International Journal of Business Analytics
Volume 9 • Issue 1

Business Analytics in Sport Talent Acquisition:

Methods, Experiences, and Open Research Opportunities

Rocio de la Torre, Public University of Navarre, Spain

Laura O. Calvet, Universitat Oberta de Catalunya, Spain

David Lopez-Lopez, ESADE, Spain

Angel A. Juan, Universitat Oberta de Catalunya, Spain

https://orcid.org/0000-0003-1392-1776

Sara Hatami, Universitat Oberta de Catalunya, Spain

ABSTRACT

Recruitment of young talented players is a critical activity for most professional teams in different sports such as football, soccer, basketball, baseball, cycling, etc. In the past, the selection of the most promising players was done just by relying on the experts' opinions but without systematic data support. Nowadays, the existence of large amounts of data and powerful analytical tools have raised the interest in making informed decisions based on data analysis and data-driven methods. Hence, most professional clubs are integrating data scientists to support managers with data-intensive methods and techniques that can identify the best candidates and predict their future evolution. This paper reviews existing work on the use of data analytics, artificial intelligence, and machine learning methods in talent acquisition. A numerical case study, based on real-life data, is also included to illustrate some of the potential applications of business analytics in sport talent acquisition. In addition, research trends, challenges, and open lines are also identified and discussed.

KEYWORDS

Business Analytics, Machine Learning, Sports, Talent Acquisition

1. INTRODUCTION

In the present day, finding and hiring talented workers has become one of the top priorities for many businesses. Inefficient hiring practices have a negative repercussion on any organization, and might impose considerable loses, both in terms of money and time. As pointed out by Davenport et al. (2010), those companies that are capable of attracting and retaining the best talented people are among the most competitive ones. According to Harris et al. (2011), in a globalized and highly competitive environment most organizations should start using data to measure and improve the contribution of their human resources (HR) to their performance.

In the sports sector, Baker et al. (2017), De Bosscher and De Rycke (2017) and Hanlon et al. (2014) state that there is an increasing interest in understanding the costs and benefits of initiatives

DOI: 10.4018/IJBAN.290406

for early identification of talented players, as well as in unleashing the factors that influence athletes' development. These authors also affirm that, while traditional statistical analysis was focused on match variables, such us goals scored o players' position on the field, recent advances in sports analytics are focused on more complex issues like talent acquisition. As pointed out by Gerrard (2017), the 2011 film 'Moneyball'[1] highlighted the possibilities of analytics as a competitive strategy, particularly for small-market teams with relatively limited resources. Following Fried and Mumcu (2016), many coaches employ data on habits and performance indicators to assess the potential of their players. In fact, it is possible to use data to: *(i)* evaluate players' performance; *(ii)* rank players (Pappalardo et al., 2019); *(iii)* estimate the value of players in transfer markets (Kim et al., 2019); *(iv)* locate the best position that a player can occupy in the field; or *(v)* forecast players' goal scoring performance in the next season (Apostolou and Tjortjis, 2019). As illustrated in Figure 1, adapted from 21st Club[2], clubs can use data to analyze the impact of a new player on the team's overall performance level.

The sports industry is being transformed by data analytics in the following dimensions: *(i) at clubs level*, e.g., soccer clubs like Liverpool, Barcelona[3],

Figure 1. Player recommendation based on data analytics (adapted from 21st Club)

Arsenal, Manchester City, or Milan are among the ones that already use data analysis to improve performance, analyze rivals, prevent injuries, optimize the management of the transfer market, and also the acquisition of new talent; *(ii)* regarding *new entrants in data management and analysis*, new platforms for data analysis and management appear to provide services to clubs, such as Wiscout[4] and Scisports[5]; *(iii)* as regards as *new entrants in data capture and generation*, large companies such as Intel have launched the creation of wearable Internet-of-Things devices, which are capable of capturing players' information in real time[6]; *(iv)* with respect to *transformation of sports management professionals*, new official university degrees dedicated to sports management have emerged, most of them with a special emphasis on data science; and *(v)* a *transformation of sports enthusiasts*, who have begun to consume complex data, both for their own digitization process as well as for their leisure and fun when using online betting applications.

This paper presents a comprehensive review of the state of the art regarding data-driven approaches for talent acquisition in sports. In addition, the paper discusses the most common used methods for talent acquisition. Some of these methods belong to the fields of artificial intelligence (AI) and machine learning (ML), which are also introduced in the context of sports analytics. Finally, the paper identifies and discusses trends, challenges, and open research lines related to this research area. The rest of the paper is organized as follows. Section 2 bring out the concept of talent analytics, specially in the sports sector. Section 3 provides a short introduction to the fields of AI and ML, thus allowing the unfamiliar reader to follow the following sections. Section 4 discusses different data-driven analytic approaches. Afterwards, Sections 5 and 6 review research papers on data analytics in recruitment and sports talent acquisition, respectively. Section 7 provides an original case study based on a real-life dataset of European soccer players, where some of the potential of ML methods is illustrated. Trends, challenges, and open research lines are discussed in Section 8. Finally, Section 9 summarizes the main contributions of this paper.

2. TALENT ANALYTICS IN SPORTS

Talent analytics (TA) represents a groundbreaking opportunity for many organizations in the sports sector. TA is defined by Bassi (2011) as "an evidence-based approach for making better decisions on the people side of the business; it consists of an array of tools and technologies, ranging from simple reporting of HR metrics all the way up to predictive modeling". During the last years, professional and elite sport teams are giving a considerable attention to implement business-oriented TA in their strategy. Hence, for Davids and Arau´jo (2019) the main challenge is not the management of talent among the players who already belong to a team, but the acquisition of young and talented players for the future. When considering young people, evolution and forecasting models that go beyond data on the current performance level should be built as well (Webb et al., 2020; Pifer, 2019; Ford and Williams, 2017). As Williams and Reilly (2000) state, multi-dimensional data (including physical, psychical, and sociological characteristics) has to be considered. Likewise, Davcheva (2014) maintains that data obtained from social networks could be used as a predictor of potential children's talent, while Martin (2015) explores the influence of the child socio-economic status on its performance and future evolution.

Other authors (Kandra˘c et al., 2019; Pickering et al., 2019; Loland, 2015; Webborn et al., 2015; Cˆot´e, 1999) discuss about whether talent is an innate skill or not, and if data analysis should be oriented towards innate capacities and genes rather than to performance. Given the number of factors that may affect decisions related to talent acquisition, authors such as Vaeyens et al. (2008), Gerrard (2017), and Bergkamp et al. (2019) conclude that talent acquisition is a multi-dimensional challenge, and one that is not only based on the skills of individual players but also on the whole team, i.e., the current team configuration has to be considered as well. In this context, Gerrard (2017) propose the use of simulation models as more effective tools than expert judgment, especially in a multi-dimensional environment like the one being considered.

Using data analytics, Gandelman (2009) was able to show that outside opportunities were higher for soccer players with a superior socioeconomic background and a better education. He also found evidence of racial dis- crimination in the Uruguayan soccer market, where obtaining a professional contract was easier to white players. Similarly, Berri and Brook (2010) used data analytics to identify some efficiencies regarding the evaluation of players' performance in the National Hockey League. Employing Bayesian analysis combined with Markov Chain Monte Carlo estimation, Rimler et al. (2010) studied the technical efficiency of a basketball team. They were not able to find significant differences, in technical efficiency levels, across teams playing in the same category. Bryson et al. (2013) investigated how the salary of a soccer player might depend upon his ability to play with both feet. After analyzing data from the five main European leagues, they concluded that the aforementioned players tend to receive a noticeable salary premium.

3. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

The popularity of AI and ML methods has been constantly increasing during the last decade (Joshi, 2019). The disciplines in which they have been originated are numerous, including: Computer Science, Mathematics, Electrical Engineering, Statistics, Signal Processing, etc. These fields find applications in a wide range of industries, such as image processing, natural language processing, and online shopping. Figures 2 and 3 shows the variety of AI-related and ML-related research disciplines, and the impact generated measured in terms of the number of indexed publications in the Web of Science (WoS).

Figure 2. WoS-indexed contributions in Artificial Intelligence according to its research area

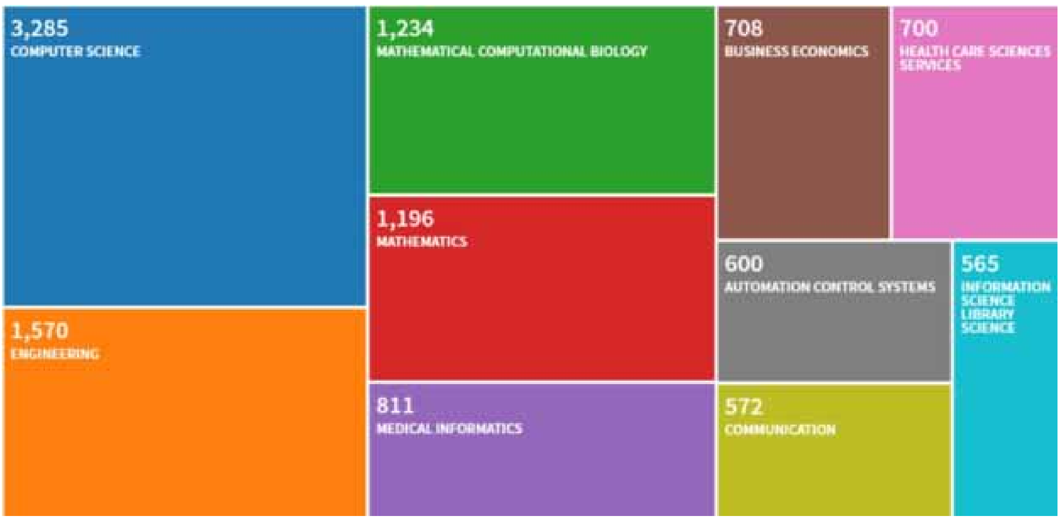
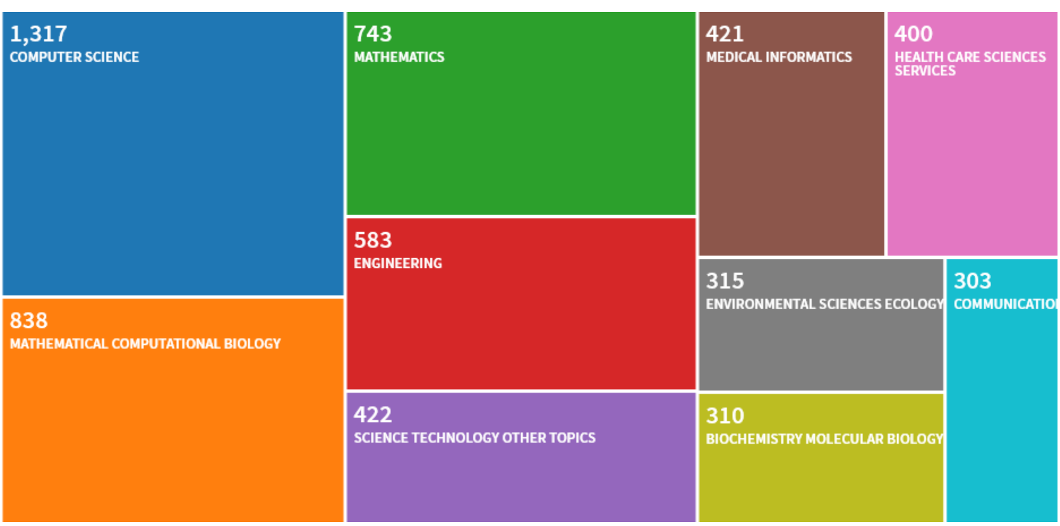


Figure 3. WoS-indexed contributions in Machine Learning according to its research area



According to Broussard et al. (2019), AI refers to machines capable of per- forming one or more tasks associated with the human nature, for instance: learn and process human language, execute mechanical tasks that require complex maneuvering, solving computer-based complex problems that may involve large amount data in a very short time lapse. Others, like Li and Du (2017), define the term as "variety of human intelligent behaviors, such as perception, memory, emotion, judgment, or reasoning, that can be realized artificially by a machine, a system, or a network". Most current applications are focused on the field of neural networks. For instance, deep neural networks are used for speech recognition (Tveter, 1997), image classification (Deepa and Devi, 2011), or prediction of words in a text (Battaglia et al., 2016).

ML as subset of AI refers to a computer program that can learn how to produce a certain behavior, which was not explicitly programmed in it (Kotsiantis et al., 2007). Indeed, it can be capable of showing behaviors from which the programmer may be completely unaware of.

As in human behavior, many aspects of learning and intelligence in AI are closely related to the representation of uncertainty. Therefore, probabilistic approaches are fundamental (Ghahramani, 2015).

Probabilistic methods try to assign an uncertainty measure to the un- known variables, as well as a certain probability to known variables. Hence, the goal is to find the unknown values using probabilistic models. These models are classified into two main types, so called generative and discriminative: discriminative models try to forecast the changes in the output just considering the changes occurred in the input, while generative models are the ones in which the changes in the output can be explained as a consequence of changes in the input as well as changes in the state (Joshi, 2019).

Current research regarding the frontier of probabilistic ML approaches (both discriminative as well as generative) is mainly focused on: *(i)* probabilistic programming (as a general framework for expressing probabilistic models as computer programs); *(ii)* Bayesian optimization (for globally optimizing unknown functions); *(iii)* hierarchical modeling for learning many related models; and *(iv)* probabilistic data compression (Ghahramani, 2015). ML approaches can be divided into supervised and unsupervised learning.

The former are involved in many applications and deal with problems related to learning with guidance. In other words, the training data in supervised learning methods needs labeled samples. Thus, for instance, samples with class labels are required in a classification problem. Hence, the mathematical model learns its parameters from labeled samples with the main goal of making predictions on samples that the model has not seen before. Then, the classifier is used for assigning class labels to the testing instances in which the values of the predictor features are known, but the value of the class label remains unknown. Since the supervised classification is one of the tasks frequently developed by 'intelligent systems', it seems logical that a great number of techniques are based on AI and statistics. Meanwhile, unsupervised learning deals with problems that involve data without labels. In this case, the machine receives inputs but obtains neither outputs nor rewards from its environment (Ghahramani, 2003). Unsupervised approaches try to find trends and some kind of structure in the training data. That is, these approaches try to understand the origin of the data itself and to build representations of the inputs that can be used for decision-making, efficiently communicating the inputs to another machine or predicting future inputs. Clustering is a typical example of unsupervised learning. In unsupervised learning, the majority of the work can be considered as a learning process of a probabilistic model. When the scenario is not able to give the machine any supervision or reward, the machine can design a model that represents the probability distribution for a new input. This is achieved by just considering a previous useful input (e.g., stock prices or weather conditions). Probabilistic models that can be used in unsupervised learning are, among others: factor analysis, independent components analysis, principal components analysis, or Gaussians models.

There are situations where the supervised methods are not the best option. The first and most important is the high cost of labeling. Moreover, having all the training data fully labeled can be practically impossible. In these cases, it is common to start with supervised methods –using a small set of labeled data–, and then improve the model in an unsupervised way–i.e., using a larger set of unlabeled data.

AI and ML techniques are present in almost all sectors, including sports. In fact, most variables that can be quantified can also be predicted using AI and ML. The sport sector is full of quantifiable elements, which makes it ideal for the use of these techniques. For example, recruitment of players is one of the areas in sports where AI and ML are increasingly employed (Chavan, 2019; Herold et al., 2019; Musa et al., 2019; Ćwiklinski et al., 2021).

4. TYPES OF ANALYSIS IN TALENT ANALYTICS

Data analytic approaches, which are typically based on ML and statistics methods, can bring insights that are critical for improving operational and business outcomes of many organizations. These approaches play a role as powerful tools in the seeking and hiring young talent. Talent analytics is a systematic process that applies statistics, technology, and expertise to large sets of people data to discover the meaningful patterns that allow for sup- porting decision-making in recruitment. Three common types of analytics

–descriptive, predictive, and prescriptive– are used in TA, people and human resource analytics frameworks to measure efficiency, effectiveness, quality of recruitment, and impact (Necula and Strimbei, 2019). From descriptive to prescriptive analysis, not only the model increases in the complexity of the data being used, but also the analysis progressively gets more sophisticated. Each of the aforementioned types are described next:

Descriptive Analytics: reporting / visualization is the first step for carrying out statistical analyses, and it is used to describe the basic features of the data via applying to the collected data through the structured questionnaire (Marrybeth et al., 2019); it plays a relevant role in providing a view into activity –such as requisition volume, talent pool size, source of hires, etc.–, as well as to reveal the levels of activity and efficiency in each candidate generation.

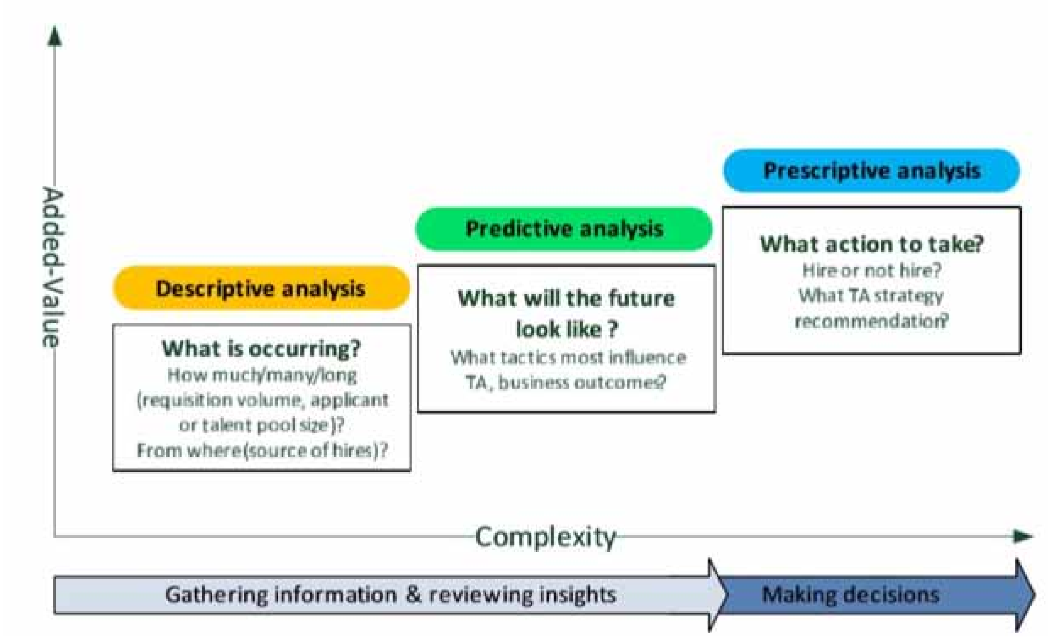
Predictive Analytics: uses data to find patterns and employs them to predict the future; it permits to identify statistical relationships between a set of activities and the expected outcomes, that will help to:

(i) forecast what will happen in the future or to explain the obtained outcome (like a candidate's likely cultural fit, level of performance, and retention); or *(ii)* notice potential talent shortages or skills gaps, and market availability (workforce planning); also, predictive analytics is mainly related to selection or rejection of candidate, acceptance of provided offer by selected candidates and root cause analysis for offer decline (Srivastava et al., 2015); predictive analysis finds answers for questions such as 'what will the future look like?', 'what tactics most influence business outcomes?', etc.; the result of this predictive analysis help shorten the entire recruiting process while improving the hiring process.

Prescriptive Analytics: goes a step further in the future and attempts to provide and suggest better decisions using data techniques such as decision modeling, ML, heuristic, simulation, neural networks; these decisions are based on the results provided by predictive analysis; it tries to evaluate the effect and impact of the provided decisions in order to modify them before implementation; it usually results in rules and recommendation for next steps (Attaran and Attaran, 2018); for example talent acquisition team receives the hire or not hire suggestions or strategy recommendation from predictive analysis.

A common data analytics framework used in TA is shown in Figure 4. The most common data-driven methodologies used in TA may classified in

Figure 4. Data analytics framework using in Talent Analytics



regression, classification, clustering, association rule mining, and anomaly detection. The reader interested in a more complete introduction to TA is referred to Davenport et al. (2010), which illustrates uses and describes the fundamentals to build a capacity in this domain, i.e.: access to high-quality data, enterprise orientation, analytical leadership, and strategic targets. In this context, Nocker and Sena (2019) discusses the advantages and costs induced (in terms of data governance and ethics) by using TA within an organization. The authors present a number of case studies to analyze the positive effect of TA usage in organizational decision-making processes and determine the key channels through which the TA adoption improve HR management and, subsequently, the whole organization function.

5. DATA ANALYTICS IN RECRUITMENT

A more data-driven culture is becoming increasingly popular among companies and governments. HR constitutes an example of a business' department that has dramatically changed during the last decades due to the use of data analytics methodologies and technologies. Indeed, companies are increasingly adopting sophisticated methods for study employee's data in order to improve the decision-making process, so they can strength their competitive advantage (Davenport et al., 2010). According to Rana et al. (2019), TA shows the potential within the decisions regarding hiring, training, improving productivity, and retaining talent, all of them with the main purpose to make a company more competitive.

Gartner, Inc.[7], points out that the volume of data and metrics available for HR has increased exponentially, while 70% of companies expect to in- crease the resources they dedicate to TA in the coming years. Even so, only 21% of the HR leaders believe that their organizations are effective at using talent data to inform business decisions. Dey and De (2015) points out five key areas where predictive analytics can create value in HR: *(i)* employee profiling and segmentation, employee attrition, and loyalty analysis; *(ii)* fore- casting of HR capacity and recruitment needs; *(iii)* appropriate recruitment profile selection; *(iv)* employee sentiment analysis; and *(v)* employee fraud risk management. Figure

5 lists the main TA applications, methodologies, and technologies. It is based on Kaur and Fink (2017), which offers a review of key approaches, competencies and tools, building on 22 interviews with academics, consultants and practitioners at 16 corporations, as well as on other TA experts.

5.1. Recruitment and Talent Acquisition

According to Wikipedia, recruitment may be defined as "the process of attracting, shortlisting, selecting, and appointing suitable candidates for jobs within an organization, and is a key function of human resource management." Another interesting definition is provided by Breaugh (2008), which defines external recruitment as "an employer's actions that are intended to:

(i) bring a job opening to the attention of potential job candidates who do not currently work for the organization; *(ii)* influence whether these individuals apply for the opening; *(iii)* affect whether they maintain interest in the position until a job offer is extended; and *(iv)* influence whether a job offer is accepted". Recruitment plays an essential role in determining the effectiveness of organizations, and it is composed of several sub-processes: *(i)* job analysis, which consists in documenting the knowledge, skills, abilities, and other characteristics (KSAOs) required for a job; *(ii)* sourcing, which is the process of attracting or identifying candidates; and finally *(iii)* screening and selection. Organizations apply recruitment strategies to identify hiring vacancy, establish timelines, and define goals throughout the recruitment process. Each organization designs its own strategies for recruitment, but there are frequent approaches such as using social networks for external recruitment advertisement or employing standard psychological tests, group discussion and a number of interviews to assess a variety of KSAOs. Thus, the recruitment process is complex and requires big amounts of effort and investment.

Bhattacharyya (2015) defines talent acquisition as a strategic approach aiming to identify, attract, and bring onboard top talent to meet dynamic business needs. According to this author, recruiting is more tactical and focuses mostly on immediate hiring needs, i.e., a process of filling the open positions. Figure 6 allows us to check the growing popularity of this research field. More specifically, it shows the evolution, from 2000 to 2019, in the number of works reported by the Web of Science when searching for: *(i)* "Talent acquisition" or "Talent recruitment" (which sometimes are used as synonyms in the literature) in "Social Science" (in orange); and *(ii)* the same but in "sports" (in blue). Clearly, there has been a positive and sustained trend during the last 10 to 15 years.

5.2. Studies on Data-driven Recruitment

Here, we describe a few recent and representative works on recruitment using data-intensive methodologies and techniques. For instance, Mohapatra and Sahu (2017) presents a case study to highlight the misconceptions in capacity planning methods (i.e. hiring process) and in the detection of bottle- necks in the hiring pipeline by using recruitment funnel technique and some

metrics to check efficiency of the aforementioned hiring process. Moreover, the authors depict appropriate sources of hiring based on the performance of candidates hired from those sources. They identify successful profiles in the company through computing correlations between selection parameters and performance scores. Finally, they provide a recruitment strategy and future road map for the case study. Kaur and Fink (2017) investigates what is re- quired to set up and run an effective TA function and the structures, system and skills that enable it. The authors discuss the answer to these questions through the collection and analysis of data from 22 interviews with academics, consultants, and different industries. The analysis shows that data infrastructure and reporting, advanced analytics, and organizational research are three components of a mature TA function.

Azar et al. (2013) provides a decision-making tool to help managers during the recruitment process. Authors claim that the tool, through using data mining techniques, is able to discover patterns

Figure 5. Scheme of talent analytics. Source: based on Kaur and Fink (2017)

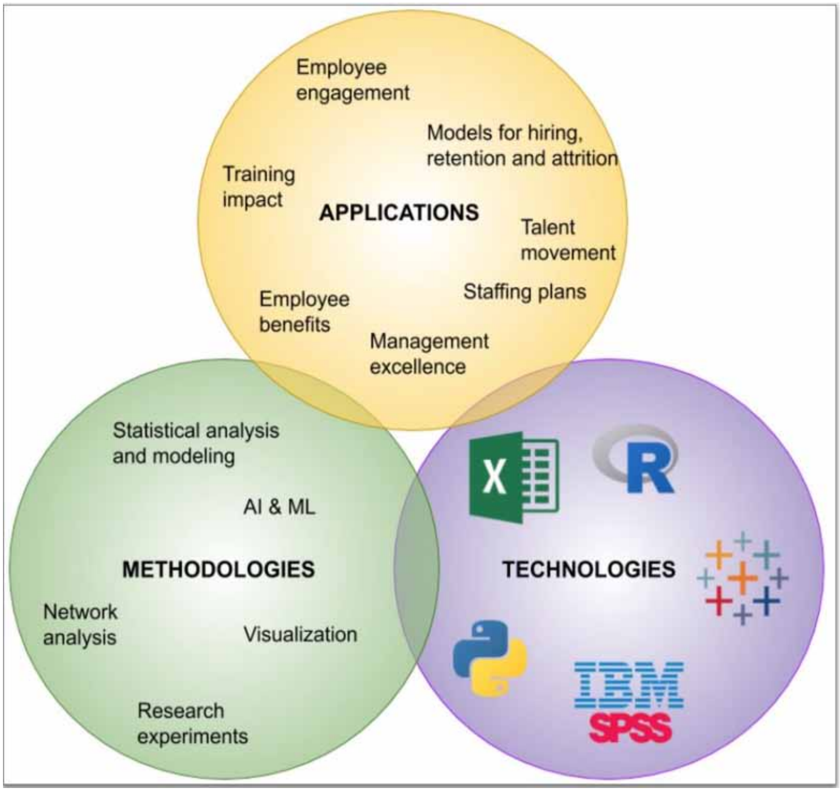
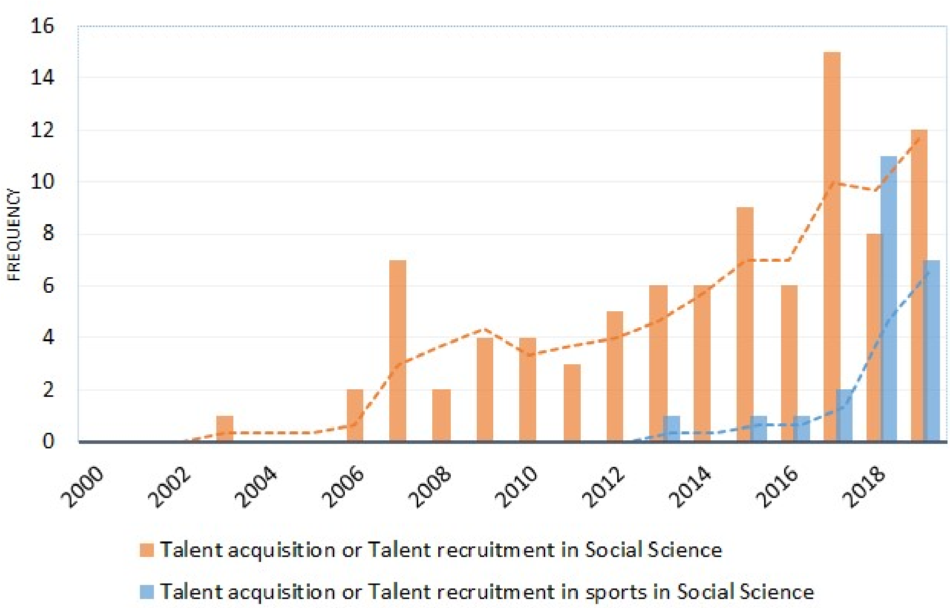


Figure 6. Evolution of the number of related works from 2000. Data source: Web of science. Dashed lines represent the tendencies calculated as moving averages.



of the relationship between employees' test scores and their performance at work. In particular, decision-trees are applied. Based on a case study of a commercial bank, they conclude that out of 26 variables only five variables (province of employment, education level, exam score, interview score, and work experience) have the most effect on the 'promotion score' target. Sivaram and Ramar (2010) apply *k*-means and fuzzy *k*-means clustering and decision-tree classification algorithms to the recruitment data of an IT industry, aiming to support the hiring decisions. The dataset used has more than 3,500 records. The goal is to maximize the accuracy to classify records in two categories: rejected or not rejected. The results (decision-trees) are discussed with the domain experts to deduce viable decision rules.

There are typically four main reasons for analyzing data: *(i)* describing it, *(ii)* explaining it, *(iii)* predicting new data, and *(iv)* optimizing a system based on this data. All the works reviewed in this section relate to the first two reasons, most also cover the third one, while optimization is not covered. Decision-trees and clustering are well-known, simple yet powerful data analytics methodologies which have been largely used.

6. DATA ANALYTICS IN TALENT ACQUISITION IN SPORTS

Sports have become one of the major businesses around the world. Hence, and since statistical analysis is an essential component in data-driven decision-making, the use of data analytics techniques in sports is raising interest (Baker and Kwartler, 2015). Sports analytics includes managing data, using predictive analysis, and informing decision-makers to provide a competitive advantage (Alamar, 2013). In their extensive literature review about ML techniques applied to baseball, Koseler and Stephan (2017) found out that the two algorithms that predominate are: *(i)* the support vector machines for classification problems; and *(ii)* the *k*-nearest neighbors for both classification and regression problems. According to its repercussion in time, there is a range of strategic, tactical, and operational decisions that can be made in the sports field. The decisions addressed in this paper are within the strategic level and related to talent acquisition. As these decisions can have repercussions on future performance, they also have features in common with other drawn-out decision-making processes like financial planning (Ofoghi et al., 2013; Brown et al., 2010). The applications are based on knowledge of past experiences, and are suited to statistical analysis and data mining (Bhandari et al., 1997). One should notice that these decisions can also be influenced by common errors and limitations of human information processing. Hence, in order to overcome such errors, researchers have used statistical techniques in modeling sports, data mining and machine learning methods (Koseler and Stephan, 2017; Edelmann-Nusser et al., 2002). Considering the aforementioned issues, Baker and Kwartler (2015) states that coaches and sport managers of the future must be skillful in the strategic application of the analysis to specific situations. They also must know how to collect appropriate data and have to be able to analyze the data accordingly. The most significant contributions within the academy have focused on those sports that have more followers and generate more money, i.e., Ameri- can football, basketball, and soccer (depending on the country). These teams compete around the world, the matches are broadcast to millions of homes and are worth billions in the betting market. Hence, modeling a predictive system for these matches output and for their recruitment process is not merely of interest in academia, but also hugely significant in economic terms (Baboota and Kaur, 2019). In fact, athlete recruiting optimization applies for multiple sports but, considering the number of leads, it is an especially severe problem for college football (Cokins and Schrader, 2017).

Dronyk-Trosper and Stitzel (2017) has used an extensive panel set covering 13 years of games along with a two-stage least squares approach to investigate the effects of the recruiting process on team performance within the National Football League (NFL), a professional American football league. The obtained evidence shows the benefits of recruiting taking into account team-specific effects –i.e., considering the complementarities of the recruits in the team. That may indicate that team success can be derived from the ability of teams to tackle and improve their recruits more than their ability to employ each athlete's raw abilities. Supporting this idea, Mankin et al. (2019) and Ramirez et al. (2017) have shown, by using multiple linear regression models, that the integration / recruitment and the training of athletes considering the idiosyncrasies of the team leads to better results. Moreover, those results would be regardless of the initial assessment ratios of the athletes and their initial raw skills.

The University of Virginia has developed a three-part consecutive study for recruiting football players and predict performance (Beckwith et al., 2019; Peng et al., 2018; Walter et al., 2017). By chronological order, the first part tries to analyze the staff's processes and metrics in order to find areas where data analytics can be effectively applied. They developed a model that forecasts the academic potential, the commitment to the university of the candidate, and the possible future competitors. The tool permits coaches to discover athletes, from lower position in the rankings, who are likely to outperform their initial ranking. In the second part of the study, Peng et al. (2018) use a dataset from over 53, 000 football recruits and more than 200 predictive attributes to model the four aspects of collegiate football recruiting defined by the university coaches. Especially, those who: *(i)* would succeed on the field at the collegiate level and will meet the academic standards; *(ii)* will fit the sports culture of the university; and *(iii)* would commit to the football team. Hence, this new tool helps coaches to recruit those athletes that really are capable to commit to the university culture, preventing them from wasting time and money in the recruitment process. In the last part, Beckwith et al. (2019) created a ranked model (point model), based on existing NFL models, in order to evaluate the team's performance and identify possible improvement areas. Finally, they have integrated the point model into a report in order to provide a better assessment of the opponent's performance. With this new tool, the coaches and the staff will be able to spend less time identifying the weaknesses and the strengths from other contestants and will have more time to exploit it.

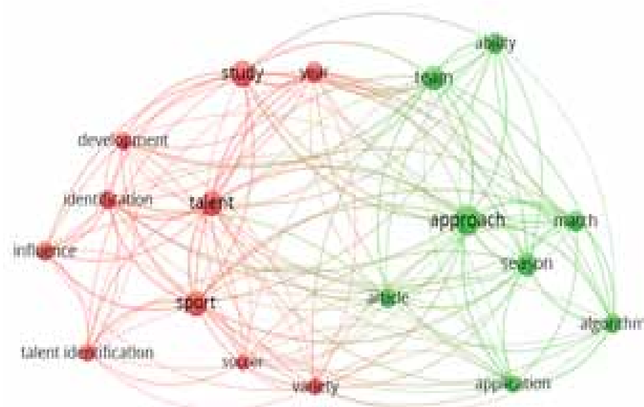
Relying on data of the Football League Championship, in Barron et al. (2020), the authors try to identify key factors from position-specific performance that will help predict the out-field players league status. For this case study, 966 players were allocated within three main categories based on where they developed most of their match time in the following season. For doing so, 340 career paths, biographical, and esteem variables were considered and analysed using a Stepwise Artificial Neural Network. The resulted models correctly predicted between 72.7% and 100% of test cases.

Inspired by the 'Moneyball' strategy, Gavi˜ao et al. (2020) focuses on finding "bargains" within the players market, that is hiring players with an outstanding trajectory ahead (because of their capabilities) but relatively low market value by selecting underrated players who have complementary skills. The authors suggest the co-joint use of four different algorithms based on the exploration of also different probabilistic preferences (i.e., ranking, classification, dynamic evaluation and regularity analysis). In that particular case, the procedure was implemented in the soccer player market.

In other sports, like cycling, Ofoghi et al. (2013) depicts the implementation of ML techniques for assisting cycling experts in the decision-making process for athlete selection and strategic planning in a track cycling competition (including the selection and formation of the teams). Using a probabilistic analysis, the authors present a model of performance prediction that provides with supporting information for assisting coaches in the process of making strategic and tactical decisions during the competition. In shot put, Babbitt (2019) depicts a model for helping in shot putters female recruitment process at the NCAA Division I Championship level. In particular, this model aims to predict the time needed by a high school female shot putter to contribute a score at the NCAA Championship meet based on her personal best high school performance. The results show (for a sample of 63 shot putters) statistical significance correlations between high school and collegiate performance for the first three years of collegiate competition. In women artistic gymnastics, Nassib et al. (2020) investigates the most appropriate aptitudes for top-level sporting results to identify physical profile of talent identified women's artistic gymnastics. 48 gymnastics participated in the study. According to the results, strength, power, flexibility, and coordination seem to be essential for good performance.

Finally, Figure 7 shows a map based on text data from the titles and abstracts of the papers reviewed in this work. It has been built with the software VOSviewer version 1.6.16 (Van Eck and Waltman, 2010). Terms are represented by their label and a circle. The size of the label and the circle of a term is determined by the weight of that term. Thus, the higher the weight of a term, the larger the label and the circle of the term. The color of a term is determined by the cluster to which the term belongs. Lines between items represent links. The distance between two words in the figure reveals the relatedness of the words in terms of co-occurrence links. In general, the closer two terms are located to each other, the stronger their relatedness. For example, it can be concluded that the term ‘influence’ appears in many works and is related to ‘talent identification’ and ‘development’.

Figure 7. Map based on text data from the titles and abstracts



7. ILLUSTRATING CONCEPTS WITH A CASE STUDY

In order to illustrate some of the aforementioned concepts, this section describes a series of numerical experiments based on real-life data. The data employed corresponds to a publicly available database on Euro- pean soccer. We have downloaded the database from the following website: https://www.kaggle.com/hugomathien/soccer. Then, the table “Player Attributes” has been converted to a CSV file. This table contains a total of 42 columns (variables), each of them measuring characteristics of 183, 978 players (rows) from different European soccer leagues. The initial dataset contains some missing values, which have been properly removed. The final dataset contains 42 variables and 180, 354 players. These variables reflect on players’ characteristics such as ‘overall rating’, ‘potential’, ‘preferred foot’, ‘attacking work rate’, ‘defensive work rate’, ‘crossing’, ‘finishing’, ‘heading

Table 1: Characteristics for each of the clusters found.

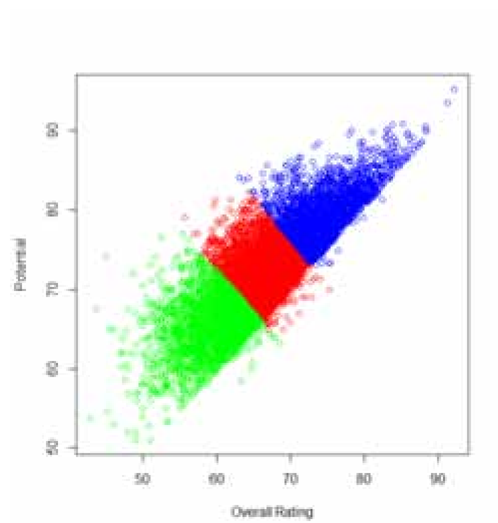
Cluster	Acceleration	Agility	Aggression	Vision	Strength
1-Green	62.425	60.075	53.925	48.168	63.257
2-Red	66.666	64.632	60.135	55.785	67.006
3-Blue	70.212	68.640	64.540	64.241	69.690

accuracy', 'short passing', 'volleys', 'dribbling', etc. For many managers, it could be interesting to have a multiple regression model that helps them to predict the estimated overall rating of new players based on a reduced number of variables, since this model could help them to quickly rank young players and bid for them before they become too expensive. With this in mind, we have developed different regression models, with the aid of Phyton Statistics version 3.4 software. While some of them are able to reach a nice predictive performance (adjusted $R^2 = 84.4\%$) by employing 24 variables, it is also possible to find models that offer a similar predictive capacity with fewer variables. In particular, one can consider the following one (adjusted $R^2 = 81.5\%$): *overall rating = -1.0446 + 0.40401 potential + 0.188931 ball-control + 0.26822 reactions + 0.118062 strength + 0.156618 gk-diving*. Notice that this is a predictive model that only employs 5 variables and, still, can be used in practice to identify promising young players and rank them. Another interesting application could be to group players in such a way that the members of the same group have similar characteristics. Figure 8 should the results of running the *K*-means algorithm to create 3 groups based on the overall rating and potential, which have a high correlation, with the aid of R-4.1.0 software. Table 1 display some characteristics per cluster, which are scores (ranging from 1 to 97) for acceleration, agility, aggression, vision, and strength. Clustering techniques may be applied to propose different training sessions, for example.

8. TRENDS, CHALLENGES, AND OPEN RESEARCH LINES

Data analytics in sports talent acquisition constitutes a relatively recent field, which is constantly growing as a consequence of a higher data-based decision-making orientation in companies as well as an increasing number of related methodologies, technologies, data sources, and successful cases. A discussion based on the analysis of the existing works is provided next.

Figure 8. Clusters of players performed with the K-means algorithm (k = 3)



8.1. Trends

Most works reviewed focus on describing, explaining, and predicting data related talent acquisition, not considering optimization. All the works study either football, soccer, or cycling teams –i.e., those sports with more followers and that generate more benefits. Classical statistics techniques (such as multiple linear regression) are popular, probably because it is easy to under- stand their fundamentals and to interpret and explain the results. However, ML is becoming increasingly used too. Also, it is worthwhile to highlight the increasing interest of universities in using these techniques to identify talent, the importance of not only considering features of the individual but also of the team (where applies), and the need to make predictions regarding children and young adults.

8.2. Challenges

Several challenges emerge from the use of data analytics in talent acquisition in general. First of all, while companies and governments are becoming increasingly aware of the value of data and the need of having data scientists, many HR departments have not made the transition yet and keep relying on more traditional practices based on intuitions and routines. According to Giuffrida (2014), the following hurdles are common: *(i)* total potential re- mains a mystery, since many HR departments are only using analytics for "basic" reporting and compliance purposes; *(ii)* full buy-in or support is hard to obtain; *(iii)* the technology solution is too rigid or cumbersome; and *(iv)* TA get pigeonholed as strictly an "operational" function when, instead, TA should support managerial and strategic functions.

As discussed before, there are plenty of technologies that eases the implementation and use of a wide range of methodologies. However, designing a data governance program may be challenging. Data governance is the overall management of the availability, usability, integrity and security of data used in an organization. For instance, ensuring that our algorithm does not take decisions without considering legal and ethical issues. While it is obvious that algorithms should not consider variables such as social class, sex orientation, race, or nationality, we must check that the results are not biased. Finally, the fact that professional or semi-professional sport teams will not disclosure their talent acquisition approach, makes it impossible for us to precisely assess the role of data analytics and is difficult for smaller teams to try to imitate the existing approaches.

8.3. Open research lines

It is possible to highlight several lines for future research. Firstly, there is a higher growing quality, quantity, and variety of data in sports. For instance, sensors are becoming increasingly affordable and powerful, allowing us to capture users vitals and emotional stress level, so trainees / recruiters can prove their ability to perform under pressure, as well as motion capture, to establish baselines, create personalized training programs, and track improvement over time. As more data is available, more advanced method- ologies (e.g., deep learning) may be applied to obtain better results. It is also worth-mentioning that software options (either free or paid) are becoming increasingly powerful and user-friendly, with very active communities supporting them. In addition, during the next years it is expected that more sports will follow the steps of football, soccer, and cycling, and start to apply data analytics for talent acquisition. Most works have focused on selecting the best candidates, considering expected performance results. However, a few consider a multi-objective approach aiming to identify players with an excellent performance, but also highly committed to the team (if applies) and who can fit perfectly with team / university / country cultures. Indeed, plenty of objectives could be proposed: expected benefits from merchandising, risk of leaving the group, risk of getting injured, etc. Similarly, many restrictions may be considered: budget limitations, types of players, etc.

9. CONCLUSION

Over recent years, businesses have drifted towards a data-driven culture. This is the result of the new data era, in which businesses may obtain large amounts of data (from different sources, with different formats, and with a greater quality) and analyze it in order to support decision-making.

In the sports field, the use of data analytics to acquire talent is becoming increasingly popular in most professional and semi-professional teams in football, soccer, cycling, etc. While in the past managers heavily rely on experts' opinion to select the most promising players, today they are guided by data scientists which employ data-intensive methods and techniques to identify the best candidates and predict their future evolution as professional players.

In this context, this paper has reviewed existing work on the use of data analytics in talent acquisition in the sports field. The literature is still scarce but increasing, employs classical statistical and machine learning approaches mainly, and is focused on those sports that generate more benefits: football, basketball, and soccer. A numerical case study, based on real-life data from the European soccer database with over 183, 978 players, has been developed to illustrate just two of the many business analytics techniques that can be used in sport talent acquisition. Hence, a multiple regression model has been built. The model allows us to predict the overall rating of a new player based on just 5 key variables. Likewise, a machine learning method has been utilized to generate a cluster of the players. The paper has also identified and discussed the main trends, challenges, and open research lines in this applied research area.

In our view, one of the most interesting lines for future work is the combination of statistical / machine learning methods with optimization- simulation approaches, since this hybridization can provide effective support for informed decision-making during talent acquisition processes.

REFERENCES

Alamar, B. C. (2013). *Sports analytics: A guide for coaches, managers, and other decision makers*. Columbia University Press.

Apostolou, K., & Tjortjis, C. (2019). Sports analytics algorithms for performance prediction. In *10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE.

Attaran, M., & Attaran, S. (2018). Opportunities and challenges of implementing predictive analytics for competitive advantage. *International Journal of Business Intelligence Research*, 9.

Azar, A., Rajaeian, A., Sebt, M. V., & Ahmadi, P. (2013). A model for personnel selection with a data mining approach: A case study in a commercial bank. *SA Journal of Human Resource Management*, *11*(1), 1–10. doi:10.4102/sajhrm.v11i1.449

Babbitt, D. G. (2019). A development model to guide the recruiting of female shot putters at the NCAA Division I Championship level. *Journal of Sports Analytics*, *5*(3), 181–190. doi:10.3233/JSA-180275

Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English premier league. *International Journal of Forecasting*, *35*(2), 741–755. doi:10.1016/j.ijforecast.2018.01.003

Baker, J., Cobley, S., Schorer, J., & Wattie, N. (2017). Talent identification and development in sport: An introduction. In *Routledge Handbook of Talent Identification and Development in Sport* (pp. 1–8). Routledge. doi:10.4324/9781315668017-1

Baker, R. E., & Kwartler, T. (2015). Sport analytics: Using open source logistic regression software to classify upcoming play type in the NFL. *Journal of Applied Sport Management*, 7.

Barron, D., Ball, G., Robins, M., & Sunderland, C. (2020). Identifying playing talent in professional football using artificial neural networks. *Journal of Sports Sciences*, *38*(11-12), 1–10. doi:10.1080/02640414.2019.1708036 PMID:31941425

Bassi, L. (2011). Raging debates in HR analytics. *People and Strategy*, *34*, 14.

Battaglia, P., Pascanu, R., Lai, M., & Rezende, D. J. (2016). Interaction networks for learning about objects, relations and physics. Advances in Neural Information Processing Systems, 4502–4510.

Beckwith, G., Callahan, T., Carlson, B., Fondren, T., Harris, R., Hoege, J., Martin, T., Menna, C., Summer, E., & Scherer, W. (2019). Systems analysis for university of virginia football recruiting and performance. In *2019 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE. doi:10.1109/SIEDS.2019.8735611

Bergkamp, T. L., Niessen, A. S. M., den Hartigh, R. J., Frencken, W. G., & Meijer, R. R. (2019). Methodological issues in soccer talent identification research. *Sports Medicine (Auckland, N.Z.)*, *49*(9), 1–19. doi:10.1007/s40279-019-01113-w PMID:31161402

Berri, D. J., & Brook, S. L. (2010). On the evaluation of the "most important" position in professional sports. *Journal of Sports Economics*, *11*(2), 157–171. doi:10.1177/1527002510363097

Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., & Ramanujam, K. (1997). Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, *1*(1), 121–125. doi:10.1023/A:1009782106822

Bhattacharyya, D. K. (2015). *The magnetic organization: Attracting and retaining the best talent*. SAGE Publications India.

Breaugh, J. A. (2008). Employee recruitment: Current knowledge and impor- tant areas for future research. *Human Resource Management Review*, *18*(3), 103–118. doi:10.1016/j.hrmr.2008.07.003

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C. H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, *96*(3), 673–695. doi:10.1177/1077699019859901

Brown, M. R., Delau, S., & Desgorces, F. D. (2010). Effort regulation in rowing races depends on performance level and exercise mode. *Journal of Science and Medicine in Sport*, *13*(6), 613–617. doi:10.1016/j.jsams.2010.01.002 PMID:20227342

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in european soccer. *Journal of Sports Economics*, *14*(6), 606–628. doi:10.1177/1527002511435118

Chavan, A. (2019). *Recruitment of Suitable Football Player by using Machine Learning Techniques* (Ph.D. thesis). National College of Ireland.

Cokins, G., & Schrader, D. (2017). The sports analytics explosion. *OR/MS Today, 44*, 28–33.

C^ot'e, J. (1999). The influence of the family in the development of talent in sport. *The Sport Psychologist*, *13*(4), 395–417. doi:10.1123/tsp.13.4.395

C´wiklinski, B., Gie-lczyk, A., & Chora's, M. (2021). Who will score? a machine learning approach to supporting football team building and transfers. *Entropy (Basel, Switzerland)*, *23*(1), 90. doi:10.3390/e23010090 PMID:33435241

Davcheva, P. (2014). Identifying sports talents by social media mining as a marketing instrument. In *Annual SRII global conference*. IEEE. doi:10.1109/SRII.2014.38

Davenport, T. H., Harris, J., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, *88*, 52–58. PMID:20929194

Davids, K., & Arau'jo, D. (2019). *Innate talent in sport: Beware of an organismic asymmetry–comment on Baker & Wattie. In Current Issues in Sport Science*. CISS.

De Bosscher, V., & De Rycke, J. (2017). Talent development programmes: A retrospective analysis of the age and support services for talented athletes in 15 nations. *European Sport Management Quarterly*, *17*(5), 590–609. doi:10.1080/16184742.2017.1324503

Deepa, S., & Devi, B. A. (2011). A survey on artificial intelligence approaches for medical image classification. *Indian Journal of Science and Technology*, *4*(11), 1583–1595. doi:10.17485/ijst/2011/v4i11.35

Dey, T., & De, P. (2015). *Predictive analytics in HR: A primer*. A White Paper, Tata Consultancy Services. Available: http://www.tcs.com/SiteCollectionDocuments/White-Papers/Predictive- Analytics-HR-0115-1.pdf

Dronyk-Trosper, T., & Stitzel, B. (2017). Lock-in and team effects: Recruiting and success in college football athletics. *Journal of Sports Economics*, *18*(4), 376–387. doi:10.1177/1527002515577885

Edelmann-Nusser, J., Hohmann, A., & Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, *2*(2), 1–10. doi:10.1080/17461390200072201

Ford, P. R., & Williams, A. M. (2017). Sport activity in childhood: Early specialization and diversification. In *Routledge Handbook of Talent Identifi- cation and Development in Sport* (pp. 116–132). Routledge. doi:10.4324/9781315668017-9

Fried, G., & Mumcu, C. (2016). *Sport analytics: A data-driven approach to sport business and management*. Taylor & Francis. doi:10.4324/9781315619088

Gandelman, N. (2009). Selection biases in sports markets. *Journal of Sports Economics*, *10*(5), 502–521. doi:10.1177/1527002509332237

Gaviao, L. O., Sant'Anna, A. P., Alves Lima, G. B., & de Almada Garcia, P. A. (2020). Evaluation of soccer players under the moneyball concept. *Journal of Sports Sciences*, *38*(11-12), 1221–1247. doi:10.1080/02640414.2019.1702280 PMID:31876264

Gerrard, B. (2017). The role of analytics in assessing playing talent. In Routledge Handbook of Talent Identification and Development in Sport. Routledge. doi:10.4324/9781315668017-30

Ghahramani, Z. (2003). Unsupervised learning. In *Summer School on Ma- chine Learning* (pp. 72–112). Springer.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelli- gence. *Nature*, *521*(7553), 452–459. doi:10.1038/nature14541 PMID:26017444

Giuffrida, M. (2014). Unleashing the power of talent analytics in federal government. *Public Management*, *43*, 7.

Hanlon, C., Morris, T., & Nabbs, S. (2014). Program providers' perspective: Recruitment and retention strategies for women in physical activity pro- grams. *Sport Management Review*, *17*(2), 133–144. doi:10.1016/j.smr.2013.04.001

Harris, J. G., Craig, E., & Light, D. A. (2011). Talent and analytics: New ap- proaches, higher ROI. *The Journal of Business Strategy*, *32*(6), 4–13. doi:10.1108/02756661111180087

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, *14*(6), 798–817. doi:10.1177/1747954119879350

Joshi, A. V. (2019). *Machine Learning and Artificial Intelligence*. Springer.

Kaur, J., & Fink, A. A. (2017). *Trends and practices in talent analytics*. Society for Human Resource Management (SHRM)-Society for Industrial-Organizational Psychology (SIOP) Science of HR White Paper Series 20.

Kim, Y., Bui, K. H. N., & Jung, J. J. (2019). Data-driven exploratory approach on player valuation in football transfer market. *Concurrency and Computation*, *33*(3), e5353. doi:10.1002/cpe.5353

Koseler, K., & Stephan, M. (2017). Machine learning applications in baseball: A systematic literature review. *Applied Artificial Intelligence*, *31*(9-10), 745–763. doi:10.1080/08839514.2018.1442991

Kotsiantis, S.B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering, 160*, 3–24.

Li, D., & Du, Y. (2017). *Artificial intelligence with uncertainty*. CRC Press. doi:10.1201/9781315366951

Loland, S. (2015). Against genetic tests for athletic talent: The primacy of the phenotype. *Sports Medicine (Auckland, N.Z.)*, *45*(9), 1229–1233. doi:10.1007/s40279-015-0352-5 PMID:26121951

Mankin, J.A., Rivas, J., & Jewell, J.J. (2019). The effectiveness of college foot- ball recruiting ratings in predicting team success: A longitudinal study. *Research in Business and Economics Journal, 14*, 1–19.

Marrybeth, J., Brackz, J., Hadson, W., & Zlong, M. (2019). Talent acquisition and talent engagement practices significant impact over employee satisfaction. *International Research Journal of Management, IT and Social Sciences, 6*, 244–252.

Martin, L. (2015). Is socioeconomic status a contributing factor to tennis players' success. *Journal of Medicine and Science in Tennis*, *20*, 116–121.

Mohapatra, M., & Sahu, P. (2017). Optimizing the recruitment funnel in an ITES company: An analytics approach. *Procedia Computer Science*, *122*, 706–714. doi:10.1016/j.procs.2017.11.427

Musa, R. M., Taha, Z., Majeed, A. P. A., & Abdullah, M. R. (2019). *Machine learning in sports: identifying potential archers*. Springer. doi:10.1007/978-981-13-2592-2

Nassib, S. H., Mkaouer, B., Riahi, S. H., Wali, S. M., & Nassib, S. (2020). Pre- diction of gymnastics physical profile through an international program evaluation in women artistic gymnastics. *Journal of Strength and Conditioning Research*, *34*(2), 577–586. doi:10.1519/JSC.0000000000001902 PMID:31386634

Necula, S. C., & Strimbei, C. (2019). People analytics of semantic web human resource r'esum'es for sustainable talent acquisition. *Sustainability*, *11*(13), 3520. doi:10.3390/su11133520

Nocker, M., & Sena, V. (2019). Big data and human resources management: The rise of talent analytics. *Social Sciences*, *8*(10), 273. doi:10.3390/socsci8100273

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Dwyer, D. (2013). Supporting athlete selection and strategic planning in track cycling omnium: A statistical and machine learning approach. *Information Sciences*, *233*, 200–213. doi:10.1016/j.ins.2012.12.050

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). Playerank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology*, *10*(5), 10. doi:10.1145/3343172

Peng, K., Cooke, J., Crockett, A., Shin, D., Foster, A., Rue, J., Williams, R., Valeiras, J., Scherer, W., & Tuttle, C. (2018). Predictive analytics for university of virginia football recruiting. In *Systems and Information Engineering Design Symposium (SIEDS)*. IEEE. doi:10.1109/SIEDS.2018.8374745

Pickering, C., Kiely, J., Grgic, J., Lucia, A., & Del Coso, J. (2019). Can genetic testing identify talent for sport? *Genes*, *10*(12), 972. doi:10.3390/genes10120972 PMID:31779250

Pifer, N. D. (2019). Data analytics in football: Positional data collection, modeling, and analysis. *Journal of Sport Management*, *33*(6), 574–574. doi:10.1123/jsm.2019-0308

Ramirez, M. C., Viloria, A., Mun˜oz, A. P., & Posso, H. (2017). Application of multiple linear regression models in the identification of factors affecting the results of the Chelsea football team. *International Journal of Control Theory and Applications*, *10*, 7–13.

Rana, G., Sharma, R., & Goel, A. K. (2019). Unraveling the power of talent analytics: Implications for enhancing business performance. In *Business Governance and Society* (pp. 29–41). Springer. doi:10.1007/978-3-319-94613-9_3

Rimler, M. S., Song, S., & Yi, D. T. (2010). Estimating production efficiency in men's ncaa college basketball: A bayesian approach. *Journal of Sports Economics*, *11*(3), 287–315. doi:10.1177/1527002509337803

Sivaram, N., & Ramar, K. (2010). Applicability of clustering and classification algorithms for recruitment data mining. *International Journal of Computers and Applications*, *4*(5), 23–28. doi:10.5120/823-1165

Srivastava, R., Palshikar, G., & Pawar, S. (2015). Analytics for improving talent acquisition processes. *Proceedings of 4th international conference on advanced data analysis, business analytics and intelligence (ICADABAI 2015)*.

Tveter, D. (1997). *The pattern recognition basis of artificial intelligence*. IEEE Press.

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports Medicine (Auckland, N.Z.)*, *38*(9), 703–714. doi:10.2165/00007256-200838090-00001 PMID:18712939

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. doi:10.1007/s11192-009-0146-3 PMID:20585380

Walter, L., Citera, A., Knowles, K., Lowen, M., Oldenburg, C., Shahin, H., Scherer, W., & Tuttle, C. (2017). Implementation of a recruit visualization tool for UVA football. In *Systems and Information Engineering Design Symposium (SIEDS)*. IEEE. doi:10.1109/SIEDS.2017.7937710

Webb, T., Dicks, M., Brown, D. J., & O'Gorman, J. (2020). An exploration of young professional football players' perceptions of the talent development process in England. *Sport Management Review*, *23*(3), 536–547. doi:10.1016/j.smr.2019.04.007

Webborn, N., Williams, A., McNamee, M., Bouchard, C., Pitsiladis, Y., Ahmetov, I., Ashley, E., Byrne, N., Camporesi, S., Collins, M., Dijkstra, P., Eynon, N., Fuku, N., Garton, F. C., Hoppe, N., Holm, S., Kaye, J., Klissouras, V., Lucia, A., & Wang, G. et al. (2015). Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement. *British Journal of Sports Medicine*, *49*(23), 1486–1491. doi:10.1136/bjsports-2015-095343 PMID:26582191

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, *18*(9), 657–667. doi:10.1080/02640410050120041 PMID:11043892

ENDNOTES

[1] https://en.wikipedia.org/wiki/Moneyball_(film)
[2] www.21stclub.com
[3] https://barcainnovationhub.com/es/event/barca-sports-analytics-summit-2019/
[4] https://wyscout.com
[5] https://www.scisports.com
[6] www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/data-powered-football.html
[7] https://www.gartner.com/en/human-resources/insights/talent-analytics

Rocio de la Torre is an Assistant Professor in the Department of Business Management at the Public University of Navarre (Spain). She is also a researcher from INARBE-Institute for Advanced Research in Business and Economics. She holds a Ph.D. and a Bachelor's Degree in Industrial Engineering from the Universitat Politecnica de Catalunya. Her major research areas are mathematical programming for strategic planning decisions in knowledge-intensive organizations (KIOs) and supply chain design.

Laura O. Calvet is a Lecturer of Statistics in the Computer Science Dept. at the Universitat Oberta de Catalunya (UOC) and Lecturer of Mathematics & Project Management at the Escola Universitària Salesiana de Sarrià (EUSS). She holds a M.S. in Applied Statistics and Operations Research completed at Universitat Politècnica de Catalunya (UPC) & Universitat de Barcelona (UB) and a Ph.D. in Network and Information Technologies completed at the UOC. She is a member of the ICSO@IN3 research group. Her main lines of research are: - Design of optimization algorithms relying on the use of metaheuristics, machine learning and/or simulation applied to sustainable logistics & computing - Applied statistics & economics: applications in health, disaster management, & learning.

David Lopez-Lopez is an academic collaborator at ESADE. He holds a PhD in digital transformation, and a joint MBA from ESADE (Spain) and the University of Duke (USA). He is also managing partner in Fhios, that employs more than 180 consultants.

Angel A. Juan is a Full Professor of Operations Research & Industrial Engineering in the Computer Science Dept. at the Universitat Oberta de Catalunya (Barcelona, Spain). He is also the Director of the ICSO research group at the Internet Interdisciplinary Institute and Lecturer at the Euncet Business School. Dr. Juan holds a Ph.D. in Industrial Engineering and an M.Sc. in Mathematics. He completed a predoctoral internship at Harvard University and postdoctoral internships at Massachusetts Institute of Technology and Georgia Institute of Technology. His main research interests include applications of simulation, metaheuristics, and machine learning methods in different fields, including: logistics & transportation, finance, and smart cities. He has published over 110 articles in JCR-indexed journals and over 250 papers indexed in Scopus.