be rated on per-minute basis accounting winning value - Replacement level metric evaluates the performance based on minutes played	- Only box-score evaluation (does not take into account other contributions) - Assumptions in replacement level, efficiency, USG	Linear formula with specific weights
VA	- Normalizes the sum of positive and negative contribution by introducing the factor of RPL and focusing on PER analytic factor.	The estimate of 'replacement player' influence the result.	Linear formula with specific weights

PLAYER	Quarter	DPI
A. Davis	Q1	99.4%
A. Davis	Q1-Q2	94.3%
R. Gobert	Q1-Q3	91.1%
R. Gobert	Q1-Q4	91.0%
R. Gobert	Q1-Q2	89.9%
A. Davis	Q1-Q3	88.7%
R. Gobert	Q1	87.0%
A. Davis	Q1-Q4	83.3%
J. Embiid	Q1	66.8%
J. Embiid	Q1-Q4	65.0%
J. Embiid	Q1-Q3	62.9%
J. Embiid	Q1-Q2	61.5%

Fig. 6. DPI forecast for 2018–19.

PLAYER	Quarter	DPI
R. Gobert	Q1	100%
R. Gobert	Q1-Q2	99.9%
R. Gobert	Q1-Q3	100.0%
A. Drummond	Q1	92.4%
A. Drummond	Q1-Q2	91.8%
A. Drummond	Q1-Q3	89.5%
G. Antetokounmpo	Q1-Q3	78.2%
G. Antetokounmpo	Q1	77.1%
G. Antetokounmpo	Q1-Q2	75.0%

Fig. 7. DPI forecast for 2019–20.

the MVP award [91]. All in all, this method is successful as it correctly predicted the MVP for 5 years running but requires several years past data.

Table 6
Data Mining algorithms and techniques *used in Sports Analytics.*

Data Mining Method	Type of Used Method	Purpose	Accuracy
K-Means clustering [55]	N nearest Trajectory embedding	Suitable for large datasets and is not as sensitive to outliers as other clustering techniques [80]	Capture a diverse and comprehensive set of player movements
Archetype Analysis (AA) and Archetypoid Analysis (ADA) [20]	- FADA for sparse time series data - ADA with h-plot for dissimilar data	Obtain outstanding players (positively and negatively). Archetypes are data-driven extreme points	- ADA can be used for better performance understanding - AA shows the extremes cases through data
Functional data analysis (FDA) [20]	- Simple linear models - ANOVA - Generalized linear models - PCA - Clustering - Classification	Modern branch of statistics that analyzes data that are drawn from continuous underlying processes,	FDA results are consistent with domain experts of sports analytics.
Neural Networks (NN) and Recurrent Neural Network (RNN) [42]		Variant of neural networks that can deal with sequential data of variable length. It was used for strategy classification in basketball through data from SportVU cameras.	- NN achieved 54.7% accuracy - RNN scored 65.6% accuracy after better understanding of data
Latent Dirichlet Allocation (LDA) [81]	LDA is a latent factor (similar to components analysis)	Organize offensive structure in possessions of basketball	Illustrated finds repeated patterns in offensive structure in basketball teams
Randomization Inference for Leader Effects (RIFLE) [82]		How much influence the wins basketball coaches	Additional Monte Carlo simulations that there is no coach effect in a team win. 20% or 30% influence of teams' success
Bayesian [66]	Regulation — Ridge regression (Tikhonov regulation)	Cross-validation (CV) technique. CV used in order to determine the optimal limit of minutes (that played) for the standard APM linear regression technique.	Reasonable ranges with the proper parameters. Due to the overfitting the model to forecast the performance degraded.
Deep Learning [55]	- K-means clustering by descriptive label in each cluster to interpret large datasets with non-sensitivity to outliers	The comparison of patterns for individual player movements on offense strategy	By finding similar positions finder, data analysts can find easily the proper positions from different seasons. The results showed accuracy from 50% to 75% between the clusters
Markov Modeling [63,83]	- Markov model used for expected point calculation - Entropy is used to quantify the unpredictability - Gibbs sampler (Markov chain Monte Carlo MCMC)	The analysis of ball movement and effective unpredictability in basketball offense	- The complex correlation of ball movement is significant important by the results and verified by the game theoretic tactics. - Use of Gibbs sampler to predict the full posterior distributions of unknown parameters
Markov Model transition to Poisson point processes [84]	Modeled Two-dimensional data	By this transition, it was extended the model from spatial statistics into flexible nonparametric methods which allows complicated patterns	In most sports data are in two-dimensions so in general the assumption in most times will be not violated
- Network Analysis (NA) - Neural Networks (NN) - Bootstraping technique [63]	- Used algorithm Latent Pathway Identification Analysis (LPIA) - Used eigenvector centrality to measure the impact (centrality or importance) of a node in the network.	- NA used to find the optimal path of game plays that generate the most points - NN used to forecast the results of NBA games - To count the statistical significance of players performance central scores	For better accuracy used adjusted p-values to classify the outliers over\under performers with threshold of 10% to avoid bias. Some interesting results are the low importance score correlated with small *p*-value

In contrast, our work correctly predicted the MVP using only current data. Furthermore, we provide predictions at the quarter level, and to the best of our knowledge, there is no current method to predict the best defender. In this work, we propose API as a sophisticated formula for MVP prediction based on approved qualitative and quantitative advanced analytics by the basketball community. API was used as an aggregated model of selected algorithms that incorporate the statistical data in a way that showed 100% accuracy for the years of 2017–18 and 2018–19. For the year 2019–20, the forecast for the MVP nominee is G. Antetokounmpo according to API, and Defensive player of the year R. Gobert based on DPI.

6. Conclusion & future work

6.1. Conclusion

Basketball is a team sport and that means the significance of analytics is not only to distinct the most efficient players and teams but also the optimal combination on pairs of a team with purpose to optimize the performance on the court [63].

Team rotation is also momentous in player selection. Hence, a team has a roster of 12 players that they are ready to be productive and efficient for each minute that they play. The new coaching trend shows that the technical staff desires 12 eligible players to be ready in each match. Over the last years there is award for the 6th player and that means the bench players can make huge difference during the game. Therefore, the proper balance of team rotation and roles distribution is a key factor for the team success and here is a huge difference in that approach comparing with the previous decade [92]

The quantification of uncertainty or luck cannot be underestimated. The purpose of this research is to evaluate the most important rating parameters in basketball and minimize the uncertainty. In addition, clutch factor which is the ability of a player to take correct decisions in critical moments or under pressure in the last seconds of a game. Hence, it is an analytics with big percentage of bias and difficult to predict [88]. Specifically,

Table 7
MVP, Best Defender, Top Scorer, Top in Assists, Top in Steals, Top in Rebounds and 3 best teams of the year for two seasons 2017–18 and 2018–19.

Players	2018-2019	2017-2018
Andre Drummond	Top Rebounder	Top Rebounder
Anthony Davis		Top Blocker - 1st Team
Ben Simmons		Rookie of the Year
Damian Lillard	2nd Team	1st Team
Giannis Antetokounmpo	MVP - 1st Team	2nd Team
James Harden	Top Scorer - 1st Team	MVP - Top Scorer - 1st Team
Jimmy Butler		3rd Team
Joel Embiid	2nd Team	2nd Team
Karl-Anthony Towns		3rd Team
Kawhi Leonard	2nd Team	GAP YEAR
Kevin Durant	2nd Team	1st Team
Kyrie Irving	2nd Team	
LaMarcus Aldridge		2nd Team
LeBron James	3rd Team	1st Team
Luka Doncic	Rookie of the Year	GAP YEAR
Nikola Jokić	1st Team	
Paul George	Top Steals - 1st Team	3rd Team
Rudy Gobert	Best Defender - 3rd Team	Best Defender - 3rd Team
Russell Westbrook	Top Assist - 3rd Team	Top Assist - 2nd Team
Stephen Curry	1st Team	3rd Team

the exploration of different sports analytics and after that the evaluation of them can boost and give extra value in the decision making in order to understand in more depth each sport.

Currently, APM and PM are the most efficient performance indicators. Especially APM uses regression model to calculate the impact of teammates and opponents while they were in the court but they do not take into account the matchup with players and their opponents [66]. In addition, both techniques (PM and APM) affect overfitting which cannot analyze circumstances of players that appeared very frequently on the floor with those who appeared very rarely.

Line-up roster selection and five players basic roster of a basketball team cannot be based only in analyzed metrics such as APM, RAPM, LinNet method and other referred analytics. This analysis is difficult to forecast accurately because there are many qualitative indicators and players skillsets that cannot specified easily. In addition, the prediction of specific matchups can take place with remarkable accuracy for few minutes but cannot remain in these high levels for a whole match [65]. The optimal choice of a line-up is very complex due to different combinations of players on the court, the difficult to find the best performance indicator in each position, time, play and opponent [23].

Based on the background research results illustrate that an aggregation of players\teams statistics, statistical modeling, visual metrics (SportVU cameras analysis, heatmaps etc.), commentary or social networks metrics, biomechanics stats, training\gym stats and wearables metrics in an optimized performance formula could be the most sophisticated solution with purpose to calculate the real performance of teams and players by minimizing the bias as much as possible [52].

The enormous amount of variable resources and different sets of data increase the complexity to interpret and develop ways and patterns of better understanding. Most of the times, the domain experts and technical staff are ex-athletes that try to interpret the game and circumstances. Hence, through sports analytics with the critical explanation can leverage them to use this knowledge as competitive advantage [14].

In the case study section, we analyzed the top5 players performance with regards to selected advanced basketball analytics and validated the yearly nominees for the awards. It was shown that our method is the only one predicting the best defender and one that requires only current data in order to correctly predict the MVP. Basketball is a team game with much complexity related to several factors such as playing, coaching decisions, team chemistry, psychology, sociology, training, marketing and health so that it is difficult to estimate the greatest NBA player for each season.

Table 8
Basketball Analytics Abbreviations.

Basketball Analytics Abbreviations	Explanation
%WIN	Win Percentage
2nd PTS	Second Chance Points
3P or TP	3 Points
3PA	3 Point Field Goals Attempted
3PA%	3 Point Field Goals Percentage
3PM	3 Point Field Goals Made
3PM%	3 Point Field Goals Percentage Made
Adj. +/−(APM)	Adjusted Plus Minus
AGE or A	Age
AI	Artificial Intelligence
AR	Assist Ratio
AST	Assists
AST RATIO	Assists Ratio
AST%	Assists Percentage
AST/A	Assists
AST/Possession	Assists per Possession
AST/TO	Assists per Turnover Ratio
BA	Basketball Analytics
BLK or B	Blocks
BLK%	Blocks Percentage
BLKA	Blocks against
BLKA%	Percent Blocked Field Goal Attempts
BPM	Box Plus Minus
DBPM	Defensive Box Plus Minus
dd2	Double double
DEF EFF	Defensive Efficiency
Defl	Deflections
DEFRTG	Defensive Rating
DM	Data Mining
DM	Data Mining
DP	Defensive Possession
DPR	Defensive Player Rating
DRB%	Defensive Rebound Percentage
DREB	Defensive Rebounds
DREB%	Defensive Rebounds Percentage
DRPM	Defensive Real Plus Minus
DS	Data Science
DWS	Defensive Win Shares
EFF	Efficiency
eFG%	Effective Field Goal Percentage
ELO	Team ELO
EPTS	Electronic Performance and Tracking Systems
EWA	Estimated Wins Added
FBPS	Fast Break Points
FG	Field Goals
FGA	Field-Goal Attempts
FGA%	Field Goal Percentage Attempted
FGA/ Possession	FGA/Possession
FGM	Field Goals Made
FGM%	Field Goal Percentage Made
FP	Fantasy Points
FT	Free Throws
FT/ Possession	FT/Possession
FT/FTA	Free throw attempts per field goal attempt
FTA	Free-Throw Attempts
FTA%	Free Throw Percentage Attempted
FTM	Free Throws Made
FTM%	Free Throw Percentage Made
FTM/FTA%	Percent of Team's Free Throws Made
FTr	Free Throw Factor
GmSc	Game Score
GP	Games Played
L	Losses
LA	League Assists
LFG	League Field Goals
LFGA	League Field-Goal Attempts
LFT	League Free Throws
LFTA	League Free-Throw Attempts
Loose Ball Rec	Loose Ball Recovered
Loose Balls Recovered	Loose Balls Recovered
LORB	League Offensive Rebounds

(*continued on next page*)

Table 8 (*continued*).

Basketball Analytics Abbreviations	Explanation
LPace or LP	League Pace
LPF	League Personal Fouls
LPPG	League Points Per Game
LPPP	League Points Per Possession
LPTS	League Points
LTOV	League Turnovers
LuPER	League uPER
M or Min	Minutes
ML	Machine Learning
MP	Minutes played
MPG/MIN	Minutes Played
MVP	Most Valuable Player
NetRtg	Net Rating
OBPM	Offensive Box Plus Minus
OFF EFF	Offensive Efficiency
OffRtg	Offensive Rating
OPOC	Opponent Points on Court
OR/P	Offensive Rebounds/Possession
OREB	Offensive Rebounds
OREB%	Offensive Rebounds Percentage
ORPM	Offensive Real Plus Minus
ORPM	Offensive RPM
OWS	Offensive Win Shares
PACE	Pace
PER	Player Efficiency Rating
PF	Personal Fouls
PFD	Personal Fouls Drawn
PFD%	% of Team's Personal Fouls Drawn
PIE	Player Impact Estimate
PIPM	Player Impact Plus Minus
PIR	Performance Index Rating
PITP	Points in the Paint
PL or P	Player
PM (+/−)	Plus Minus
Pos	Position
Poss	Possession
PP	Points Produced
PPP	Points Per Possession
PRA%	Points Rebounds Assists Percentage
PRL	Position Replacement Level
PTS	Points
PTS 2PT%	% of Points (2-Point Field Goals)
PTS 3PT%	% of Points (3-Point Field Goals)
PTS FBPS%	% of Points (Fast Break Points)
PTS FT%	Percent of Points (Free Throws)
PTS OFF TO	Points Off Turnovers
PTS%	% of Team's Points
PTS/Possession	Points/Possession
Real +/−(RPM)	Real Plus Minus
REB or TREB	Total Rebounds
REB%	Total Rebounds Percentage
RPM	Real Plus Minus
SA	Sports Analytics
Screen Assists	Screen Assists
Screen Assists PTS	Screen Assists to Points
SDM	Sports Data Mining
STL or S	Steals
STL%	Steals Percentage
TA	Team Assists
td3	Triple Double
TDP	Team Defensive Possessions
TFG	Team Field Goals
Tm or T	Team
TMP	Team Minutes Played
TO or TOV	Turnovers
TO RATIO	Turnovers Ratio
TOV%	Turnovers Percentage
TP or TPace	Team Pace
TPOC	Team Points on Court
TPOffC48	Points off Court per 48 min
TPOnC48	Points on Court per 48 min
TRB	Total Rebounds

(*continued on next page*)

Table 8 (*continued*).

Basketball Analytics Abbreviations	Explanation
TRB%	Total Rebound Percentage
TS%	True Shooting Percentage
USG%	Usage Percentage
VORP	Value Over Replacement
W	Wins
WAR or WARP	Wins Above Replacement
WINS or RPM WINS	Real Plus Minus WINs
WS	Win Shares
WS/48	Win Shares per 48 min
AST/FGM	Assist per Field Goal Made

6.2. Future work

Sports Analytics can be used in innumerable types, such as Social engagement, Performance biomechanics analysis, psychological and physical metrics and the aforementioned critical analysis of advanced sports statistics so that technical staff and domain experts can understand more the game and improve the processes and methodologies [9,23].

Predictive Analytics can be applied for forecasting purposes through different factors in order to understand teams and opponent's strength and weaknesses.

In addition to the previous research a further supplementary work can pay attention into the physical, psychological or injury part of the available metrics for players and team prediction. For the technical team and coaches, it is a big opportunity to predict such bad possible situations that could cost in team performance. There is a study that focus in NBA players psychology and behavior which measure the athlete's social networks activity and correlates with players performance for future games. Sentiment analysis performed on those online social posts in order to understand the thoughts and behavior of the players and conclude in useful reports.

Motion capture technologies it is already the trend with tremendous data collections and be able to track every team or player movement on the court in milliseconds. The game statistics, sensors' data from wearable devices, Computer vision through SportVU cameras and aforesaid valuable information can be used as aggregated outcomes for useful statements using DM and ML techniques that could affect in a noteworthy level in sports decision [93].

Basketball is a sport of decisions and that means the team selection, the way of training, the psychological part of a player, a possession, a deflection, a pass or a shoot can give serious impact to the performance and directly as sequence to the result of a game [71]. An additionally potential future work could be the analysis and calculation of expected possession value (EPV), which rates and evaluate each taken decision during the basketball game.

According to the current research and the two suggested formulas (API and DPI), a further optimization based on empirical results can apply specific weights on each algorithmic parameter and produce the relevant results.

Finally, sports analytics of different domains can feed directly betting intelligence systems in order to forecast players and teams results for maximizing their profits and accuracy [16,62,94] and [95]. Future research will focus on the construction of a formula with combined weighted basketball ratings based on play by play data with the purpose to optimize not only the performance evaluation but also to forecast more accurate.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The only known to us conflict of interest is with staff working at the International Hellenic University.

Appendix

We include here various tables with basketball metrics, for Rating KPIs (Table 1), Defensive criteria (Table 2), Offensive criteria (Table 3), Overall Performance criteria (Table 4), a Comparison matrix for basketball performance analytics (Table 5), as well as DM algorithms used in Sports Analytics (Table 6) classified and categorized through multiple sources such as [79,85,86,92,96–98] and [99]. Table 7 shows the MVP, Best Defender, Top Scorer, Top in Assists, Top in Steals, Top in Rebounds and the 3 best teams of the year for two seasons (2017–18 and 2018–19). We also include two Radial Charts on percentage values and logarithmic normalization (Chart 1) for the top5 performed players in Season 2018–19 and a list of Basketball Analytics Abbreviations (Table 8).

References

[1] E. Turban, R. Sharda, D. Delen, Decision Support and Business Intelligence Systems, Vol. 9, ninth ed., Pearson, 2011.
[2] A. Senderovich, A. Shleyfman, M. Weidlich, A. Gal, To aggregate or to eliminate? optimal model simplification for improved process performance prediction, Inf. Syst. (2018) 1–16.
[3] B. Gerrard, Moneyball and the role of sports analytics: A decision-theoretic perspective, in: North American Society for Sport Management Conference, NASSM 2016, 2016, pp. 2010–2012, no. Nassm.
[4] K. Pelechrinis, E. Papalexakis, Athlytics, in: Proc. Elev. ACM Int. Conf. Web Search Data Min, WSDM '18, 2018, pp. 787–788.
[5] O. Kostakis, N. Tatti, A. Gionis, Discovering Recurring Activity in Temporal Networks, Vol. 31, No. 6, Springer US, 2017.
[6] W. Tichy, Changing the Game, ACM, 2016, pp. 1–10, no. May.
[7] V.C.P. Chen, S. Bum, K. Asil, D. Sundaramoorthi, Preface?: Data Mining and Analytics, Springer Sci. Media, 2018.
[8] G. Vinue, I. Epifanio, Forecasting basketball players' performance using sparse functional data, Stat. Anal. Data Min. (1) (2018) 1–26.
[9] P.O. Donoghue, Research Methods for Sports Performance Analysis, 2009.
[10] S.G.H. Victor Blanco, Roman Salmeron, A multicriteria selection system based on player performance. case study: The spanish ACB basketball league, 2018, pp. 1–17.
[11] M. Dunham, Data Mining - Introductory and Advanced Topics, Prentice Hall, 2003.
[12] I. Katakis, et al., Mining urban data (Part B), Inf. Syst. 57 (2016) 75–76.
[13] E. Siegel, Predictive Analytics, Wiley, 2017.
[14] J.F. Drazan, A.K. Loya, B.D. Horne, R. Eglash, From sports to Science?: Using basketball analytics to broaden the appeal of math and science among youth, in: 2017 MIT-Sloan Sport. Anal. Conf, no. March, 2017, pp. 1–16.
[15] R.P. Schumaker, A.T. Jarmoszko, C.S. Labedz, Predicting wins and spread in the Premier League using a sentiment analysis of twitter, Decis. Support Syst. 88 (2016) 76–84.
[16] R. Silva, Minusha, Sports Analytics, 2016.
[17] B.C. Alamar, Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers, Columbia University Press, 2013.
[18] E. Morgulev, O.H. Azar, R. Lidor, Sports analytics and the big-data era, Int. J. Data Sci. Anal. 5 (4) (2018) 213–222.
[19] I. Bhandari, E. Colet, J. Parket, Z. Pines, R. Pratap, K. Ramanujam, Advanced scout: Data mining and knowledge discovery in NBA data, Data Min. Knowl. Discov. 125 (1997) 121–125.
[20] G. Vinué, I. Epifanio, Archetypoid analysis for sports analytics, Data Min. Knowl. Discov. 31 (6) (2017).
[21] T.B. Swartz, Working in sports analytics, 2009, pp. 12–15.
[22] B. Marr, Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance, John Wiley & Sons, 2015.
[23] T.H. Davenport, Analytics in sports: The new science of winning, Int. Inst. Anal. 2 (February) (2014) 1–28.
[24] H. Arrieta, et al., Relative age effect and performance in the U16, U18 and U20 European basketball championships championships, J. Sports Sci. 0414 (March) (2016).
[25] F. Bianchi, T. Facchinetti, P. Zuccolotto, Role revolution?: towards a new meaning of positions in basketball, electron, J. Appl. Stat. Anal. 10 (03) (2017) 712–734.

[26] Aurélien Géron, Hands-on Machine Learning with Scikit-Learn & TensorFlow, O'Reilly Media, 2017.
[27] J. Fernández, From training to match performance: An exploratory and predictive analysis on F.C. Barcelona, in: Facultat d'Informàtica de Barcelona, 2016.
[28] J. Foreman, Data Smart, Wiley, 2014.
[29] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, third ed., Morgan Kaufmann, 2011.
[30] C.M. Bishop, Machine Learning and Pattern Recoginiton, Springer US, 2007.
[31] P. Tzirakis, C. Tjortjis, T3C: improving a decision tree classification algorithm's interval splits on continuous attributes, Adv. Data Anal. Classif. 11 (2) (2017) 353–370.
[32] C. Cao, Sports Data Mining Technology Used in Basketball Outcome Prediction, 2012.
[33] I. Witten, E. Frank, M. Hall, Data Mining: Practical Machine Learning Tools and Techniques, No. 3, third ed., Elsevier Inc., 2011.
[34] R. Ul Mustafa, M.S. Nawaz, M.I.U. Lali, T. Zia, W. Mehmood, Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning methods, Malaysian, J. Comput. Sci. 30 (1) (2017) 63–76.
[35] T.M. Mitchell, Machine Learning, McGraw-Hill Science, 1997.
[36] S. Raschka, Python Machine Learning: Unlock Deeper Insights Into Machine Learning with this Vital Guide to Cutting-Edge Predictive Analytics, 2016.
[37] M.R. Berthold, Journeys to Data Mining, Springer US, 2012.
[38] D. Miljković, L. Gajić, A. Kovačević, Z. Konjović, The use of data mining for basketball matches outcomes prediction, in: SIISY 2010-8th IEEE International Symposium on Intelligent Systems and Informatics, Vol. 30, No. September, 2010, pp. 309–312.
[39] S.M. Ghafari, C. Tjortjis, A survey on association rules mining using heuristics, WIREs Data Min. Knowl. Discov. 9 (4) (2019).
[40] S. Yakhchi, S.M. Ghafari, C. Tjortjis, M. Fazeli, ARMICA-Improved: A New Approach for Association Rule Mining, No. 3, Springer International Publishing, 2017, pp. 296–306.
[41] L. Dong, C. Tjortjis, Experiences of using a quantitative approach for mining association rules, in: Handbook of Research on Emerging Rule-Based Languages and Technologies, No. May, 2003, pp. 693–700.
[42] K. Wang, R. Zemel, Classifying NBA offensive plays using neural networks, in: MIT Sloan Sport. Anal. Conf., 2016, pp. 1–9.
[43] P. Foster, T. Fawcett, Data Science for Business What You Need to Know About Data Mining and Data Analytic Thinking, O'Reilly, 2013.
[44] J. Ledolter, Data Mining and Business Analytics with R, Wiley, 2013.
[45] J. Gudmundsson, M. Horton, Spatio-temporal analysis of team sports – a survey, 2016.
[46] C. Tjortjis, J. Keane, T3: A Classification Algorithm for Data Mining, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2412, 2002, pp. 50–55, No. August 2002.
[47] A. Franks, A. Miller, L. Bornn, K. Goldsberry, Characterizing the spatial structure of defensive skill in professional basketball, Ann. Appl. Stat. 9 (1) (2015) 94–121.
[48] L. Lefebure, Understanding Player Positions in the NBA, 2014, pp. 1–5.
[49] R. Nagarajan, L. Li, Optimizing NBA player selection strategies based on salary and statistics analysis, in: Proc. - 2017 IEEE 15th Int. Conf. Dependable, Auton. Secur. Comput. 2017 IEEE 15th Int. Conf. Pervasive Intell. Comput. 2017 IEEE 3rd Int. Conf. Big Data Intell. Compu, 2018-Janua, 2018, pp. 1076–1083.
[50] V. Madhavan, Predicting NBA game outcomes with hidden Markov models, 2017.
[51] D. Cervone, L. Bornn, K. Goldsberry, NBA court realty intro: the basketball court is a real estate market, in: MIT Sloan Sport. Anal. Conf., No. 1, 2016,
[52] A. Franks, A. Miller, L. Bornn, K. Goldsberry, Counterpoints: Advanced defensive metrics for NBA basketball, in: MIT Sloan Sport. Anal. Conf., 2015, pp. 1–8.
[53] D. Lutz, A cluster analysis of NBA players, 2010.
[54] L. Sha, et al., Interactive sports analytics: An intelligent interface for utilizing trajectories for interactive sports play retrieval and analytics, ACM Trans. Comput. Interact. 25 (2) (2018).
[55] A. Nistala, J. Guttag, Using deep learning to understand patterns of player movement in the NBA, in: MIT Sloan Sport. Anal. Conf. Conf., 2019, pp. 1–14.
[56] M. Cohen, M. Sloane, Predictive modeling and statistical analysis in sports, 2015.
[57] E. Papalexakis, K. Pelechrinis, THoops: A multi-aspect analytical framework spatio-temporal basketball data, 2017.
[58] A. Miller, L. Bornn, K. Goldsberry, Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball, Vol. 32, 2014, pp. 1–9.
[59] Y. Kanellopoulos, P. Antonellis, C. Tjortjis, C. Makris, N. Tsirakis, K-attractors: A partitional clustering algorithm for numeric data analysis, Appl. Artif. Intell. 25 (2) (2011) 97–115.
[60] R.P. Schumaker, O.K. Solieman, H. Chen, Sports knowledge management and data mining, in: Annual Review of Information Science and Technology, Vol. 44, Springer, 2010, pp. 115–158.
[61] T. Chartier, Embracing imperfection, J. Corp. Account. Financ. 27 (4) (2016) 67–68.
[62] P.O.S. Vaz de Melo, V.A.F. Almeida, A.A.F. Loureiro, C. Faloutsos, Forecasting in the NBA and other team sports: Network effects in action, ACM Trans. Knowl. Discov. Data 6 (3) (2012) 1–27.
[63] L. Pham, S. Anand, J. Piette, Evaluating basketball player performance via statistical network modeling, in: MIT Sloan Sport. Anal. Conf., no. June, 2011.
[64] A.M. Franks, A. D'Amour, D. Cervone, L. Bornn, Meta-analytics: Tools for understanding the statistical properties of sports metrics, J. Quant. Anal. Sports 12 (4) (2016) 151–165.
[65] K. Pelechrinis, Linnet: Probabilistic Lineup Evaluation Through Network Embedding, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 11053 LNAI, 2019, pp. 20–36.
[66] J. Sill, Improved NBA adjusted + / - using regularization and out-of-sample testing, in: MIT Sloan Sports Analytics Conference, 2010, pp. 1–7.
[67] www.bball-index.com. [Online]. Available: https://www.bball-index.com/player-impact-plus-minus/. (Accessed 15 September 2019).
[68] Y. Chen, Y. Gong, X. Li, Evaluating NBA player performance using bounded integer data envelopment analysis, Infor Inf. Syst. Oper. Res. 55 (1) (2017) 38–51.
[69] M. van Bommel, L. Bornn, Adjusting for scorekeeper bias in NBA box scores, Data Min. Knowl. Discov. 31 (6) (2017) 1622–1642.
[70] Databall - Advanced stats in the NBA for dummies. [Online]. Available: https://nbastatsgeeks.wordpress.com/newinnovative-apbrmetrics/expected-possession-value-epv/.
[71] D. Cervone, A. D'Amour, L. Bornn, K. Goldsberry, POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data, in: MIT Sloan Sport. Anal. Conf., Vol. 64, 2012, pp. S1081–S1082.
[72] D. Cervone, A. D'Amour, L. Bornn, K. Goldsberry, A multiresolution stochastic process model for predicting basketball possession outcomes, J. Amer. Statist. Assoc. 111 (514) (2016) 585–599.
[73] S.J. Miller, A derivation of the pythagorean won-loss formula in baseball, Vol. 30, 2006, pp. 1–13.
[74] Y. Yang, Predicting regular season results of NBA teams based on regression analysis of common basketball statistics, 2015.
[75] M. Sachanidi, N. Apostolidis, D. Chatzicharistos, M. Sachanidi, N. Apostolidis, D. Chatzicharistos, Passing efficacy of young basketball players: test or observation? Int. J. Perform. Anal. Sport (2013).
[76] T. Chartier, Coachable business results, J. Corp. Account. Financ. 27 (1) (2015) 83–85.
[77] M. Goldberg, Evaluating lineups and complementary play styles in the NBA the harvard community has made this, 2018.
[78] T. Chartier, Valuing Data, Vol. 28, No. 2, Wiley Online Library, 2017, pp. 88–89, Jan.
[79] NBA.com. [Online]. Available: https://stats.nba.com.
[80] M. Kontaki, Y. Manolopoulos, A. Gounaris, A.N. Papadopoulos, K. Tsichlas, Efficient and flexible algorithms for monitoring distance-based outliers over data streams, Inf. Syst. (2015) 1–17.
[81] A.C. Miller, L. Bornn, Possession sketches?: Mapping NBA strategies, in: MIT Sloan Sport. Anal. Conf. Conf., 2017, pp. 1–12.
[82] C.R. Berry, A. Fowler, How much do coaches matter? in: MIT Sloan Sport. Anal. Conf. 2011, 2019, pp. 1–25.
[83] A. D'Amour, D. Cervone, L. Bornn, K. Goldsberry, Move or Die?: How ball movement creates open shots in the NBA, in: MIT Sloan Sport. Anal. Conf. Conf., 2015, 2015.
[84] J. Mortensen, L. Bornn, From Markov models to Poisson point processes: Modeling movement in the NBA, in: MIT Sloan Sport. Anal. Conf. Conf., 2019, pp. 1–10.
[85] NBA basketball reference, 2019, [Online]. Available: https://www.basketball-reference.com/. (Accessed 20 August 2019).
[86] ESPN NBA stats, 2019, [Online]. Available: https://www.espn.com/nba/stats. (Accessed 20 August 2019).
[87] J. Rosen, P. Arcidiacono, F. Advisor, K. Kimbrough, F. Advisor, Determining NBA free agent salary from player performance, 2016.
[88] E. Whitmoyer, Measuring greatness in the NBA, 2019.
[89] C. Marmarinos, T. Bolatoglou, K. Karteroliotis, N. Apostolidis, Structural validity and reliability of new index for evaluation of high-level basketball players, Int. J. Perform. Anal. Sport 19 (4) (2019) 624–631.
[90] J.C. Scheide, M.J. Krebs, Generating relative pick value in the NBA draft and predicting success from college basketball generating relative pick value in the NBA draft and predicting success from college basketball, 2019.
[91] J. Hu, H. Zhang, J. Qiu, Prediction of MVP attribution in NBA regular match based on BP neural network model, in: ACM Int. Conf. Proceeding Ser, 2019.
[92] Statathlon.com, 2019, [Online] Available: https://statathlon.com/ (Accessed 20 August 2019).

[93] F. Sun, C. Yi, W. Li, Y. Li, A wearable H-shirt for exercise ECG monitoring and individual lactate threshold computing, Comput. Ind. 92–93 (2017) 1–11.
[94] H. Manner, Modeling and forecasting the outcomes of NBA basketball games, J. Quant. Anal. Sports 12 (1) (2016) 31–41.
[95] B.T. Foster, M.D. Binns, Analytics for the Front Office: Valuing Protections on NBA Draft Picks, in: MIT Sloan Sport. Anal. Conf., 2019, pp. 1–26.
[96] APBRMetrics, [Online]. Available: http://www.apbr.org/metrics/index.php.
[97] Fivethirtyeight, 2019, [Online]. Available: https://fivethirtyeight.com/. (Accessed 20 August 2019).
[98] Boxscoregeeks.com, 2019, [Online]. Available: https://www.boxscoregeeks.com. (Accessed 20 August 2019).
[99] Pivot sports, 2019, [Online] Available: https://blog.pivotsports.com/. (Accessed 20 August 2019).