Research Article

Sport Analytics: A Review

Nitin Singh*

Indian Institute of Management Ranchi, Meur’s Road, Ranchi, 834008, India

ARTICLE INFO

Article History
Received XX XXX XXXX
Accepted XX XXX XXXX

Keywords
Sport analytics
systematic review
business research
sport management

ABSTRACT

This paper offers a systematic review of research in the emerging field of sport analytics, which is receiving increasing attention in practice and research circles. The purpose of this study is to understand the state of the research on the application of sport analytics and its emerging sub-fields in business. Various publications are analysed by applying a structured search in databases; which are classified based on business context and analytical methodology. The discussion presents key findings and the synthesis of review in sport analytics.

© 2020 The Authors. Published by Atlantis Press International B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (http://creativecommons.org/licenses/by-nc/4.0/).

1. INTRODUCTION

Over the past years, a variety of data-capturing technologies have become available in sport business [1]. These technologies allow sport management businesses to capture and collect data on games, bidding, bookmaker odds, playing styles, scores, and many other sport attributes [2]. Such a repository of data allows firms to garner invaluable insights through the leveraging of data analytics. There have also been several discussions around this issue in literary and business circles [3–7]. The studies suggest that a data-driven approach to sport business and marketing is an interesting area to investigate. In this context, data analytics could be of immense value.

Data analytics in sport has become an integral part of sport business [1]. Data mining-based models are also being explored in sport [8]. Other forms of analytical techniques being explored are, page rank models [9], numerical algorithms, machine learning [9–11]. There are other techniques and methods also explored as well.

Recently, the Fédération Internationale de Football Association approved the use of Electronic Performance and Tracking Systems (EPTS) during competition, which made novel data related to physical player performance available. Now, the physical data collected using EPTS, from both training sessions and live matches, can be used to evaluate the extent to which match performance can be predicted [12]. Scholars have proposed a normalised root mean square error metric for analysing the results obtained from the application of machine-learning algorithms to the data collected for various physical variables. Specific physical variables can also act as representatives of several other variables, which are in turn highly correlated, to further reduce the number of variables that must be periodically analysed by coaches [12]. Moreover, continuously growing sports platforms related to games incentivize bookies and bettors to bet on match results as a game changes ball by ball. Hence, attempts have been made to predict match results based on historical match data [13].

Further, the emotional expressions of sports team members, and their correlation with the team’s performance, can be analysed to draw conclusions about the psychological mind-set of players [14]. Design of experiments, is applied to frame experiments explaining the variation under certain conditions to predict the outcome in certain variables [7]. Finally, researchers have introduced variables such as possession (ball occupation) and territory (dominance of territory) and a novel visual analytics system to analyse tactical transitions in a continuous ball match [15].

The rest of the paper is organised as follows. Section 2 presents the discussion on systematic review. Section 3 presents the methodology including the process of literature delimitation and classifications. The analysis and systematic review of sport analytics is presented in Section 4. Key findings and synthesis of the analysis is presented in Section 5. It is followed by Conclusion in Section 6.

2. SYSTEMATIC REVIEW

Sport analytics has been receiving significant research attention and studies have suggested that sport analytics could be used to a greater degree in businesses engaged in sport [1,3–7]. A review in this field is of relevance as it can lead to understanding the state of research and classifications of study within this topic [16]. There have been scoping reviews in sport management in the areas of sport governance and spectator sport [18,19]. However, a Systematic Literature Review (SLR) is an appropriate approach as it enables the researcher to structure a research field and understand the state of research and emerging research themes [19]. This, in
turn, discovers the application of the area and provides guidance toward understanding or evaluating the area [16,20]. SLR is also considered to be a valid approach as it is an integral part of research and a vital process in structuring a research field [21]. It identifies the conceptual content of the field and found to be an effective approach to understanding the intellectual foundations of a research field and to classify studies [22].

With this context, the motivation for this study is to present an integrated view of the literature on different aspects of sport analytics so as to facilitate further research, study, and practice. We conduct a meta-analysis or systematic review where we systematically search for relevant papers that discuss or deliberate on the analytical or quantitative modelling in sport. In the next stage, we identify the studies (papers/books/articles) with designs according to a specific criterion. In the last stage, we attempt to draw conclusions (synthesis) on this topic of sport analytics based on an analysis of the reviewed studies.

We follow the review method suggested by Searcy and Mentzer [23], which can be classified as an archival research method. Literature review has been found to be an effective approach to understand the intellectual foundations of concepts and classify related research [22]. Second, we analyse and classify the literature based on context and methodology. A taxonomical approach to context and methodology has been adopted in the review studies and found to be immensely helpful to researchers in understanding the state of the art in research [22]. Third, the context classification is mapped according to the techniques adopted in the studies. Last, we draw key findings and synthesis from the systematic review and identify challenges and opportunities in the area of sport analytics. The review follows a structured approach and is driven by practical considerations, which allows us to draw conclusions on the review.

3. METHODOLOGY

We collect publications, or material, by applying a structured keyword search in library databases, including Scopus, Springer, Wiley, EBSCO, Business Source Primer, ABI/INFO, Science Direct, Emerald, and the databases of major conferences such as INFORMS, ACM, AIS, and IEEE. We have identified the studies based on a delimitation criteria. In the next sub-section, we describe the approach to delimit publications and discuss the reasons for delimitation.

We delimit the literature search by establishing clear criteria. To gather more context and background, we also include conferences, books, and chapters in edited volumes. This reflects the fact that research encompasses specific terms, such as ‘sport AND analytics’ and ‘sports AND analytics.’ These two terms are used interchangeably, and thus, both are included in the search as keywords. The search criteria use the string operator AND to delimit the search and scan relevant research within the field of sport business management. The two phrases in the keyword string are used together as we want to delimit the search for journal articles, books, and proceedings to research analytics in the context of sport business management.

We conduct a search for publications for a 10-year window (Jan 2009–June 2019) to gain better understanding of the concept, context, and development. The starting point was set at 2009 to establish a time span of 10 years from 2009 to 2019. We use the terms ‘publications’ and ‘material’ interchangeably and define the terms as a unit of the literature being considered for review in this study. We collate publications or material by applying a structured keyword search in library databases, including Scopus, Springer, Wiley, EBSCO, Business Source Primer, ABI/INFO, Science Direct, Emerald, and the databases of major conferences. This creates a sub-selection of the literature and determines the material to include or exclude through another check of the criteria as above to remove any duplicates. We carefully review all the identified material (publications) leading to peer-reviewed journal papers, conference papers, books, and chapters in edited volumes used for the literature review. To reduce potential bias, all unique publications included in the sub selection are analysed in the review.

4. ANALYSIS

In some cases, some author(s) have written different papers on similar area published in different journals. There are five papers on online marketing, by different authors [5,24–27]. There are two papers on social media marketing which use variety of analytical methods [28,29]. We identified three papers on betting marketing that used analytical methods [3,4,30].

Betting marketing have also been studied in sport management through analytical methods. There are two studies on fan-base marketing [31,32]. We also discuss papers that use data driven techniques to assess player performance and bookmaker odds [1,33].

There are seven papers on match outcome by different set of authors [6,34–39]. One study is observed on sport leadership selection [40]. Three papers reflected on match strategy, team selection, match outcome and related areas of strategic relevance in sport [31,41,42].

We observe that the papers use variety of analytical methods/technique to address business questions in sport management. These analytical methods leverage primary and secondary data both and cover a wide spectrum depending upon the kind of business question being addressed. The methods range from data envelopment analysis to machine-learning methods. We present a tabulation of these methods, their application areas and year of publications in Table 1. Table 2 organises the literature in different areas based on classifications, such as the nature, methods, year of the study, to identify the breadth of readership.

Predictive analytics method is used by seven different studies [7,30,34,35,38–41]. Predictive analytics includes different types of econometric and statistical techniques that range from machine learning, classification, multivariate regression and other related techniques. Such techniques attempt to analyse historical data to make prediction, which could be about an event in the future or an outcome about the rank of a player. For example, Nalbantis et al. [31] and Toma [36] use econometric methods like multivariate regression and classification techniques to make predictions. Nalbantis et al. [31] apply statistical techniques to predict the Fans’ perception of competitive balance. Toma [36] analysed the data through classification techniques to predict missed shots at the free-throw shots of games in professional basketball. Technical efficiency is estimated using data envelopment analysis to address the problem of measuring sporting results as output in knockout competitions [37]. Mixed methods are applied by few studies as well [5,28]. Mixed methods involve the methods where research
Table 1  Classifications by area of application

<table>
<thead>
<tr>
<th>Paper (numbered as per reference list)</th>
<th>Journal/Conference</th>
<th>Application area</th>
</tr>
</thead>
<tbody>
<tr>
<td>5, 16</td>
<td>Journal of sport management</td>
<td>Social media marketing</td>
</tr>
<tr>
<td>1, 9</td>
<td>MIT sloan sport analytics conference</td>
<td>Data analytics in sport management</td>
</tr>
<tr>
<td>6</td>
<td>Twenty-second Americas conference on information systems, San Diego</td>
<td>Technology and data analytics in sport management</td>
</tr>
<tr>
<td>3</td>
<td>Journal of the royal statistical society</td>
<td>Football betting market</td>
</tr>
<tr>
<td>33</td>
<td>Journal of sport management</td>
<td>Perception study</td>
</tr>
<tr>
<td>4</td>
<td>Journal of the royal statistical society</td>
<td>Football betting market</td>
</tr>
<tr>
<td>35</td>
<td>Journal of sports economics</td>
<td>Performance of teams</td>
</tr>
<tr>
<td>41</td>
<td>Journal of sports economics</td>
<td>Payroll of football players</td>
</tr>
<tr>
<td>3, 25</td>
<td>International journal of sports marketing and sponsorship</td>
<td>Bidding for sport events/tournaments</td>
</tr>
<tr>
<td>18</td>
<td>Journal of sport management</td>
<td>Scoping review</td>
</tr>
<tr>
<td>36</td>
<td>Journal of advanced research in computer and communication engineering</td>
<td>Multi-target identification in sport</td>
</tr>
<tr>
<td>43</td>
<td>Journal of sport management</td>
<td>Contest success function</td>
</tr>
<tr>
<td>42</td>
<td>Journal of sports economics</td>
<td>Performance assessment</td>
</tr>
<tr>
<td>36</td>
<td>Journal of sports economics</td>
<td>Marketing and promotions</td>
</tr>
<tr>
<td>30</td>
<td>Malaysian journal of computer science</td>
<td>Competitive balance</td>
</tr>
<tr>
<td>37</td>
<td>Journal of sports economics</td>
<td>Performance evaluation</td>
</tr>
<tr>
<td>38</td>
<td>Journal of sports economics</td>
<td>Data analytics to predict match outcome</td>
</tr>
<tr>
<td>31</td>
<td>Journal of quantitative analytics in sport</td>
<td>Betting odds</td>
</tr>
<tr>
<td>34</td>
<td>Journal of sports economics</td>
<td>Scoping study</td>
</tr>
<tr>
<td>19</td>
<td>Journal of sport management</td>
<td>Visual analytics</td>
</tr>
<tr>
<td>16</td>
<td>Visual informatics</td>
<td>Consumer innovativeness on sports team</td>
</tr>
<tr>
<td>28</td>
<td>Journal of sport management</td>
<td></td>
</tr>
</tbody>
</table>

Table 2  Classification by analytical methods and area of study

<table>
<thead>
<tr>
<th>Analytical methods</th>
<th>Area of study</th>
<th>Reference (numbered from the reference list)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental studies</td>
<td>Team composition</td>
<td>14</td>
</tr>
<tr>
<td>Confirmatory factor analysis</td>
<td>Predicting athletic performance</td>
<td>13</td>
</tr>
<tr>
<td>Predictive analytics/ predictive modelling</td>
<td>Match outcome</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Classify draws</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Match outcome</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Match strategy</td>
<td>13, 38</td>
</tr>
<tr>
<td></td>
<td>Various areas of application in sports organizations</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Betting</td>
<td>5, 15</td>
</tr>
<tr>
<td></td>
<td>Teams</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Social media marketing</td>
<td>25, 28, 43</td>
</tr>
<tr>
<td></td>
<td>Bookmaker odds</td>
<td></td>
</tr>
<tr>
<td>Bayesian probability</td>
<td>Match outcome</td>
<td>30</td>
</tr>
<tr>
<td>Multivariate regression</td>
<td>Impact of star power in major league baseball</td>
<td>39</td>
</tr>
<tr>
<td>Design of experiments</td>
<td>Sport outcome</td>
<td>15</td>
</tr>
<tr>
<td>Machine learning</td>
<td>Various areas of application</td>
<td>11, 13, 30</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>Social media marketing</td>
<td>13, 29</td>
</tr>
<tr>
<td>Contest success function</td>
<td>Match strategy</td>
<td>42</td>
</tr>
</tbody>
</table>

is conducted by analysing and integrating quantitative and qualitative (e.g., interview, focus groups discussion) techniques. Baena [24] discuss and apply principal components analysis in one study. Cordes and Olfman [24] have applied machine learning to perform classifications. Bharathan et al. [1] have applied optimization model to assess player performance team selection. Machine learning has been discussed and applied by two different studies [25,29]. The latter study has also used classification methods [29]. Experimental studies have been used to study the effect of consumer innovativeness on sports team applications acceptance [27]. The research studies have applied stochastic process to estimate probabilities for outcome of soccer matches played between any two teams [4]. Swanson and Kent [40] have used confirmatory factor analysis. Conjoint analysis and contest success function have been studied and applied respectively [40,42]. Štrumbelj [33] has presented the application of stochastic processes and predictive models in betting markets. Table 2 presents broad classification of studies based on application of sport analytics.

The classification is based on application and technique adopted. There are applications in different areas of sport, including modelling, bid outcome, match outcome, player assessment, team composition, and others. Specific papers within each classification are reviewed. These are also cross tabulated, and the key findings are presented in the ‘Discussion’ section.

We used a wide level of taxonomy and identified the years of their publications to understand the level of attention that respective publications are gaining. We complement the taxonomy with a discussion of the extent research (in ‘Discussion’) to understand sub-areas of study and further details of the studies.

It is observable from the aforementioned studies that analytics has been applied increasingly in sport management. With sports becoming more competitive, researchers are turning to sport analytics for newer models to understand the relevance of data analytics to sports across different areas, including bidding, player performance, team performance, decision-making, entertainment, and attracting fans more effectively. Table 1 organises the literature in different areas based on classifications, such as the nature, methods, year of the study, to identify the breadth of readership and longitudinal analysis.
A summary of these studies suggests that there is a combination of data-capturing technology and adaption of newer data analytics models within the sport industry. We also find that data driven methods in sport are becoming increasingly relevant, and it has been used in several studies within sport management. We discuss these in detail in next section and present specific studies within each classification (context and mathematical tools/techniques). We obtain a useful taxonomical based review that helps researchers to identify specific areas of work and the method used. We also find that data analysis applied is on primary data or secondary data. We find that, recently, several studies have been using secondary data (from social media platforms and databases) to apply data analytics in different contexts of sport.

5. DISCUSSION

In this section, we present key findings and synthesis of the review. Specifically, we focus on the key discoveries that we could ascertain from the review and these are presented in a sequential order. The sequence, however, does not refer to the relative importance of the area of research. Also, the findings are synthesised from the systematic review and are carefully considered such that they cover major facets of research in this area. By no means, they are exhaustive in nature. We have classified the literature based on application of sport analytics and analytical methods that have been applied. We could find that various analytical tools and techniques have been applied in different circumstances and context. The studies have covered diverse ways of analyzing the data that was related to application of sport analytics and the data being used. Although, analytical methods leverage both primary and secondary data, we observe growing number of studies relying on secondary data gathered from such media. In the recent past, secondary data, particularly data generated from Internet and mobile commerce has significantly increased and presents opportunity to derive further insights that can be useful for the business. We could also ascertain that the studies covered different techniques and analytical methods like predictive analytics, machine learning, mixed methods, experimental studies, stochastic process, classification, support vector machine, data envelopment analysis, multivariate regression, predictive modelling, association rules, discriminant analysis, confirmatory factor analysis, conjoint analysis and contest success function.

We observe that different methods are applied to practical problems of business interest in sport. The business applications in different functional areas of management include strategy, operations, leadership, finance and marketing. Business issues deliberated upon are leadership selection, fan-base marketing, social media analytics, match outcome, bookmaker odds, team composition, online marketing, match strategy and others. Among these, few areas of increasing interest include social media marketing, prevention of sport injury and bidding for games. Data analytics also gets applied to evaluate sport injuries and derive strategies for prevention of injuries. In such applications, secondary data on sport injuries is analysed to predict injuries. This kind of research proposes methods to mitigate sport injuries and thereby protect the health of players. It is observable from the review that sports analytics has been applied increasingly in bidding for the events like Olympics and Commonwealth Games. These studies take into account perspectives of both host and non-host communities and their relative preferences.

Also, researchers are increasingly using machine-learning methods to process the opinion of people on social media [29]. Some studies have also investigated the feasibility of using classification algorithms on information collected on various social media platforms to predict match outcome. Additionally, it was found that the technique of Support Vector Machine technique is best suited for such kind of problems. Abeza et al. [28] have done exhaustive study on the social media scholarship and have identified research approaches, platforms, theories, and topic areas receiving the least/most attention from the social media.

The papers reviewed indicate that data analytics based on modern machine-learning techniques is gaining strong acceptance in sport marketing. Researchers have applied it in developing and applying analytical procedures to aid in strategic, tactical and operational issues in sport marketing like fan-base marketing, sponsorship, promotions and social media marketing. Associations, clubs, managers, and coaches increasingly apply data analytics in sport and this is an increasing trend considering the frequency of such studies over the recent years. Several studies present that better conceptualisation of sport management and data analytics will result in better applications. It is also found that, while much of the data used in research to date come from secondary datasets (which are proprietary), few papers are based on publicly available data. We find that it is, by far, possible to use information in the public domain for application of sport analytics.

The studies also indicate that it is possible to effectively utilise immense amount of data being collected, both on the field and off, through continuously improving data collection technologies. Various techniques are being effectively applied that enable better understanding and analysis of the available data. The application of these techniques is being discussed in the literature and improved continuously. Several such applications, and the motivation behind each, are presented in different studies.

This study, through a systematic review, identifies how sport analytics has been used in sport research and helps assess the present state of research on sport analytics. While the author attempts to present the review, but it is not possible to present a review that is exhaustive over all the years. Also, the review does not attempt to assess research quality. However, it indicates current position of research in the field and sub-fields. For example, the review shows that studies have sought to (1) assess game outcomes, (2) study the behaviour of betting markets, (3) examine relationships between payroll and player performance, (4) analyse the challenges and opportunities facing a specific sport, (5) determine elements essential to victory, (6) discern consumers’ emotions while watching sports, (7) establish elements that lead to player choking, and (8) observe the link between sport-related knowledge and skill and the perception of leaders.

The papers reviewed indicate that companies, clubs, managers, and coaches wish to use technology and data analytics in sport. In summary, we find that both the literature and practitioners continue to emphasise the use of data analytics in the future of sports. Several studies present that better conceptualisation of sport management and data analytics will result in better applications. Further, we find that, while sport analytics has attracted significant attention from scholars, actual sport analytics research is in a nascent stage, and there is an urgent need for more research in this area. In addition, while much of the data used in research to date comes from
secondary datasets, which are proprietary, we also find that few papers are based on publicly available data. We find that it is, by far, possible to use information in the public domain, thereby improving the efficiency and effectiveness of data analytics application in sport management.

6. CONCLUSION

We find that different kinds of analytical methods and techniques have been applied in diverse contexts of sport. We also observe that frequency of research in sport analytics is consistently increasing as we could gather from the increasing frequency of such studies in the recent years.

Several analytical methods and techniques, through which data analytics has evolved in the field of sport business management, have been studied and explored in the literature. The paper contributes to theory in two specific ways. This study has used systematic analysis of the literature in journals, conferences and other fora in the field of sport analytics. The analysis helps identify impactful studies in the field for scholarship and theoretical understanding. Second, the study classifies and conducts a taxonomy on variety of analytical methods used in different sport contexts. It helps in development of a structure to classify the studies and also understand the theoretical base.

Certain practical implications are drawn from the review of studies. The review shows that several authors have demonstrated the application of data analytics in a variety of practical contexts like fan-base marketing, consumer sentiments, player bidding, sport injury, player performance, promotions, bidding for games and others. It is also observed that variety of analytical methods have been used. These methods include multivariate regression, descriptive analysis, optimization and even machine-learning methods (logistic regression, support vector machines, random forest etc.). Increasingly, these methods are being adopted in practice with the application of open source technologies like Python, R, Gephi and others.

There are certain limitations to the study and future research directions. The study uses a systematic review to determine the current position of studies published in conferences proceedings, journals, books and chapters in edited volumes. While a systematic literature review is useful to develop a scholarship base, it is still limited by the sample of literature as selected by the researcher. A systematic review cannot always be exhaustive as it is not possible to enumerate all the research historically. Further, the author does not attempt to perform a longitudinal analysis in this study as the author’s focus was on taxonomical review and also uncovered range of analytical methods and business contexts. As a future research, longitudinal analysis of review can be attempted to discover the trend in research.

As future research direction, it would also be useful to use semantic analytics and bibliometrics to perform citation and co-citation analysis on sport analytics. Such an analysis, based on citations, co-citations and co-authorship, would uncover intellectual structure of this field while also discovering emerging trends, challenges and prospects.

REFERENCES


