

The transformative impact of big data applications in sport marketing: current and future directions

Big data
in sport
management

Yoseph Mamo

Tennessee State University, Nashville, Tennessee, USA

Yiran Su

University of Georgia, Athens, Georgia, USA and

Department of Kinesiology, The Institute for Artificial Intelligence,

University of Georgia, Athens, Georgia, USA, and

Damon P.S. Andrew

College of Education, Florida State University, Tallahassee, Florida, USA

Received 25 March 2021

Revised 18 July 2021

22 September 2021

Accepted 5 October 2021

Abstract

Purpose – As big data (BD) has increasingly become an important tool for managers and researchers to transform sport management practices, the purpose of this research is to highlight diverse data sources and modern analytical techniques that will leverage BD as a means to advance scholarship in sport management.

Design/methodology/approach – A comprehensive review of existing BD literature in sport management outlines new perspectives on BD research method and the application of BD in sport management.

Findings – First, through a thorough review of the literature, a domain-specific conceptualization that incorporates the field's mission and priorities was developed. Second, potential data sources and different types of analytical opportunities was identified, highlighting strategies for developing methodological approaches that leads to novel research questions. BD analytics can allow for more flexibility in improving methodological capability to analyze data and, thus, provide more granular and predictive insights. Finally, this paper concludes with a discussion of BD's impact on three domains of sport management, whereby the organizations yield data-driven decisions.

Originality/value – BD has the potential to transform the sport management operations and bridges the research-practice gap. BD research in sport management is instrumental for accumulating new knowledge and/or testing existing theories, either in a deductive fashion or by taking an inductive approach, as the field embarks to advance scholarship.

Keywords Big data, Big data analytics, Machine learning, Data-rich environments, Analytics, Social media

Paper type Research paper

Introduction

The term “big data (BD)” is used to describe datasets and analytical techniques in applications ranging from social media to the Internet of Things (IoT) that require advanced and unique storage, management and analysis techniques (Boyd and Crawford, 2012). It should be noted that “big data is just another way of saying analytics” (McAfee and Brynjolfsson, 2012, p. 4) with the addition of key characteristics that include volume, velocity and variety (the 3 Vs). The use of BD has gained popularity in social science (George *et al.*, 2014; Humphreys and Wang, 2018; Lazer and Radford, 2017; Tonidandel *et al.*, 2018) because data have become more readily available and the tools for analysis support improving science and practices. Correspondingly, the current interest in BD research among various domains of sport research has witnessed an increasing reliance on data-driven strategies with the fusion of BD and data analytics knowledge (Burton, 2019; DeSchriver *et al.*, 2021; Gong *et al.*, 2021; Lopez, 2020; Watanabe *et al.*, 2021).



Yet despite that many sport leagues and clubs have utilized BD and modern analytical techniques in marketing, there has been comparatively little sport marketing dialogue to extend theoretical knowledge and practices. For example, the advent of consumer-generated content and social media has revolutionized marketing research (Nam *et al.*, 2017). The high volume, variety and velocity of data have been demonstrated to be useful to assess consumers' opinions and attitudes, consumer segmentations, preferences for products and services, and promotional campaigns. As this article will demonstrate, BD in sport marketing is still in its infancy, requiring future research to combine perspectives and insights from different fields of research traditions. Recognizing a need for research, this paper is intended to illustrate the most promising opportunities that can be derived from diverse data sources and contemporary analytical techniques that will leverage BD as a means to advance scholarship in sport marketing.

Drawing on George *et al.* (2016, p. 1), the adaptation of BD in academic research is about "collecting and managing large varied data." As new technologies and platforms become ubiquitous, more convenient means of accessing BD arise, particularly through social media, mobile communications [Digital Social Media and Mobile (DSMM)] channels, sensors, cameras, transactions and neuro/psychophysiological tools (e.g. functional magnetic resonance imaging, electroencephalography). Johnson *et al.* (2019) argued that BD research depends on researchers' access to data. Although there are many new data sources to acquire BD from data exhaust, community data and self-quantification data sources (George *et al.*, 2014), research in sport marketing can significantly benefit from three specific developments in data availability: community, private and public (see Figure 2).

The new data sources further pose challenges to the use of traditional analytics approaches as data from the community and private data usually come as unstructured and do not fit into a relational database. These data sources are mainly composed of text and images that come in the form of unstructured data; thus, researchers interested in these data streams will need to familiarize themselves with modern analytical techniques. Some of the commonly used general analytical approaches include text analytics (e.g. computational linguistics, machine learning, sentiment analysis), image analytics, audio analytics (e.g. speech analytics) and video analytics (e.g. server-based analytics and edge-based analytics). Although these tools are advanced in computer sciences and modern business analytics, there are plenty of studies in social sciences that employ advanced predictive techniques in areas related to organizational science, management, psychology and sociology.

The rest of the paper is in four parts. First, we briefly review the existing BD literature in sport research to assess, critique and synthesize the knowledge base to provide a potential domain-specific conceptualization that incorporates the field's mission and priorities. Second, we discuss the potential data sources and features that are appropriate for sport marketing disciplines. Third, we discuss different types of analytical opportunities that may be associated with the various dimensions of BD and highlight potential strategies for developing methodological approaches via careful analysis and substantive research questions. Via these mechanisms, BD analytics can allow for more flexibility in improving methodological capability to analyze data and, thus, become more precise than ever before. Finally, we conclude with a discussion of three domains of sport marketing which we predict that BD impacts.

Characterizing big data

Most data scientists and scholars define BD according to three characteristics, the 3Vs of volume, variety and velocity (Laney, 2001). Large volumes of digital data are generated continuously due to various technological advancements, such as cloud computing, multiple data centers, the proliferation of smart devices and the ever-growing access to the Internet

across the globe. International Business Machines (IBM) has estimated that more than 2.5 quintillion bytes of data are generated every day (Marr, 2021), which is equivalent to 50 billion filing cabinets filled with text. McAfee and Brynjolfsson (2012) asserted that the amount of data was doubling almost every three years. Variety refers to a wide variety of data units (e.g. video, images, blogs, audio and face recognition). Such data types are generated in structured, semi-structured and unstructured forms. Velocity is associated with transfer speed, scalability and timing.

Though the Vs might be a useful approach for explaining the characteristics of BD, scholars still seek a clear definition of BD since it has changed over time, and its complexity has increased due to the advance of computational techniques, such as machine learning and artificial intelligence (Sena *et al.*, 2019). For instance, it is not clear what “big” means or the minimum number of instances in a dataset that can define data as BD. While BD typically require tetra-bytes of disk space (All too much, 2021), customer transaction data (e.g. purchase history) in sport organizations may not seem to have a large number of measurements per individual compared to other industries, such as retail, healthcare, supply chain or those in the service sectors. On the other hand, an uncompressed 1080p high-definition video of a single football game can easily exceed hundreds of gigabytes (imagine the data generated in a single football season) and may be considered a “big” dataset. Furthermore, the use of the variety and velocity of BD in sport-related business is unfamiliar territory apart from performance-related studies (e.g. Franks *et al.*, 2016; Lopez, 2020). Thus, rather than thinking about the volume, variety or velocity of data, we suggest that the classification of BD depends on the context in which it is used, and it is discipline-dependent.

Big data research in sport management

To determine the occurrence and practice of BD, the focus was set on finding the most relevant journal articles on BD in sport scholarship. We reviewed articles that were published in the *Journal of Sport Management (JSM)*, *Sport Management Review (SMR)*, *European Sport Management Quarterly (ESMQ)*, *Sport Marketing Quarterly (SMQ)*, *Communication and Sport (CS)*, *International Journal of Sport Communication (IJSC)*, *International Journal of Sport Marketing and Sponsorship (IJSMS)*, *Journal of Sports Economics (JSE)*, *International Journal of Sport Finance (IJSF)* and *Journal of Global Sport Management (JGSM)* over the last nine years (August 2012 to May 2021). These journals are often considered the principal specialist journals in the targeted context of sport. We also included the *Journal of Quantitative Analysis in Sport* and *Journal of Sports Analytics* given these journals have been focusing on novel methods to understand data to address topics, such as game outcomes models, within-game strategy and analysis of sporting technologies.

To identify the relevant articles, we used the following search terms in multiple combinations: *Big Data*, *Big Data analytics*, *analytics*, *large dataset*, *large-scale data*, *large dataset*, *tall data*, *multiple data*, *volume of data*, *variety of data*, *velocity of data*, *web analytics*, *social media analytics*, *multiple data formats*, *high rate*, *data flow* and *non-homogenous structure*. These keywords were selected based on three factors: (1) their proximal meaning to the idea of BD given that all terms have been used by past studies to refer to some sort of BD, (2) previous scholars’ work in other fields (Sivarajah *et al.*, 2017) and (3) insights from experts. Next, for the sake of simplicity, we based our classification on the 3 Vs of volume, velocity and variety. Some studies encompassed more than one V.

Volume, velocity and variety

Given that volume generally refers to the storage space for enormous quantities of data, we outline three important factors to determine large datasets in sport management: (1) the aggregation of a large number of variables, (2) a large set of observations for each construct

and (3) volumes of data extracted from non-survey data, in which analytical techniques are specified after data collection. In other words, our approach to categorizing studies under volume is mainly based on the sample size, the number of variables, the research settings in which the data have been collected and the analytical techniques to manage the data. These factors are by no means exhaustive criteria of volume. Rather, they are used for illustrative purposes and were selected due to our familiarity with the field. Regarding velocity and variety, most of the discussed criteria for selecting large datasets for volume data also apply to velocity and variety characteristics of BD, with the notable additions of the frequency of data generation for velocity and heterogeneity of the data for variety.

The initial search yielded 136 articles that satisfied the specified criteria. As presented in Figure 1, the largest number of publications was recorded for the year 2020 ($n = 33$), followed by 2019 ($n = 25$) and ($n = 23$) in 2021. However, it should be noted that the search was

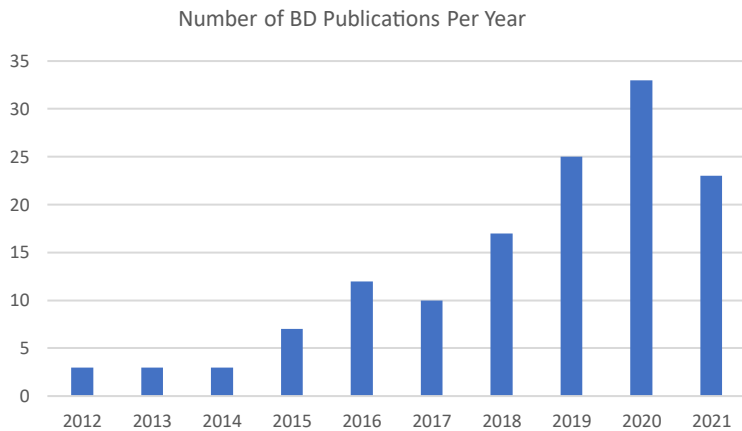


Figure 1. Big data publications

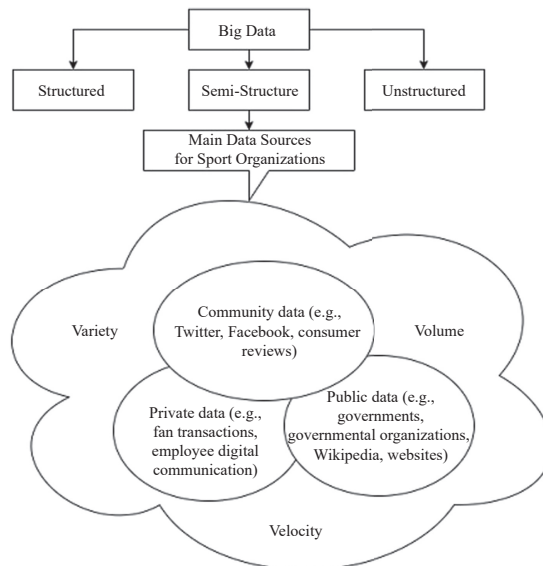


Figure 2. Big data source in sport management

conducted in May 2021, and it is reasonable to assume the number of publications will grow in the next seven months. The rapid increase in publications highlights BD's awareness and importance in the sport context. Most articles were published in *Journal of Quantitative Analysis in Sport (JQAS)* (21), *CS* (19), *IJSF* (16), *JSM* (16) and *IJSC* (12). It is also notable that *JSM* has a special issue devoted to BD in 2021. The remaining journals have fewer than ten articles each.

Of the 136 articles, 108 articles emphasized the business side of sport (e.g. economics, communication, finance, marketing and management). The remaining articles were focused on measuring and analyzing game strategies, player ability, completion probability and team performance. In regard to sport marketing components, we drew on [Peetz and Reams' \(2011\)](#) categorizations of sport marketing topics (e.g. analysis of spectators, marketing research, marketing segmentation, industry segmentation, price and product). Next, the research team carefully selected some articles to illustrate the current application of BD in the sport marketing areas.

As reported in [Tables 1 and 2](#), respectively, the presence of BD research is evident in sport marketing. Most sport marketing research meeting the specified criteria was closely

Authors and publication year	Purpose of study	Size and source of datasets	Study findings
Matti (2021)	Assessed the role of emotional cues developed from unexpected game outcomes on perceptions of local businesses	Data were collected from two different sources. First, information nearly 30,000 Yelp-rated establishments in the greater Phoenix, Arizona area. Second, 394 Suns games across five seasons from the sports betting website covers.com	The result suggested that the impact of unexpected losses is concentrated in home games, with no effect for away games
Watanabe et al. (2019)	Examined the effect of college football players' protests against racial injustice and police brutality on consumer interest in attending games	Data were extracted from regular-season college football game attendance for the 2014–2016 seasons	The results revealed that the voting demographic of the local sport market was affected by the protests, and in some cases, attendance declined by up to 10%
Watanabe et al. (2016)	Investigate factors that determine consumers' interest in using the team's social media account	Data were collected from Major League Baseball Twitter accounts for the 12 months by the NodeXL software package	The findings revealed that varying results of the short- and long-term consumer interest in teams on Twitter
Tainsky et al. (2015)	Investigated potential discrimination among MLB umpires in terms of ball-strike calls as associated with race, specifically, the similarity of the umpire's and pitcher's races	Data were drawn from multiple seasons and millions of pitches	They found no significant relationship between ethnic discrimination in the calling of balls and strikes by MLB umpires

Note(s): Select articles are published in 2012–2021 in the *Journal of Sport Management (JSM)*, *Sport Management Review (SMR)*, *European Sport Management Quarterly (ESMQ)*, *Sport Marketing Quarterly (SMQ)*, *Communication and Sport (CS)*, *International Journal of Sport Communication (IJSC)*, *International Journal of Sport Marketing and Sponsorship (IJSMS)*, *Journal of Sports Economics (JSE)*, *International Journal of Sport Finance (IJSF)*, *Journal of Global Sport Management (JGSM)*, *Journal of Quantitative Analysis in Sport (JQAS)* and the *Journal of Sports Analytics (JSA)*. The table is not meant to be comprehensive

Table 1.
Volume studies in
sport marketing
components from
08/2012 to 06/2021

Authors and publication year	Purpose of study	Size and source of datasets	Study findings
<i>Velocity</i>			
Chang (2019)	To measure spectators' emotional response during the 2016 Super Bowl game	A total of 328,000 real-time tweets during the 2016 Super-Bowl game	The findings showed that fans expressed positive emotion while the team scored, whereas they showed negative emotion when the opponent's team scored
Burton (2019)	Examined consumer attitudes towards ambush marketing and official event sponsorship	Twitter data were accessed between a week before the commencement of the 2018 FIFA World Cup and terminating one week following the tournament's final	The study's findings showed that consumer attitudes towards ambush marketing appear to be more positive than previously assumed
Yan <i>et al.</i> (2019)	Investigated patterns of attention and power from the European Football Championship League final match	By retrieving data from Twitter networks (#UCL), these researchers collected 19,869 pre-match posts, 3,276 halftime posts, and 5,691 post-match posts	Large sport teams and superstar athletes had a stable and privileged status in organizational organizing the networks
O'Hallarn <i>et al.</i> (2018)	Examined secondary market price through sport team's Twitter hashtags	A Twitter aggregator, Topsy, was used to track eight "official" NFL's team Twitter handle for a month	The results revealed that increased Twitter hashtag use is a significant positive antecedent of ticket prices on the secondary market.
Yan <i>et al.</i> (2018)	How and why users engage in sport digital communication?	A social-media analytic software program was used to collect data from 123 FBS teams. A total data set composed of 2,822,850 pieces of content over 21 weeks	The findings showed that the presence and timing of college football games along with game outcome, are significant factors for user's online content generation
<i>Variety</i>			
DeSchriver <i>et al.</i> (2021)	Examined derived demand in the lodging industry by analyzing hotel revenue changes that are attributable to market- and game-specific factors in cities that host major college football	Data were collected from the hotel management analytics firm Smith Travel Research (STR). The original data set obtained from STR featured over 4.4 million daily performance metrics for the 1,153 hotels in the SEC markets over the 15-year time period (2003–2017)	Results indicated that team performance, opponent, school events, and hotel class positively affected demand
Delia (2019)	Examined Syracuse University Men's Basketball fan perceptions of Coach Jim Boheim's 1,000th win	Data were collected from news articles from the school website, its Facebook page, and the SB Nation Syracuse Athletics blog	Fans may form their own truth and maintain a positive social identity while enduring an identity threat
Table 2. Velocity and variety studies in sport marketing components from 08/2012 to 06/2021	Note(s): Select articles are published in 2012–2021 in the <i>Journal of Sport Management (JSM)</i> , <i>Sport Management Review (SMR)</i> , <i>European Sport Management Quarterly (ESMQ)</i> , <i>Sport Marketing Quarterly (SMQ)</i> , <i>Communication and Sport (CS)</i> , <i>International Journal of Sport Communication (IJSC)</i> , <i>International Journal of Sport Marketing and Sponsorship (IJSMS)</i> , <i>Journal of Sports Economics (JSE)</i> , <i>International Journal of Sport Finance (IJSF)</i> , <i>Journal of Global Sport Management (JGSM)</i> , <i>Journal of Quantitative Analysis in Sport (JQAS)</i> and the <i>Journal of Sports Analytics (JSA)</i> . The table is not meant to be comprehensive		

associated with social media data and panel data concerning volume. A large number of variables were collected from open sources that underline the availability of volume data as a structured form. That is, most of the data are ready for use, or researchers may write a scraping code to access the data. Few studies were employed in the sport marketing space (e.g. Burton, 2019; Chang, 2019).

The majority of reviewed articles draw BD from panel data and individual Tweets, referring to “what” and “who” questions with a structured data type. Given that almost all of the work on BD has focused on limited characteristics of BD (i.e. volume), community data sources, relatively narrow sets of methodological approaches (e.g. descriptive, correlation and regression) and numeric data type, one can argue that the concept in the field is still in its nascent stages. It should be noted that the selected articles did not comprehensively fulfill all the specified criteria for each characteristic of BD (i.e. 3Vs). Nevertheless, the goal is to provide a broad overview of BD compared to more traditional forms of small dataset across the domains of sport marketing.

Defining big data in sport

[Tonidandel and his colleagues \(2018\)](#) suggest that “a more productive [way] to define data as [big data] is when the characteristics of the data require you to change your mindset” (p. 526). In light of this perspective, it is important to examine in what ways the landscape of BD can influence the mindset of sport marketing researchers in the pursuit of scanning, identifying, acquiring and processing data to address complex issues in order to contribute to the knowledge base.

Sport is a “unique cultural institution that operates in a commercial environment” ([Smith and Stewart, 2010](#), p. 2), which, in turn, is visible to large audiences. Evolving from its traditional emphasis on physical education, modern sport management research also places a strong emphasis on sport-related business activities ([Pitts and Stotlar, 2013](#)). To achieve this, almost all professional sport teams, international sport governing bodies, collegiate athletics and non-profit sport organizations have embraced digital technology to interact with customers and facilitate engagement. This dynamic interaction fuels circulation of sport-related content spread through digital channels (e.g. social media posts, user discussion forums and blogs) across a variety of levels, which, in turn, allows sport organizations to be situated in a data-rich environment. [Wedel and Kannan \(2016\)](#) describe a data-rich environment as structured and/or unstructured data that generate from consumers, firms’ digital communication channels and content produced from customer to firm.

Given that sport marketing is an applied field that could be expanded and impacted through its data-rich environment, researchers need to glean different data types to offer both substantive and theoretical developments. In general, we propose that many benefits of BD in sport will be realized with a thoughtful understanding of the new data sources, modern analytical techniques and their increasing use in research practice. Therefore, there is a need to start a conversation pertaining to changing our mindset as to how we should “improve our game” ([Funk et al., 2016](#)). In doing so, exploring and adopting new data streams and applying new computational techniques to address those gaps are appropriate and require a change in mindset.

As stated above, espousing a clear meaning of BD has a strong association with the field’s scholarly strengths and priorities. A concise working definition of BD in sport is the process of identifying and analyzing information from data-rich environments, which may emerge as multiple data types with variable speeds to make decisions regarding sport-related business, athletic activities and products. Our proposed definition of BD in sport marketing is intended to capture the following concepts. First, the central idea concerning BD in sport is its association with data-rich environments, both structured and unstructured. Second, BD is a phenomenon that requires a change in mindset; that is, data may be added constantly and heterogeneously. Finally, BD is a strategic tool that contributes to a sport organization’s internal and external decision-making purposes.

Big data sources for sport marketing research

Sport benefits from increased community data availability through microblogging sites, social network trends and content communities. For example, among the world's ten most searched stories on the Internet in 2019, three of them are sport-related (Editorial, 2021). Additionally, sport leagues are now turning to more digital programming to connect and engage with fans (e.g. 81% of sport viewers interact on social media; [The Huge Statistics of Sports on Social Media | Rosy Strategies, 2021](#)). Teams such as Real Madrid and Barcelona each have more than 210 million Facebook and Twitter followers. [Brock and Khan \(2017\)](#) asserted that social media is considered a vast source of BD.

[Yan et al. \(2019\)](#) investigated the patterns of attention and power in the European Football Championship League final match. NodeXL software was used to collect Twitter data, and the software packages allow researchers to gather, analyze and visualize social media data. By using the hashtag #UCL, these researchers retrieved 19,869 pre-match posts, 3,276 halftime posts and 5,691 post-match posts from Twitter. As social media have become a significant source of volume and velocity data, this study reveals that large sport clubs have the ability to dominate the social network due to their stable privilege position while certain actors are influential in a specific time. Relatedly, [Chang \(2019\)](#) collected 328,000 real-time tweets to measure spectators' emotional responses during the 2016 Super Bowl game. This investigation extended our understanding of sport spectators' emotional fluctuations during viewership on sport consumer behavior. In this sense, analyzing fans' unfiltered real-life expressions (e.g. dynamic, urgent and contextual) during a live sport game can unlock insights to identify potential trends and publicity crises.

More recently, [Matti \(2021\)](#) scraped almost one million Yelp reviews to examine the impact of emotional cues emanating from unexpected outcomes on perceptions of local business. Yelp is one of the most popular Internet ratings and review sites for local businesses, and it is highly applicable in tourism and services research. This study collected nearly one million Yelp reviewers and descriptive statistics for 394 Phoenix Suns games from the sport betting website [covers.com](#). Such insights inform our understanding of loss aversion through applying volume and variety data. Similarly, [Mao \(2020\)](#) employed Yelp reviews to understand the retail quality of sporting goods stores. [Mao \(2020\)](#) collected 27,793 online reviews of 1,481 stores across seven states: Arizona, Illinois, North Carolina, Nevada, Ohio, Pennsylvania and Wisconsin. This investigation used BD to cross-validate on retail literature and indicated that knowledgeable staff beyond a salesperson can provide other value-added services.

The second development involves using private data collected within the sport organization through recording consumer's transactions, including the use of clickstream to categorize visits by location, purchase, search and use of mobile application. The popularity and emergence of technology have contributed to digital life in sport contexts, that is, access to customers' and firms' digital interactions ([Lazer and Radford, 2017](#)). As sport leagues and customers are connected through digital platforms, teams and sponsors will have unique benefits to gather and share information. For example, [DeSchriver et al. \(2021\)](#) obtained daily performance data on hotel occupancy rates and average daily rates to examine the impact of Southeastern Conference college football games on local hotel demand from 2003 to 2017. Data were collected from the hotel management analytics firm Smith Travel Research (STR). This data source is an example of proprietary data in the tourism and hospitality industry. This study advanced demand literature in sport and hospitality using volume and variety data.

The third area of data availability comes from the government, research funding agencies, professional societies, universities, individual researchers and many other public data repositories. [Uhlir and Schröder \(2007\)](#) asserted that "many publicly funded data can be of great value for reuse by a broad range of public and private researchers, other types of

socioeconomic applications and the general public (p. 38). [DeSchrive *et al.* \(2021\)](#) collected population data from the US Census Bureau. Furthermore, the National Collegiate Athletic Association, sports-reference.com, and official university websites have published large datasets. [Jensen *et al.* \(2020\)](#) gathered 169,479 observations to explore donor behavior in the intercollegiate athletic industry. The results highlight that the probability of donor contraction increases with decreasing economic growth. In addition, the public data repository (<https://libguides.geneseo.edu/>) publishes historical and other data for professional and Olympic sports as well as all co-educational postsecondary institutions that receive Title IV funding (i.e. those that participate in federal student aid programs) and that have an intercollegiate athletics program. In general, sport marketing researchers can benefit from utilizing public data produced by various entities and enhance the quality and productivity of research.

Given that there has been a tremendous effort to demonstrate the uniqueness of sport (e.g. stakeholders and business operations and the value co-creating consumption process, and its uncertain nature), new data sources may be useful for answering many research questions about consumer, fan and market behavior, as well as society at large. Despite progress in collecting data from community, private and public data sources, many opportunities for continued contributions remain. For example, the emergence of sport blog websites and fan-hosted podcasts can help sport marketers to uncover fans' opinions and enable new business opportunities (e.g. [Kwak *et al.*, 2010](#)). In addition, sport marketing scholars may extend the analysis from text data to image, audio and video data from various data sources to enhance our understanding of marketing effectiveness (e.g. [Su *et al.*, 2020](#); [Rivers and Ross, 2021](#)) (see [Figure 2](#)).

Big data and modern analytical techniques in sport marketing

As evident in sport scholarship, much work completed within sport marketing relies on linear relationships (e.g. multiple regression and structural equation models [SEMs]), small sample size and primarily focuses on cross-sectional studies and attitudinal surveys to collect information. Often, community, public and private data come as unstructured and do not fit into a relational database, which, in turn, poses challenges to the use of traditional approaches. These data sources are mainly composed of text, images, audio, metadata and video that come in the form of unstructured data; thus, researchers interested in these data streams may need to familiarize themselves with modern analytical techniques. Numerous techniques and tools have recently emerged, such as natural language processing techniques, machine learning, random decision forests, neural networks, deep learning and text mining, along with many other applications that help better explain complex phenomena and provide deeper insights. For example, text mining is a broader term used to describe the process of deriving meaning from textual materials to capture key concepts, trends and hidden relationships.

Existing studies in sport management have applied text mining techniques to analyze tweets (i.e. unstructured data) in sponsorship and ambush marketing ([Burton, 2019](#)), spectators' emotional tweets during the Superbowl ([Chang, 2019](#)) and tanking ([Gong *et al.*, 2021](#)). [Gong *et al.* \(2021\)](#) investigated the impact of tanking on consumer interest in attending sport games. Unlike previous sentiment analysis practices within sport management that used a preprogrammed approach (e.g. a lexicon-based approach) or manual coding of manageable data, [Gong *et al.* \(2021\)](#) took a further step and trained a text classifier by employing a linear support vector machine (LSVM) algorithm. While LSVM requires more effort in a human-labeled document (i.e. training data), the experimental result showed that the technique produces higher precision in predicting sentiments on text documents. Through employing LSVM, [Gong *et al.* \(2021\)](#) trained 5,000 tweets out of 166,875 and reported

a 70% prediction accuracy on predicting the orientation of sentiments. Subsequently, they computed regression analysis to predict game attendance as a function of sentiment analysis and other control variables. Relatedly, [Mao \(2020\)](#) utilized a patented probabilistic topic modeling algorithm from Leximancer 4.0 to identify critical quality dimensions associated with sporting goods stores.

Other studies utilized modern analytical techniques in structured data. [Koenigstorfer and Wemmar \(2019\)](#) employed machine learning techniques and a random forest tree to predict factors that contribute to a nonprofit sport club's success in recruiting and retaining members. They found that manager-perceived service quality and trust are the most important predictors of recruiting and retaining members. [Du et al. \(2021\)](#) compared two supervised machine learning algorithms (i.e. random forest and support vector machines [SVMs]) with logistic regression to provide actionable heuristics to optimize the managerial capacity of churn management programs for season ticket holders. Findings suggest that the random forest technique was 40% more accurate than conventional logistic regression. Similarly, the SVM approach was 19% more accurate than the logistic model.

Despite increased attention to modern analytical techniques in sport marketing literature, significant gaps remain that could support the analysis of both unstructured and structured data. The following sections will discuss new perspectives on descriptive techniques, predictive techniques and construct measurement that could challenge the status quo and transform current sport management research.

New perspectives on descriptive techniques

Many areas of sport marketing and practice are inherently interested in uncovering patterns and correlations to better understand sport organizations and individuals without imposing a preconceived structure. This is sometimes referred to as exploratory analysis. Although exploratory analysis is the common recommendation from introductory statistics and research methods courses to summarize, plot and screen data, BD analytics have taken these techniques to a new height of importance and relevance. For example, descriptive analytics provide a static view of the past by applying algorithms to perform clustering and categorization to summarize the main characteristics of the data. Descriptive analytics employs supervised, semi-supervised and unsupervised learning models ([Rehman et al., 2016](#)). Supervised machine learning models infer a function from given labeled input data assigned by the researcher, whereas unsupervised learning allows the computer to find its own structure without human guidance.

Like exploratory analysis, descriptive analytics may not precisely identify pattern distributions and variability of data and relationships, but they can shed important light on agreeable starting points and boundary conditions, thereby generating novel research questions. Core functions of descriptive analytics, such as text mining, clustering models, ratio analysis and data visualization, have been used to enhance operations and facilitate business decision-making ([Balducci and Marinova, 2018](#)). [Oswald et al. \(2020, p. 512\)](#) referred to data visualization as "the strongest tool to explore and understand data and models interactively, dynamically, transparently and intelligently, above and beyond what overall statistical summaries might provide". Data visualization techniques process unstructured data into a graphic format (e.g. semantic network, word mapping, network diagram and social networks) that adopts incremental benefits for managerial decision-making processes. For example, [Chang \(2019\)](#) illustrated fans' emotional changes exhibited in tweets during the Super Bowl 50 game based on location and time interval. Given that sport organizations operate in a digital-rich environment, significant insights may be extracted by visualizing social media posts. Exploratory analysis, in many contexts, is not the ultimate goal of the research, but is instead used to determine predictive models that test relationships between or among variables.

New perspectives on predictive techniques

Predictive techniques used in the sport marketing literature are largely applied to traditional statistical methods that often incorporate traditional assumptions of linearity and are designed to handle small sample sizes and structure data. Such approaches then focus on a limited number of variables involved in the analysis to maintain a good model fit or theory congruence explanation. Whereas traditional statistical techniques are common practice in theory testing and refinement within existing research paradigms, many of these techniques do not accommodate nonlinearity, overfitting, balancing flexibility and robustness, or scale to longitudinal/dynamic measurement (Putka *et al.*, 2018; Yarkoni and Westfall, 2017).

Today's era of algorithm culture affords new opportunities to enhance detection and superior model granularity. For example, one of the primary benefits of modern analytical tools over traditional methods is "their ability to create more effective prediction models in the face of small sample size (n) to a number of predictor (p) ratios (i.e. n/p)" (Putka *et al.*, 2018, p. 691). For instance, suppose one seeks to understand which factors of external forces influence individuals' attitudinal formation and behavioral intentions using a sport product and service. There exists a wide range of potential predictors, such as socio-cultural, personality, perception constraints and motivation variables, one could examine. Previous research in sport marketing predominantly adopted a reduced-form approach, selecting a limited number of predictors based on one or part of one theoretical framework and conceptual model to balance complexity-parsimony. Such an approach has not only limited the number of variables to be computed in a model, but it may also engender a model that underutilizes potential variables; therefore, theories have often remained relatively simple (Putka *et al.*, 2018).

Putka *et al.*'s (2018) compared modern data analytics techniques with traditional methods while measuring personality and motivational constructs to predict physical performance. The findings showed that BD methods produce predictions that result in a higher level of precision than traditional ordinary least squares (OLS) regression. Du *et al.* (2021) compared two supervised machine learning algorithms (i.e. random forest and SVMs) with logistic regression to provide actionable heuristics to optimize the managerial capacity of churn management programs for season ticket holders. Findings suggest that the random forest technique was 40% more accurate than the conventional logistic regression. Similarly, the SVM approach was 19% more accurate than the logistic model. Perhaps the most appealing benefit of modern predictive techniques within sport marketing is the enhancement of accurate predictions and classifications.

New perspectives on measuring constructs

Sport marketing researchers routinely interested in constructs that change over time (e.g. satisfaction, etc.), BD measurement techniques may play prominent roles to accurately measure those constructs. For example, by evaluating Facebook "likes", Youyou *et al.* (2015) found that the computer-derived personality judgments had higher criterion validities than self-reported personality. Similarly, Park *et al.* (2015) employed language processing to assess personality from 70,000 Facebook users and found a greater degree of incremental validity than traditional techniques. Such research has laid the foundation to go beyond gathering cross-sectional structure data and leverage more multifaceted measures through social media to assess employees' behavior changes and psychological connections.

Within sport marketing, constructs such as team identification, satisfaction, fan engagement and advertising goals have been explored through survey and scale measurements (e.g. Funk and James, 2001). Although numerical rating scales are very popular, future studies might consider unstructured data to complement and extend traditional rating scales. Understanding such dynamic constructs is important for targeting

and tracking changes at the individual, firm and cultural levels (Berger *et al.*, 2020), thereby enhancing the researchers' ability to suggest appropriate interventions. For example, one might describe and differentiate sport fans' emotions and attitudes from their social media interactions using Linguistic Inquiry and Word Count (LIWC) and Latent Semantic Analysis sentiment analysis. LIWC is a transparent text analysis program that counts words in psychologically meaningful categories. Such results could help sports teams make predictions on appropriate tipping points for individuals to stimulate their fan identity according to the psychological continuum model (PCM). To this end, we argue that data-rich environments afford an opportunity to expand measurement approaches and enhance validity and reliability, improve sample size, and minimize misapplication of measurement instruments.

Building big data driven strategies in sport business

The distinctive characteristics of sport, such as its intangibility, inconsistent and uncertain nature, the value co-creating consumption process (Hedlund, 2014) and the connection between the training of athletes and business operations (Sharpe *et al.*, 2018), highlight the importance of BD in addressing the unique challenges faced by sport organizations. To develop an agile data-driven strategy, it is essential to integrate different units of the organization to derive a unified and accurate view of the organization's current operations and consumer demands. Sport teams' social media content, for example, is heavily influenced by their on-field performance. Through the use of BD analytics across units, the team can effectively manage content creation to meet the evolving needs of fans. As shown in Figure 3, we identify three areas (i.e. business analytics, human resource analytics and athletic and team analytics) where BD will help sport organizations navigate the sport business world. Additionally, we outline the challenges facing sport organizations as they undergo digital transformation.

Athletic and team analytics and marketing strategy

In sport business, team or athlete performance is closely related to the profitability of the sport organization. As indicated in Figure 3, the application units of BD in the area of athlete training and performance enhancement could greatly inform what strategies sport teams and athletes could proactively adopt for better effectiveness. Though BD is well integrated into performance analytics, little effort has been put into analyzing how athletic and team analytics can improve sport marketing. Traditionally, the applications of sports field data to market-related outcomes have been focused on TV viewership. Researchers have utilized player or team performance indicators, coupled with psychological measurements, to predict the viewership across various sports (e.g. Grimshaw and Larson, 2020). Using analytic stimulation technique, recent research has explored the relationship between player value, including performance quality and match significance with sport team market performance (e.g. Buraimo *et al.*, 2021). The application of metrics traditionally used in sport analytics could enhance sport marketing research as it offers an objective indicator that could have a spillover effect on fans' psychometric responses to the team.

Future research avenues

At the managerial level, little is known about how sport organizations can exploit player behavioral data or team strategy data to gain a marketing edge. Researchers are encouraged to apply analytics and stimulation to derive the player's value, significance of the match and geolocation of the game to predict a marketing-related outcome, such as brand value, consumer engagement rate and sponsorship value. In order to realize this potential, industry-

Big data in sport management

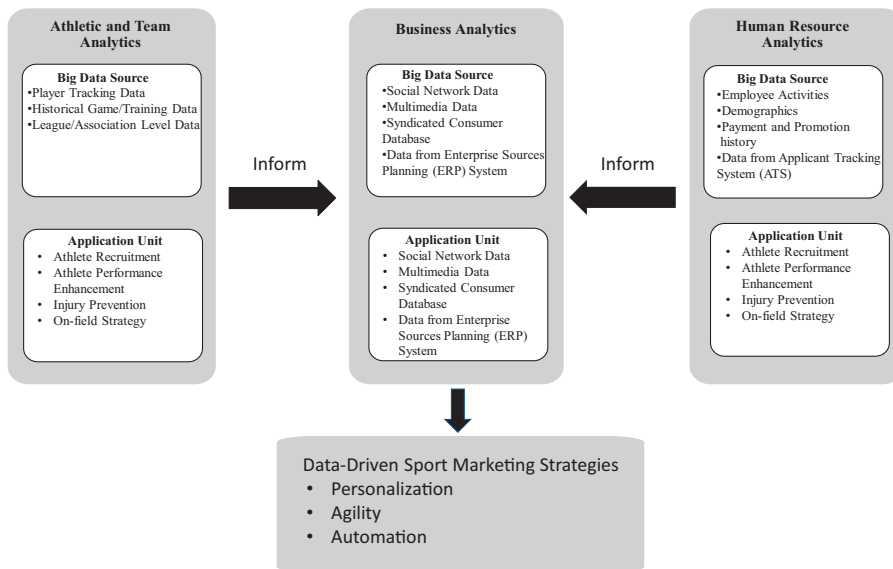


Figure 3. Conceptual model of big data application in sport organizations

academia collaboration will be required, including sport management scholars, data scientists, information technology (IT) specialists and team managers.

Moreover, further explorations could be conducted in some emerging contexts, such as sport betting and esports. For example, in sport betting, researchers could examine the possible impact of publicly available information about players' and teams' performance metrics on sports betting behavior and its downstream effects on consumer engagement. In the area of esports, the enormous amount of data generated by a digital game environment can help players improve their performance, resulting in increased customer satisfaction and retention since players are also consumers.

Human resource analytics and marketing strategy

Human resource analytics is also an important contributor to the data-driven organizational strategy (Putka *et al.*, 2018). The meaning of the term "human resource analytics" spans from describing the use of data to evaluate employees' work performance to the much more comprehensive designation of optimizing human capital and predicting the future demand for and supply of talent (Hamilton and Sodeman, 2020). BD has shown to have an effect on human resource management in organizational culture (Alharthi *et al.*, 2017), diversity and inclusion (Morgan *et al.*, 2015), and employee well-being (Axtell *et al.*, 2019). While limited attention has been paid to this research area, human resource analytics plays a paramount role in optimizing the role of BD in sport organizations because it strives to simultaneously meet the needs of team operations and business operations. Research has identified that the application of human resource management greatly impacts the marketing organization (Olson *et al.*, 2018). The structure, culture, processes, influence and leadership of sport organizations are critical factors for achieving marketing goals. It is particularly relevant in cases where salespeople play a major role in the outcome. Using human resources analytics in marketing strategy development can provide context about who can execute the strategy effectively, as illustrated in Figure 3.

Future research avenues

Researchers are encouraged to utilize human resources analytics as a tool to gauge the effectiveness of internal marketing. For example, what characteristics of frontline employees respond more positively to the internal branding strategies, which could potentially have a positive spillover effect to consumer behavior. Another area is how human resource analytics could be used to improve the brand image of sport organizations. For example, how does the recruitment process impact the effectiveness of cause marketing? BD in human resources could be used to create diversity indices, which can then be used to study the connection between minority employee percentages and the perceived authenticity of a marketing campaign.

In particular, researchers should pay attention to the potential digital divide created by implementing BD and data analytics in sport organizations. Sport organizations that lack diversity and inclusion can face particular challenges due to a digital divide. Thus, the increasing level of BD literacy within sport organizations can influence long-term issues of employee diversity and equality. It is important to note that, even if the research in this area does not directly involve using BD, it has implications for marketing research because it identifies the behavior change in sport organizations as a result of implementing BD.

Business analytics and marketing strategy

The analytics landscape has shaped the sport organizations' business strategies, especially fan engagement and revenue generation. However, BD has recently gained momentum in the field, the concept of business analytics dates back to 1990. The advent of BD application helps sport organizations leverage the unique characteristics of sport related to outcome uncertainty, experience variability, socialization and intense emotion. Marketing managers can glean insights from real-time social media responses (e.g. likes and comments), historical ticket sales data and longitudinal consumer spending patterns (Troilo *et al.*, 2016). For instance, fan-generated reviews of the organizations' services on social media platforms can be used to shed light on fans' over satisfaction with the service. With advanced marketing data collection methods, sport organizations can collect data at a more granular level for personalization marketing initiatives.

Future research avenues

As indicated in the previous section, researchers have started to use BD, particularly social media data, to yield insights for consumer behaviors that surveys have traditionally measured. By applying text analytics to these data to categorize fans' experiences, sport organizations can delve into the factors that drive satisfactory or unsatisfactory ratings. In a similar vein, data-driven metrics can be applied to dynamic pricing implementation, sponsorship effectiveness measurement, customer relationship management and corporate partnerships evaluation (Mondello and Kamke, 2014).

In the future, researchers are encouraged to integrate multiple sources of BD data to get a holistic view of consumer behavior. The sport teams could, for example, propose a model to predict fans behavior based on ticket and merchandise sales information, as well as unstructured user-generated data from their social media channels throughout the season. By combining information from different sources, data analytics allows teams to prioritize marketing strategies efficiently, such as identifying the optimal timing to send out promotional messages for team merchandise.

Another area that is worthy of investigation is challenges associated with consumer privacy and data maltreatment as sport organizations rush to implement BD analytics to keep up with their competition in the business world (Fried and Mumcu, 2016). An example might be what privacy regulations sport teams abide by and how they affect their fans'

relationships. While research examining privacy and ethics concerns might not directly involve BD or data analytics, its examination is vital to the sustainable and healthy development of sport organizations. BD can be a double-edged sword when a sport organization does not clearly assess its potential risks.

Conclusion

In light of the optimistic view that BD can offer exciting opportunities for the field of sport marketing, scholars and practitioners need to embrace new ways of data collection and analytics techniques. Two specific areas of BD applications have been discussed: BD sources and modern analytical techniques. Our review suggests that BD has transformative potential for solving complex problems through new data sources and related computational advancements. Whether we are thinking about business intelligence, consumer segmentation, fan engagement or communication strategies, BD can provide unique insight for sport marketing scholars. Furthermore, modern analytical techniques are crucial to process complex insights and absorb their meaning to make decisions quickly.

Despite all of its promise, BD in sport marketing may invite several unique challenges. Most of the user-generated data sport marketers access is pertaining to highly engaged fans, which is often not representative of the total target consumer base. While BD analyses can result in predictions and recommendations, algorithmic analyses typically struggle to answer “why” questions (e.g. why fans purchased one type of team merchandise rather than another) in a way that sport researchers and practitioners understand. Aside from these limitations, we firmly believe that recent BD advancement offers exciting possibilities for uncovering new ideas and questions and rich opportunities for interdisciplinary research.

References

- Alharthi, A., Krotov, V. and Bowman, M. (2017), “Addressing barriers to big data”, *Business Horizons*, Vol. 60 No. 3, pp. 285-292.
- All too much (2021), “The economist”, available at: <http://www.economist.com/node/15557421> (accessed 28 June 2021).
- Axtell, C., Taylor, M. and Wessels, B. (2019), “Big Data and employee wellbeing: walking the tightrope between utopia and dystopia”, *Social Sciences, Multidisciplinary Digital Publishing Institute*, Vol. 8 No. 12, p. 321.
- Balducci, B. and Marinova, D. (2018), “Unstructured data in marketing”, *Journal of the Academy of Marketing Science*, Vol. 46 No. 4, pp. 557-590.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W., Netzer, O. and Schweidel, D. (2020), “Uniting the tribes: using text for marketing insight”, *Journal of Marketing*, Vol. 84 No. 1, pp. 1-25.
- Boyd, D. and Crawford, K. (2012), “Critical question for big data”, *Information, Communication and Society*, Vol. 15 No. 5, pp. 662-679.
- Brock, V. and Khan, H. (2017), “Big data analytics: does organizational factor matters impact technology acceptance?”, *Journal of Big Data*, Vol. 4 No. 1, doi: [10.1186/s40537-017-0081-8](https://doi.org/10.1186/s40537-017-0081-8).
- Buraimo, B., Forrest, D., McHale, I.G. and Tena, J.D. (2021), “Armchair fans: modelling audience size for televised football matches”, *European Journal of Operational Research*. doi: [10.1016/j.ejor.2021.06.046](https://doi.org/10.1016/j.ejor.2021.06.046).
- Burton, N. (2019), “Exploring user sentiment towards sponsorship and ambush marketing”, *International Journal of Sports Marketing and Sponsorship*, Vol. 20 No. 4, pp. 583-602.
- Chang, Y. (2019), “Spectators’ emotional responses in tweets during the Super Bowl 50 game”, *Sport Management Review*, Vol. 22 No. 3, pp. 348-362.

-
- Delia, E. (2019), “‘You can’t just erase history’: coping with team identity threat”, *Journal of Sport Management*, Vol. 33 No. 3, pp. 203-214.
- DeSchraver, T., Webb, T., Tainsky, S. and Simion, A. (2021), “Sporting events and the derived demand for hotels: evidence from Southeastern Conference football games”, *Journal of Sport Management*, Vol. 35 No. 3, pp. 228-238.
- Du, J., McLeod, C., Baker, B., Gulavani, S. and James, J. (2021), *Churn Management for Season Ticket Holders Using Machine Learning, Presentation*.
- Editorial, R. (2021), “The world’s most-googled news stories of 2019 | Reuters Video”, *Mobile.reuters.com*, available at: <https://mobile.reuters.com/video/watch/the-worlds-most-googled-news-stories-of-id640338041?chan=77jbcblg> (accessed 28 June 2021).
- Franks, A., D’Amour, A., Cervone, D. and Bornn, L. (2016), “Meta-analytics: tools for understanding the statistical properties of sports metrics”, *Journal of Quantitative Analysis in Sports*, Vol. 12 No. 4, doi: [10.1515/jqas-2016-0098](https://doi.org/10.1515/jqas-2016-0098).
- Fried, G. and Mumcu, C. (Eds) (2016), *Sport Analytics: A Data-Driven Approach to Sport Business and Management*, Taylor & Francis, New York, NY.
- Funk, D. and James, J. (2001), “The Psychological Continuum Model: a conceptual framework for understanding an individual’s psychological connection to sport”, *Sport Management Review*, Vol. 4 No. 2, pp. 119-150.
- Funk, D., Lock, D., Karg, A. and Pritchard, M. (2016), “Sport consumer behavior research: improving our game”, *Journal of Sport Management*, Vol. 30 No. 2, pp. 113-116.
- George, G., Haas, M. and Pentland, A. (2014), “Big data and management”, *Academy of Management Journal*, Vol. 57 No. 2, pp. 321-326.
- George, G., Osinga, E., Lavie, D. and Scott, B. (2016), “Big Data and data science methods for management research”, *Academy of Management Journal*, Vol. 59 No. 5, pp. 1493-1507.
- Gong, H., Watanabe, N., Soebbing, B., Brown, M. and Nagel, M. (2021), “Do consumer perceptions of tanking impact attendance at National Basketball Association games? A sentiment analysis approach”, *Journal of Sport Management*, Vol. 35 No. 3, pp. 254-265.
- Grimshaw, S. and Larson, J. (2020), “Effect of star power on NBA All-Star Game TV audience”, *Journal of Sports Economics*, Vol. 22 No. 2, pp. 139-163.
- Hamilton, R. and Sodeman, W. (2020), “The questions we ask: opportunities and challenges for using big data analytics to strategically manage human capital resources”, *Business Horizons*, Vol. 63 No. 1, pp. 85-95.
- Hedlund, D. (2014), “Creating value through membership and participation in sport fan consumption communities”, *European Sport Management Quarterly*, Vol. 14 No. 1, pp. 50-71.
- Humphreys, A. and Wang, R. (2018), “Automated text analysis for consumer research”, *Journal of Consumer Research*, Vol. 44 No. 6, pp. 1274-1306.
- Jensen, J.A., Wanless, L., Hedt, K. and Wayland, E. (2020), “Analyzing Big Data in intercollegiate athletics: an application to athletic donor behavior”, *Journal of Issues in Intercollegiate Athletics*, Vol. 13 No. 1, pp. 292-311.
- Johnson, S., Gray, P. and Sarker, S. (2019), “Revisiting IS research practice in the era of Big Data”, *Information and Organization*, Vol. 29 No. 1, pp. 41-56.
- Koenigstorfer, J. and Wemmer, F. (2019), “What makes sports clubs successful at recruiting and retaining members from the perspective of managers? Results from a random forest analysis”, *Journal of Global Sport Management*, pp. 1-20, doi: [10.1080/24704067.2019.1701952](https://doi.org/10.1080/24704067.2019.1701952).
- Kwak, D.H., Clavio, G.E., Eagleman, A.N. and Kim, K.T. (2010), “Exploring the antecedents and consequences of personalizing sport video game experiences”, *Sport Marketing Quarterly*, Vol. 19 No. 4, pp. 217-225.
- Laney, D. (2001), “3D data management: controlling data volume, velocity and variety”, *META Group Research Note*, Vol. 6 No. 70, p. 1.

-
- Lazer, D. and Radford, J. (2017), "Data Ex Machina: introduction to big data", *Annual Review of Sociology*, Vol. 43 No. 1, pp. 19-39.
- Lopez, M. (2020), "Bigger data, better questions, and a return to fourth down behavior: an introduction to a special issue on tracking data in the National Football League", *Journal of Quantitative Analysis in Sports*, Vol. 16 No. 2, pp. 73-79.
- Mao, L. (2020), "Understanding retail quality of sporting goods stores: a text mining approach", *International Journal of Sports Marketing and Sponsorship*, Vol. 22 No. 2, pp. 330-352.
- Marr, B. (2021), "How much data do we create every day? The mind-blowing stats everyone should read", [online], *Forbes*, available at: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/> (accessed 23 March 2021).
- Matti, J. (2021), "Frustrated customers: the effect of unexpected emotional cues on yelp reviews", *Journal of Sport Management*, Vol. 35 No. 3, pp. 203-215.
- McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012), "Big Data: the management revolution", *Harvard Business Review*, Vol. 90 No. 10, pp. 60-68.
- Mondello, M. and Kamke, C. (2014), "The introduction and application of sports analytics in professional sport organizations", *Journal of Applied Sport Management*, Vol. 6 No. 2, pp. 1-11.
- Morgan, W.B., Dunleavy, E. and DeVries, P.D. (2015), "Using big data to create diversity and inclusion in organizations", in Tonidandel, S., King, E. and Cortina, J. (Eds), *Big Data at Work: the Data Science Revolution and Organizational Psychology*, Routledge, New York, NY, pp. 310-335.
- Nam, H., Joshi, Y. and Kannan, P. (2017), "Harvesting brand information from social tags", *Journal of Marketing*, Vol. 81 No. 4, pp. 88-108.
- O'Hallarn, B., Shapiro, S. and Pegoraro, A. (2018), "Hashmoney: exploring Twitter hashtag use as a secondary ticket market price determinant", *International Journal of Sport Management and Marketing*, Vol. 18 No. 3, p. 199.
- Olson, E., Slater, S., Hult, G. and Olson, K. (2018), "The application of human resource management policies within the marketing organization: the impact on business and marketing strategy implementation", *Industrial Marketing Management*, Vol. 69, pp. 62-73.
- Oswald, F., Behrend, T., Putka, D. and Sinar, E. (2020), "Big Data in industrial-organizational psychology and human resource management: forward progress for organizational research and practice", *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 7 No. 1, pp. 505-533.
- Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H. and Seligman, M.E. (2015), "Automatic personality assessment through social media language", *Journal of Personality and Social Psychology*, Vol. 108 No. 6, p. 934.
- Peetz, T.B. and Reams, L. (2011), "A content analysis of sport marketing quarterly: 1992-2011", *Sport Marketing Quarterly*, Vol. 20 No. 4, pp. 209-218.
- Pitts, B. and Stotlar, D. (2013), *Fundamentals of Sport Marketing*, Fitness Information Technology, Morgantown, WV.
- Putka, D., Beatty, A. and Reeder, M. (2018), "Modern prediction methods", *Organizational Research Methods*, Vol. 21 No. 3, pp. 689-732.
- Rehman, M., Chang, V., Batoool, A. and Wah, T. (2016), "Big Data reduction framework for value creation in sustainable enterprises", *International Journal of Information Management*, Vol. 36 No. 6, pp. 917-928.
- Rivers, D.J. and Ross, A.S. (2021), "'This channel has more subs from rival fans than Arsenal fans': Arsenal Fan TV, football fandom and banter in the new media era", *Sport in Society*, Vol. 24 No. 6, pp. 867-885.
- Sena, V., Bhaumik, S., Sengupta, A. and Demirbag, M. (2019), "Big Data and performance: what can management research tell us?", *British Journal of Management*, Vol. 30 No. 2, pp. 219-228.

-
- Sharpe, S., Beaton, A. and Scott, O. (2018), "Considering ongoing professionalization in sport organizations: a case study of the ACT Brumbies Super Rugby Club", *Journal of Global Sport Management*, Vol. 3 No. 3, pp. 215-236.
- Sivarajah, U., Kamal, M., Irani, Z. and Weerakkody, V. (2017), "Critical analysis of Big Data challenges and analytical methods", *Journal of Business Research*, Vol. 70, pp. 263-286.
- Smith, A. and Stewart, B. (2010), "The special features of sport: a critical revisit", *Sport Management Review*, Vol. 13 No. 1, pp. 1-13.
- Su, Y., Baker, B.J., Doyle, J.P. and Yan, M. (2020), "Fan engagement in 15 seconds: athletes' relationship marketing during a pandemic via TikTok", *International Journal of Sport Communication*, Vol. 13 No. 3, pp. 436-446.
- Tainsky, S., Mills, B. and Winfree, J. (2015), "Further examination of potential discrimination among MLB umpires", *Journal of Sports Economics*, Vol. 16 No. 4, pp. 353-374.
- The Huge Statistics of Sports on Social Media | Rosy Strategies (2021), "Rosy strategies", available at: <https://rosystrategies.com/the-huge-statistics-of-sports-on-social-media> (accessed 28 June 2021).
- Tonidandel, S., King, E. and Cortina, J. (2018), "Big data methods", *Organizational Research Methods*, Vol. 21 No. 3, pp. 525-547.
- Troilo, M., Bouchet, A., Urban, T. and Sutton, W. (2016), "Perception, reality, and the adoption of business analytics: evidence from North American professional sport organizations", *Omega*, Vol. 59, pp. 72-83.
- Uhlir, P. and Schröder, P. (2007), "Open data for global science", *Data Science Journal*, Vol. 6, pp. OD36-OD53.
- Watanabe, N., Yan, G. and Soebbing, B. (2016), "Consumer interest in major league Baseball: an analytical modeling of Twitter", *Journal of Sport Management*, Vol. 30 No. 2, pp. 207-220.
- Watanabe, N., Yan, G. and Soebbing, B. (2019), "Market disruption as a regime for athlete activism: an economic analysis of college football player protests", *Sport Management Review*, Vol. 22 No. 5, pp. 600-612.
- Watanabe, N., Shapiro, S. and Drayer, J. (2021), "Big Data and analytics in sport management", *Journal of Sport Management*, Vol. 35 No. 3, pp. 197-202.
- Wedel, M. and Kannan, P. (2016), "Marketing analytics for data-rich environments", *Journal of Marketing*, Vol. 80 No. 6, pp. 97-121.
- Yan, G., Steller, D., Watanabe, N. and Popp, N. (2018), "What determines user-generated content creation of college football? A Big-Data analysis of structural influences", *International Journal of Sport Communication*, Vol. 11 No. 2, pp. 219-240.
- Yan, G., Watanabe, N., Shapiro, S., Naraine, M. and Hull, K. (2019), "Unfolding the Twitter scene of the 2017 UEFA champions league final: social media networks and power dynamics", *European Sport Management Quarterly*, Vol. 19 No. 4, pp. 419-436.
- Yarkoni, T. and Westfall, J. (2017), "Choosing prediction over explanation in psychology: lessons from machine learning", *Perspectives on Psychological Science*, Vol. 12 No. 6, pp. 1100-1122.
- Youyou, W., Kosinski, M. and Stillwell, D. (2015), "Computer-based personality judgments are more accurate than those made by humans", *Proceedings of the National Academy of Sciences*, Vol. 112 No. 4, pp. 1036-1040.

Corresponding author

Yoseph Mamo can be contacted at: ymamo@tnstate.edu

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com