



# The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports

Devansh Patel<sup>1</sup> · Dhwanil Shah<sup>2</sup> · Manan Shah<sup>3</sup>

Received: 24 July 2019 / Revised: 19 November 2019 / Accepted: 28 December 2019 /  
Published online: 11 January 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Big data, artificial intelligence, data analytics, machine learning, neural networks are promising prospects in the industry right now and subsequently the posterity for the current technological landscape. Initially, we looked into the overwhelming amount of sectors it is involved in. In the auditing process, we were able to conclude; the sports industry seems to be one of the most promising applications of these modern technologies. Hence, this paper offers a deeper look into how the entire sports industry has been affected in a multi-faceted way. Not only, are the on field antics that have been impacted but also the business implications and immersion of fans. The following is a comprehensive review expanding on the aforementioned aspects.

**Keywords** Big data · Sports · Analytics, artificial intelligence

## 1 Introduction

Big Data is massive and complex; consisting of structured and unstructured data sets. The 5 main properties of Big data are infamously its 5 Vs; namely Volume, Velocity, Variety, Veracity and Value. Volume pertains to the sheer size of the data [1, 2]. Velocity is the sheer rate at which data is accumulated and processed, especially in real time applications or near real time applications like sensor event log from live matches. Variety refers to the diversity in composition of this data and the newer data sources which provide data in a plethora of formats. External data sources from third-party content, often via cloud based providers, can change their data structure without notification to downstream organizations using that

✉ Manan Shah  
manan.shah@spt.pdpu.ac.in

<sup>1</sup> Department of Computer Engineering, Indus University, Ahmedabad, Gujarat, India

<sup>2</sup> Department of Computer Engineering, L.J. Institute of Engineering and Technology, Ahmedabad, Gujarat, India

<sup>3</sup> Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar, Gujarat, India

information; so it becomes difficult to further process the data [3, 4]. Veracity refers to the truthfulness of the data and questions the sources, methodologies and technologies of the data [5]. It is often found that the data accumulated has biases, ambiguities and inconsistencies. Value extraction becomes difficult due to the aforementioned reasons and data has intrinsic value but it is of no use until discovered. It is due to these reasons that traditional databases and software techniques are incapable of operating on this data [6–8].

Outlining the procedure of processing big data; it starts with collecting information, succeeded by cleaning of the data and deciding whether it is suitable for performing further processing then finally performing analytics to mine meaningful relations and results which are further tested for veracity. Needless to say that the process is complex and gaining a deeper understanding has motivated the main focuses of the paper; how big data, sports analytics and other related technologies have affect the current state of sports. For the intents of this paper, “sports analytics” will be defined as, the management of organized historic data, the use of predictive analytical models that employ the data, and the usage of information systems to inform decision makers and permit them to aid their establishments in attainment of a competitive benefit on the ground [9].

The main goals of Sports Analytics are to gather data from varied sources which is stored in a cohesive presentation displaying all the relevant information in an integrated manner [10]. Thus, helping in collecting and showcasing the suitable information in an organized manner speeding up the decision making process.

The other objective of any good sports analytics program is to provide information that is not obvious to the decision maker; information that is new and useful. But as the data gets more and more humongous and complex, to derive information from it also becomes a difficult task. Analytical models provide information that is not obtainable without advanced statistical analysis and combining this new information with the insights of sports experts leads to a better judgment of the game eventually leading to better results [11].

Sports Analytics not only involves the players and professionals, it also involves the fans. Big data, if applied correctly, plays a vital role here by improving their engagement with the game.

During the game big-data could help further enhance the live feel and enrich supporter involvement. Live features involve fans which could create a more hearty experience for the people involved and deliver a greater entertainment value. Looking beyond this aspect, greater knowledge of attendees could make pursuing fans more efficient. Once the attendees’ profiles have been processed specific sponsor can be targeted that pertain to the interests of the attendees and look to understand the psychology of the non-attendees. This in turn will help increase business revenue to bolster the monetary side of any sports organization. Sports analytics not only affects the game but the overall fiscal performance of any sports team when viewed as a business organization. Sports and business analytics result in increase in year on year revenue, reduce overhead costs, calculated risk taking and efficient utilization of available resources whether it be man-power, time or monetary [12].

Based on the current state of the field, it is apparent that decision making, fan engagement and business strategies are the most affected area of the sports world

whether it be positive or negative. Decision making here pertains to decisions made by the coaches on and off the field. It is said that decision making is one of the most impacted areas because it can change the outcomes of games drastically. Sports decisions affect the coaches' signing of players by studying the market, it ranks the players of the teams by judging performance, and it affects in-game decision and assesses the opposition team which is helpful in devising game-day strategy. Fan engagement is an important aspect of any sports organization because it increases revenue and gets the fans to the stadium. Lastly, business strategies are affected widely which can be supported by a simple example [13].

The weekday games of the Atlanta Hawks started at 7:30 pm and then in the following season the time was altered to 8 pm. Steve Koonin (Atlanta Hawks Chief Executive Officer) explained that the decision was "data based". The change in starting time was a direct consequence of analyzing the relevant data. An example of the data analyzed is the times of arrival of the attendees. The organization also analyzed local traffic patterns to determine whether pushing back the game start time would result in a more efficient fan commute to Philips Arena. This data-driven business decision aligned with customer input, and the change resulted in an increased number of fans in arena seats at game tip-off time. It is due to this reason that these sections will be focused on.

## 2 Decision Making On & Off the Pitch

For the first section we started with how Bigdata, AI and other latest technologies have actually impacted the game i.e. the playing part of the sport. The decisions that are made on and off the pitch are based upon the outcome of processing huge amounts of data and various algorithms as elaborated below.

Computer Intelligence algorithms, like Artificial neural networks (ANN), Swarm Intelligence (SI) etc., are algorithms inspired by natural systems. These algorithms are extensively applied in solving problems in the sports domain. Mobile tracking devices produce large amounts of data in the form of TCX files. Data is categorically sorted about all the athletes' performance over a specific short interval of time but it is too complicated to be understood by the coach manually and it is even more to make a decision based on the data. Computer Intelligence algorithms are used to mine data and help deciding the amount of training sessions and their period, detect over training leading to injuries, fatigue and other side effects, also help pick the right nutrition plans for the athletes. Fister proposed the bat algorithm for planning the amount of training sessions, Novatchkov & Baca introduced the fuzzy logic algorithm for assessment of power exercise [14].

Intensity, collisions, and fatigue data sets of each athlete is monitored and IBM's predictive analysis is used to provide us with accurate and dependable measure of forecasting each and every player's boundary and their respective injury risk in order to adjust their drills [15]. SAP's Injury Risk Monitor is another similar technology used by elite sports' teams to monitor the health of their players and thus prevent any injuries [16].

Players' personal reactions to high pressure situations are also gathered which help understand the players' physiological patterns. This statistic is vital in Basketball where it is a crucial factor in analyzing a players shooting performance. Insights from this data is plotted in graphs for the experts to analyze [17].

Furthermore Analyses like Movement and Constellation-based Analysis, Single player Analysis Multiplayer Analysis, Event-Based Analysis, Cluster Analysis in Weka and Shot-Event Feature Pattern Analysis are performed. These kinds of semi-automatic analyses are very useful and this is supported by the account of soccer adept, who has been involved with soccer since 31 years of which he spent 23 coaching and 9 playing. Right now, he is working for the successful German Club Bayern Munich. He is sure that the repercussions of a semi-automatic tool would be positive and he is certain that these tools could be applied in soccer. He says that the semi-automatic analysis will assist managers and cases alike, in cases where there isn't enough time for a manual analysis and it allows analyzing larger data in a similar time frame [18].

Baseball team, Oakland Athletics are currently relying heavily on analytics to sign players and pick the playing team and they have been outplaying teams with much deeper pockets i.e. more resources. Many organizations use exponential amounts of data and refined machine learning algorithms to meet customer and managerial needs [19].

Machine learning techniques on soccer match data can be used to identify many things in the game, starting with the most important attributes of a player's performance which determine the ratings associated by soccer experts. The second piece of information that can be analyzed is that which performance attributes of the players of the two competing teams affect the match outcome and to what extent the match outcome is impacted by these characteristics. Among other things team rating and individual player ratings can also be generated and finally based on a few games as training for the model the prediction of future games [20].

In addition to this, machine learning in soccer can present two approaches to deriving insights into the sport by analyzing characteristics of passing. Firstly, the passing style of a team is considered to create heat maps which in turn create a fingerprint that can be used to identify teams with better passing and identify a pattern of passing which remains consistent for a certain team through the entire season. The second approach shows the location of the pass origins and destinations which predict whether the play will end in a shot at goal [21].

In the sport of baseball, Pitchf/x technology from Sport vision has been tested and implemented in all the Major League Baseball (MLB) stadiums to track of pitches thrown during the course of the entire season. Sport vision has a lot of other technologies for other sports including baseball, football and motor sports. However, nothing has replaced the sometimes insanely hard, judgment calls umpires have had to make in real-time. Hank Adams, CEO of Sportvision says "Sportvision technology is being adapted to use for referees and umpires. We can very accurately determine if something is a strike or a ball."

### 3 Fan Engagement

In this section authors moved on the fans side of any sports organization which is crucial to any team because a recurring fan base is what generates the revenue needed for any club to function. Sports analytics, big data, AI, machine learning and other technologies are applied to make the experience for any fan more engaging and immersive so as to keep them coming back and also to expand the fan base exponentially which results to a direct increment in year on year revenue.

As stated before, fan engagement results in an increase in ticket sales, but it could be argued that the opposite might be true as well. To increase ticket sales, sports organizations are employing methods to learn arrangements and analyze tendencies for impactful forecast of upcoming customer comportment. In a particular paper, the research was done in two stages. Firstly a qualitative approach was taken to answer the, “what” and “how” questions, like how to increase revenue or sales. The latter part is a quantitative approach using data mining methods such as clustering, classification, rule mining, relationship modeling to try to understand how ticket sales has performed over time. The main objective of this study was to understand past sales and use this information to predict future sales and implement any recommendations or interventions that might be suggested [22].

Artificial Intelligence helps make the viewing more immersive and overall “better” using Chatbots, Automated journalism, Wearable tech and computer vision. Chatbots answer the queries of the fans when the game is live without any human input. Computer vision is a kind of technology used in motor racing sports. Media channels use artificial intelligence to widen the scope of their sports coverage reportage. Wearable technology is extensively used in the sports training world to track training of players as well as their in-game performance. KAI is an example of a chatbot which is currently being used by the NBA team Sacramento Kings. It can answer questions about the histories of the organization, the current state of the team regarding team status and statistics and details about their home stadium. Another example of a popular chatbot is the one that Tampa Bay Lightning use, called Lightning. It can answer questions about their home stadium, parking info and game tickets. The racing sport of NASCAR applies algorithms of artificial intelligence and neural networks to make out the issues with cars before the race so as to avoid dangerous situations during the course of the race. Thus, reducing the risk to themselves and the drivers considerably [23].

In the near future Big Data, AI and natural language processing will completely revolutionize live video production companies and the technology they use by using technologies from Google, IBM, NVidia and Convivia. All these companies produce what they call Video Artificial Intelligence software. Google offers Cloud Video Intelligence, IBM has Watson on offer, NVidia has their very own DLA and last but not least Convivia has Video AI Architecture. AI aims to help the video production teams by tracking the subject in focus without the need of any additional hardware. The technical director is a human in the video production team who does this manually, commanding to change camera angles

and focus lights. AI aims to completely put these people out of a job. AI will also be able to create the highlights reel on its own by recognizing the best viewing angles for the current action being performed. It will take into account facial expressions, gestures, emotional cues, body language, color of all the elements, clothing of the people in the shot and other imaging data cues. All of this will work towards enhancing the spectator experience.

With the current state of the market these has been a steady reluctance to adopt these AI based technologies. Once a concrete correlation is established with using AI technologies and improved performance the reluctance will slowly wither [23].

The introduction of big data has also enabled fans to newer means of evaluating and discussing their players and teams of interest. NBA has collected nearly every statistic in league history and made it accessible as an online search engine. This engine is driven by SAP's HANA platform—which is capable of handling 4.5 quadrillion combinations of data [24]. Manchester City, an English soccer club, released extensive player and team data with the intension of allowing the common fan to analyze these stats like a professional expert [25].

Within the stadium, richer data helps in solidifying fan engagement and animate the live game experience. What if fans could submit song requests for the in-game playlist, then be featured on the jumbotron when their request is played. Looking beyond the in-game features, collecting data about the profiles of the attendees can help better target the fans who attend live games throughout the season and those who don't. Data from mobile application can also be collected and further be used to analyze fans behavior and offer personalized offerings and promotions. In general the data needed is out there but now combined with the latest technology it can be used to obtain considerate insights not only about the game but also regarding the fans that make up the game [12, 15].

National Football League (NFL) made available a vast array of data sets to Fantasy Football participants through systems that make use of in-memory processing and data analytics capabilities. This peaked the interest of the Fantasy Football users and the platform was able to grow exponentially after its implementation. Similarly, data analytical tools have become critically important to engage the tech-savvy sport consumers, who now don't have to rely on expert analysts to study their favorite team and/or player's performance [26].

## 4 Business Side

Moving on to the impact of Sports analytics, big data and other modern technologies on the side that generates revenue and enables the sports organization to stay afloat i.e. Business side. It is interesting to note at this point that, in sport a multitude of dimensions have been affected by technology and the more we dive into the subject different sides to this can be revealed including the extent of impact.

Getting back to the subject of the section, Business Analytics do increase revenue as observed by sports managers. This was tested by a group using the Wald Chi squared test, the pseudo R-squared statistic and Analytics Index. Professional sports organizations in the continent of Northern America are in an exclusive business that

allows duopoly i.e. power divided over just two or three businesses in their local markets. However, to maximize profits and revenue these organizations are forced to turn to analytics in light of recent findings support the hypothesis that analytics do improve the revenue and over economic rationality. A recent finding states that the adoption if analytics increases revenue up to 7.2% in the year following the adoption which is a rational result since examining data trumps guesswork.

Apart from increasing revenue, sports organizations are being more and more compelled to incorporate analytics so as to get out of the increased debt that they are taking on. In earlier times, sports organizations use to rely a fair amount on the public subsidies to secure new facilities but since municipalities re moving towards more and more privatization, it has become difficult for them to rely on public funds [27].

It is tremendously significant to create a “good” fan experience so as to drive more revenue and create more revenue and value. The monetary, advertising and sales departments try to better their strategy based on data available. The ROI for any sports organization is improved by data [23].

When analyzing the business side of sport, it is quite encompassing and covers a plethora of fields. It covers economic analysis like one would expect but also the allotment of resources when picking pro teams. The pricing of tickets is also determined through such an analysis. For instance, a study was done into the overall long term growth of a city in which the Olympics have been hosted. The growth of a nation was gauged by picking the host cities that were hosting the finals but were not selected by the International Olympic Committee. The study resulted in showing that there was no direct correlation in hosting the Olympics and growth in terms of revenue, Gross Domestic Product (GDP) and trade willingness [28].

Use of technologies to gather and analyze big data relative to the business side of sports has become widespread which can be observed in the information. We can consider examples in different sports like the National Football League (NFL), a study carried out by Borghesi inspected the association of compensation and performance during the 10 consecutive seasons of NFL, it exposed that teams that paid the players the maximum disproportionately were revealed to be the worst performing teams. Big data sets are also analyzed by management for efficient ticket appraising in such a way so as to increase year on year incomes for sports organizations. Since the 2010s a dynamic pricing strategy was introduced in sports, in which the values of tickets would change on a daily basis to meet the current market conditions.

The sports betting industry has also used the introduction of the latest technologies to their advantage. Both consumers and supplier use machine learning algorithms to predict the outcome of the game. Large data sets consisting the previous outcomes are fed into cutting edge algorithms to accurately predict the outcome of the future games so that the bets of winning can be swayed in favor [28].

This field has created new work opportunities as big data analytics of sporting activities have been the foundation of many third party companies. One such company is Opta (recently acquired by the Perform Group) which entitles itself to be ‘the world’s leading sports data company’ servicing enumerable audiences in over 3 dozen countries. It stores data of a variety of sports but mainly focuses on soccer and rugby clubs. Its clientele thus includes soccer and rugby clubs but also media



houses, betting sites and video coverage companies. It also provides this data to the general public by representing it in user friendly graphs and charts.

Similarly, STATS that started out as non-profit organization is now providing huge amounts of information to some of the most affluent leagues of sports internationally, which includes the National Football League (NFL), Major League Baseball (MLB), the Professional Golfers' Association Tour (PGA Tour), National Association for Stock Car Racing (NASCAR), Union of European Football Associations (UEFA), Nippon Professional Baseball League and the Canadian Football League. These contracts deliver 'personnel management, player evaluation and game preparation' systems to leagues and teams that promise 'objective analysis', 'operational efficiency' and an 'integrated data platform' [29].

Hence, we can observe here, that not only the playing part of the sport, or the processes on the pitch but also off the pitch for the fans and sports organization (financially). Big Data has revolutionized the way everything works and has completely changed the vision moving forward and based on current evidence, we can easily state that this 'new' technology will continue to revolutionize the sport for years to come and will completely transform the way we think about the any sport and how the sport is played.

## 5 Case Studies

Moving forward, we decided to do case studies on the sports of Football & Formula 1 to better understand the impact of the newer technology on this sport particularly.

For this authors have used four sources. Fried and Mumcu [30] talked about the impact made by Big Data Analytics in NJ Youth Soccer when he was the Executive director of the aforementioned association. Authors said that the huge statistical information impacted the association in many ways. They took frequent surveys to improve customer satisfaction and tweeted on trending topics like the Women's World Cup to ensure that their tweet reached the maximum amount of people. They employed data mining to learn about the players enrolling for the Olympic Development program. This resulted in answers of several question which included; which demographic was applying more and why. Using the data to identify patterns helps them make changes to increase the turnout. Similar approach was applied in the association soccer to understand where to invest their promotion funds for best results [30].

Barros and Leach [31] stated that for a football club success on the pitch is essential but it doesn't necessarily translate into positive financial results. Such financial deficit reverts back to affect the performance on the pitch. Researchers, Carlos Barros and Stephanie Leach, apply the Data Envelopment Analysis (DEA) to gauge the performance of English Premier League football clubs from 1998/99 to 2002/03 taking into account game and fiscal variables. DEA is a linear coding method that allows the administration to yardstick the best-practice decision units. Also it identifies ways to improve the inefficiencies in these management units.



The result from the DEA indicates that a larger population base and economically more capable areas enable clubs to enjoy greater revenue. Scale of the club is the biggest factor in sporting efficiency, highlighting the importance of a strong local fan base in addition with an international fan base. Clubs showcase different managerial styles however what remains constant is that inhabitants in the team area is a main motorist of financial performance [31].

Chiappori et al. [32] aimed to test the predictions of game theory using penalty kicks in soccer as the given data which is mined from the French and Italian leagues. A general model is developed allowing for heterogeneity across players and demonstrate that the most basic outcomes that the model gets right so as to assure its viability for further analysis. The analysis provides us with 3 propositions, the first being that the kicker will choose to strike the ball in the center are of the goal more frequently than the goalie will guess to stay center for the save. The second is the kicker and goalie are more likely to go to the left(kickers left) than right and the third being that the goalie play left(the kickers natural side) more than the kickers do. Finally an additional prediction is that the strategy of picking a side should not be based on the previous outcomes, that is, it shall be truly random. In conclusion the paper gives us 3 predictions and proves that the empirical results are consistent with the predictions of the model [32].

Tactics have been a key component for success in modern soccer. With the introduction of advanced tracking technologies there has been an influx of data which has become difficult to manage. Rein and Memmert [33] discussed how big data technologies have opened up new chances to study tactical behavior in top football. Tracking data, Physiological data, coaching data, Scouting data and Crowd data from various sources are stored in a database after structuring of the respective data sets. Machine Learning techniques and big data technologies are then used to obtain models from the data sets influencing the tactics of the team. In conclusion, big data technologies have helped soccer teams understand tactics better and also innovate new strategies on the basis of the data obtained. The amalgamation of computer with soccer has led to refinement of tactics and strategies throughout the soccer community [33].

We believe that big data and along with it the technology to analyze that data has provided football clubs with greater power to measure every metric available to them in order to maximize their performance on and off the pitch. Football is an extremely competitive sport and one where there is a lot of money being invested in it. Teams are now willing to invest in these technologies as well to get an edge over their competition. Initially it started with gathering player biometrics to ensure that each individual is performing to their level best. Later, gathering of team data lead to influencing team tactics and in-game decisions. Recently, off field issues; such as atmosphere inside a stadium, fan engagement with its team & players and business decisions have all been influenced by the incorporation of latest innovations in big data analytics. Every aspect of this beautiful sport is now regulated by data analytics whether for the good or the bad of the sport. At the end, success is what'll football clubs aim for and big data analytics has provided these clubs with another tool to sway the results in their favor and they are not shying away from its use.

Continuing onto the sport of Formula 1; where big data, machine learning, AI, sports analytics have already been in use for some years and have already caused a major change in the current scenario of the sport. For this, again, we used four sources.

In this paper, the main objective of the authors is to scheme a forecasting system for live use throughout the course of a pro motor race. The author potentially aims to greatly influence how contesting sides enhance their racing plan by adjusting tire-pressure to improve the final outcome. The work done is something that builds upon previous work by deeply diving into the field. The experiment resulted in building of prediction models for each group. Further conclusions drawn from these were that the later parts of races tend to have a changed style than the starting parts of the race. For example, the distance to finish is lesser, so the amount of risk willing to be taken changes. The consequence to this is more violent driving and greater amount of cautions. Therefore, two-tire decisions are comparatively of greater importance and are thus observed more at that time. The experiment came to an end, concluding the problems in creating a prediction problem that results into the enterprise of planning and decision making tools for strategy building purpose by using knowledge in the domain and converting time-series data into a supervised learning platform [34].

Bouchet et al. [35] discussed an insight into a multi-billion dollar industry that is the Formula 1 racing sport being watched by hundreds of millions of viewers all over the world. The author aims to compare data and determine the fastest lap times, speed of F1 cars, best drivers and rivalries, etc. As a result, box plot analysis and outlier analysis were some of the analytical processes performed on the data used on the data and interesting outcomes appeared. In general we see a decrease in the average fastest lap per year, for almost all circuits. It looks like there was a trend, with an increase in the time to complete a single lap. Then from 2014 and onwards, it plateaus a bit. Some other interesting outcomes based on the results were; Michael Schumacher holds the highest number of victories as a single driver but Ferrari has the highest number when constructors are considered and Prost-Senna was the greatest rivalry that ever existed. Conclusions drawn from this are the all-time greatest drivers of the sport were identified including the best constructor and driver partnerships where McLaren, Ferrari and Williams dominated.

In this study the author talks about the enormous number of variables involved in winning a race. A minor tweak to one variable has a cascading effect on the remaining variables. Mercedes-AMG uses Data processing and analysis techniques to optimize the configurations and find a winning formula. As a result, 500 GB of data is collected in a single race weekend, amounting to 10 TB over the course of the season. Analytics of this data is used by Mercedes-AMG team to make critical decisions which the team hopes puts it in a position to win. In conclusion, as Matt Harris (Head IT AMG) believes, finding the right balance between data analysis and human intuition to reach to the very top of the competition is a key aspect. Emerging machine learning and deep learning frameworks have thus provided the team of Mercedes with newer capabilities to process and predict from the previously collected big data sets (<https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powers-f1.html>.)

Woodie [36] highlighted that F1 racing is a multi-million dollar sport, where every second is crucial to the winning of a team, and hopes to help them do the same. Since the teams are constantly looking for an edge that enables them to win, they have incorporated data analytics to do so in an effort to overtake their rivals on and off the track. The result is that F1 has become the pinnacle of vehicular technology with the aid of data analytics. Data analytics is used extensively in every aspect of F1 racing like in no other sport. Ranging from the making of cars, driving them to the broadcasting of the race. Today, there are approximately 200 sensors fitted onto the F1 car. Data analysts process the big data to reveal insights, obscure to the naked eye. This data has been a boon for the broadcasters as they use this real time data, provided by the constructors, and display it in an understandable format to the audience resulting in a more engaging experience for the viewer. The conclusion drawn from all of this is that now, data analytics is used to enhance the performance of the car, devise the optimal strategy and also to a greater extent assist in decision making. Success is now not only down to the drivers but a collective effort of the entire team.

Formula1 racing is a sport which depends on the race car as much as the driver and its crew. It didn't take long for big data to impact each of these aspects of Formula1 racing. From car design, to practice runs and driver biometrics, every detail is stored for further analytics. Every Formula1 side has a dedicated data analytics team whose sole purpose is to store and analyze these massive datasets. On race weekends, it becomes vital to process the incoming data in near real time so that quick decisions can be made proving to be vital in the team's success. It is a sport where competition is as fierce as it can be; even a millisecond can determine the outcome of the race. Data analytics is hence employed to achieve success even if it means by saving a second here and there.

## 6 Future Scope and Challenges

Big Data can surely have its pitfalls. The word "Big Data" in itself is an understatement, the data that comes through is massive not only in terms of its size but many other aspects such as the sheer different number of formats it comes in. So, an initial challenge is developing a data set on which processing can actually be done. So with the massive data set, many statisticians and computer scientists say that there could be an alarming number of incorrect relations. Which means that the relations mined out of the data set can be erroneous. As is, Big Data Analysis is like looking for a needle, a meaningful one, in haystack. A statistics professor at Stanford Trevor Hastie, builds on this by saying, "many bits of straw look like needles" [37].

To understand the challenges more comprehensively, one must take a look at how data is gathered. The game is recorded in sequence of Who, What, Where, When. Who meaning the team or member of team in possession of the ball. What meaning the action that is linked to the "Who". Where meaning the exact co-ordinates of the action being performed and when meaning the time in the game at which point the action took place. The difficulty is in understanding the Why and How of the game

which are critical to comprehend the underlying meaning of the actions and mine actual cause and consequence relationships [38].

The data that is gathered is then used to mine meaningful relationships, which is something that has been established. Predictions are made based on the relationships that mined. The problem with this is that, the predictions made might not always come true which is attributed to human behavior which is always unpredictable and the team might change their style or pattern of play on purpose so as to not fall into a predictable rut which can be countered [39].

Player performance indicators are a popular result of processing big data. What these “performance indicator” fail to take into account is opponent interactions.

As for example, a player might be judged to be an excellent performer even if he is playing for the losing team but might be judged as an average performer if he is on the winning side. If the performance indicators of a certain player are strong they attribute to a strong performance and weak ones attribute to a weak performance. The performance indicators of a certain player take from or adds to the whole performance of the team and also takes away or adds to the performance of the opposition team. To gauge team performance and provide accurate performance indicators based on the dynamics of both teams is indeed very difficult [40].

An ex Basketball pro who is currently an analyst for Inside the NBA, Charles Barkley, said I’ve always believed analytics were crap. They’re just some crap that some people who are really smart made up to try to get in the game ‘cause they had no talent” [41]. The problem with Big Data is that data must be collected on and off the pitch. This incurs a privacy problem for the athletes when data is to be collected regarding players health. Even if the player consents to this health data, the study calls for going deeper in terms of understanding the genome of any pro athlete to understand how some people are naturally better at sports and also an injury analysis. Moreover, the data for the personal health of these athletes could be misused [42].

Enumerable managers and coaches have deterred from adopting this data driven approach. However, data analytics is here to stay, in spite of its drawbacks, data will always reign supreme.

Establishing that Big Data is here to stay we talk about the scope of its applications in sports.

Not only in sports but, in the last decade, the implementation of Big Data and the latest technology have overtaken all sectors of the industry, whether it be business, healthcare or any other industry [43]. For example, in medicine the Human Brain Project and the US BRAIN Initiative, are trying to mimic the activities of the brain’s inner workings in a supercomputer simulation. Also, designing of the billions of neurons helps identifying solutions to problem like Alzheimer’s and Parkinson’s. In addition, data sets of all patients can be analyzed to discover cures to previously incurable diseases and also help prevent other patients from getting them [44].

Moving on to the crux, in sports the recent surge in data and latest innovations in technology keep pushing the teams to implement these tools to outperform their opponents on and off the field. As [45] writes, “numbers determine who holds power, and whose claim to power is justified [45]. Big Data in the future can be seen impacting other aspect of the sporting world like:

- Optimization of sports' equipment
- Optimization of different player positions on the field
- Detection of doping in sport
- Generation of training course
- Detection of athlete crisis during endurance competitions
- Avoiding pain and over training
- Deciding the dates of finals

Also, there are many researches going on in developing an artificial sports trainer that can eventually replace the real trainer in the future. A virtual trainer has large data sets trained so it can predict the next training schedule of the player according to his/her current biological factors. Big Data in sports analytics is estimated to become a 4 billion \$ industry by 2022 thus, and we can't stress this enough, it is safe to say that it is here to stay [46].

## 7 Conclusion

The innovation of big data and its associated technologies have influenced a wide range of markets. It did not take long for sports organizations to take advantage of this widespread technology to get an edge over the opposition. As is teams are constantly on the lookout for any factor that can provide them with an edge over their competitors and the introduction of big data has done just that. Data mining was performed previously, but now the tools are present that can utilize the data to realize patterns and meaningful relations that can shed light onto the many aspects of the game and any sports organization would be remiss if they aren't already using these tools available at their disposal in doing so. This in turn leads to gathering of more data and the cycle continues. First noticeable impact of data analysis in sports was when a baseball team Oakland A's general manager used it to scout and recruit players and build a team that won the World Series despite of having a tight budget. This incident is depicted in the film Moneyball and it made the world take notice of the use of this technology. Since then, it has spread across every aspect of the game ranging from player monitoring, in game decision making, strategy planning, off field management, monetary decisions, fan engagement and so on. Using these technologies have yielded successful results making more and more teams willing to adopt these strategies. Also, all sports have carved out an application of big data to suit their demands. The exponential growth in demand of these technologies have led to further innovation in this sector leading to newer technologies for the teams to use. This data driven approach has led to many positive as well as negative consequences as highlighted in the previous sections of this paper. Data mining and analyzing is always going to come with its caveats but through the entire research time period of this paper we noticed that it does definitely provide results, whether it be year on year monetary growth or better passing patterns in game or an enhanced experience for the fans. It is everywhere and it has been proven, decidedly so. Data mining always comes in with the privacy and misuse afflictions but when have there not been two sides to a coin? It is our belief that with the right security measures

some amazing results can yielded out of such a beautiful technology. To end the paper, we must emphasize the fact that, the analytics approach provides a method to madness. All kinds of decisions are now made not solely on intuition but also backed by vast data which is analyzed into patterns. Human emotion and biases can now be minimized effectively. Hence, we can say that Big Data in sports analytics is here to stay even with its hitches. Simply put, the pros outweigh the cons.

**Acknowledgements** The authors are grateful to Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University and Indus University for permission to publish this research.

**Author Contributions** All the authors make substantial contribution in this manuscript. DP, DS and MS participated in drafting the manuscript. DP and DS wrote the main manuscript, all the authors discussed the results and implication on the manuscript at all stages.

**Funding** Not applicable.

**Data Availability** All relevant data and material are presented in the main paper.

## Compliance with Ethical Standards

**Conflict of interests** The authors declare that they have no conflict of interests.

**Consent for Publication** Not applicable.

**Ethics Approval and Consent to Participate** Not applicable.

## References

1. Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. *Artif Intell Agric* 2:1–12
2. Shah G, Shah A, Shah M (2019) Panacea of challenges in real-world application of big data analytics in healthcare sector. *Data Inf Manag*. <https://doi.org/10.1007/s42488-019-00010-1>
3. Kakkad V, Patel M, Shah M (2019) Biometric authentication and image encryption for image security in cloud framework. *Multiscale Multidiscip Model Exp Des*. <https://doi.org/10.1007/s41939-019-00049-y>
4. Pandya R, Nadiadwala S, Shah R, Shah M (2019) Buildout of methodology for meticulous diagnosis of k-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence. *Augment Hum Res*. <https://doi.org/10.1007/s41133-019-0021-6>
5. Rubin V, Lukoianova T (2013) Veracity roadmap: is big data objective, truthful and credible? *Adv Classif Res Online* 24(1):4–15
6. Jain A, Bhatnagar V (2016) Olympics big data prognostications. *Int J Rough Sets Data Anal* 3(4):32–35
7. Zikopoulos P, Eaton C (2011) *Understanding big data analytics for enterprise class hadoop and streaming data*. McGrawHill, New York, pp 1–166
8. IBM (2017) *Making sense of big data a day in the life of an enterprise architect*. IBM, New York, pp 1–8
9. Alamar BC (2013) *Sports analytics a guide for coaches, managers, and others decision makers*. Columbia University Press, New York, pp 1–127
10. Katal A, Wazid M, Goudar RH (2013) Big data: issues, challenges, tools and good practices. In: 2013 sixth international conference on contemporary computing (IC3), pp 404–409

11. Miller TW (2015) Sports analytics and data science winning the game with methods and models. Pearson Education Inc, Prentice Hall, pp 1–225
12. Deloitte (2017) Upping your game how data can help drive sports sponsorship and fan engagement. Deloitte, New York, pp 1–8
13. Harrison CK, Bukstein S (2016) Sport business analytics: using data to increase revenue and improve operational efficiency. CRC Press, Boca Raton, pp 1–218
14. Fister I, Ljubic K, Suganthan PN, Perc M, Fister I (2015) Computational intelligence in sports: challenges and opportunities within a new research domain. *Appl Math Comput* 262:178–186
15. Millington B, Millington R (2015) The datafication of everything: toward a sociology of sport and big data. *Soc Sport J* 32:140–160
16. Joong KN, Keun PJ (2015) Sports analytics and risk monitoring based on Hana platform sports related big data & IoT trends by using HANA in-memory platform. 2015 international SoC design conference (ISOCC). *Gyungju* 2015:221–222
17. Zuccolotto P, Manisera M, Sandri M (2017) Big data analytics for modeling scoring probability in basketball: the effect of shooting under high-pressure conditions. *Int J Sports Sci Coach* 13(4):569–589
18. Janetzko H, Sacha D, Stein M, Schreck T, Keim DA, Deussen O (2014) Feature-driven visual analytics of soccer data. In: *IEEE symposium on visual analytics science and technology*, pp. 13–22
19. Parikh RB, Kakad M, Bates DW (2016) Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA* 315(7):651–652
20. Kumar G (2012) Machine learning for soccer analytics. *KuLeuven, Leuven*, pp 1–99
21. Brooks J, Kerr M, Guttag J (2016) Using machine learning to draw interfaces from pass location data in soccer. *Stat Anal Data Min ASA Data Sci J*. 9:338–349
22. Bedely RT, Benbow M, Iyer LS (2016) An analytic-based strategy for college sportsticket sales and management. In: *Twenty-second Americas conference on information systems, San Diego, 2016*, pp. 1–5
23. Pottala M (2018) Artificial intelligence in Sports. Thesis, Cenria University of Applied Sciences International Business, pp. 1–37
24. Beck H (2013) Advanced statistics added to revamped NBA site'. *New York Times*. [http://www.nytimes.com/2013/02/15/sports/basketball/nbas-site-to-feature-updated-statistics-database.html?\\_r=2&](http://www.nytimes.com/2013/02/15/sports/basketball/nbas-site-to-feature-updated-statistics-database.html?_r=2&). Accessed 29 Jul 2013
25. Furnas A, Lezra G (2012) Make way for the soccer geeks. *The Atlantic*. <http://www.theatlantic.com/technology/archive/2012/08/makeway-for-the-soccer-geeks/261634/>. Accessed 29 Jul 2013
26. Sellitto C, Hawking P (2015) Enterprise systems and data analytics: a fantasy football case study. *Int J Enterp Inf Syst* 11(3):1–12
27. Troilo M, Bouchet A, Urban TL, Sutton WA (2016) Perception, reality and the adoption of business analytics: evidence from North American professional sport organizations. *Omega*. 59(Part A):72–83
28. Morgulev E, Azar OH, Lidor R (2018) Sports analytics and the big-data era. *Int J Data Sci Anal* 5(4):213–222
29. Hutchins B (2015) Tales of the digital sublime: tracing the relationship between big data and professional sport. *Int J Res New Media Technol* 22(5):494–509
30. Fried G, Mumcu C (2016) Sport analytics: a data-driven approach to sport business and management. Taylor & Francis, London, pp 1–256
31. Barros CP, Leach S (2007) Performance evaluation of the English Premier Football League with data envelopment analysis. *Appl Econ* 38(12):1449–1458
32. Chiappori P-A, Steven L, Timothy G (2002) Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. *Am Econ Rev* 92(4):1138–1151
33. Rein R, Memmert D (2016) Big data and tactical analysis in elite soccer: future challenges and opportunities for sport science. *Springer Plus* 5:1410
34. Tulabandhula T, Rudin C (2014) Tire changes, fresh air, and yellow flags: challenges in predictive analytics for professional racing. *Big Data* 2(2):1–16
35. Bouchet A, Doellman TW, Troilo M, Walkup BR (2017) Pre-empting the competition: how do shareholders view sponsorships in the sport apparel industry? *J Sport Manag* 31(3):275–287
36. Woodie A (2018) Go fast and win: the big data analytics of F1 racing, pp 1–2. <https://www.datanami.com/2018/04/19/go-fast-and-win-the-big-data-analytics-of-f1-racing/>
37. Lohr S (2012) The age of big data, big data's impact in world. *The New York Times*, New York, pp 1–5



38. Travassos B, Davids K, Araujo D, Esteves TP (2013) Performance analysis in team sports: advances from an ecological dynamics approach. *Int J Perform Anal Sport* 13(1):83–95
39. Baerg A (2016) Big data sports, and the digital divide: theorizing how athletes might respond to big data monitoring. *J Sport Soc Issues* 41(1):3–20
40. Mcgarry T (2009) Applied and theoretical perspectives of performance analysis in sport: scientific issues and challenges. *Int J Perform Anal Sport* 9(1):128–140
41. Curtis B (2015) Moneyball II: Charles Barkley, the sports media, and the second statistical war. Grantland.com. Retrieved from <http://grantland.com/the-triangle/moneyball-advanced-statistics-charles-barkley-sports-media-daryl-morey-al-leiter-rob-neyernba-mlb-nfl-nhl/>. Accessed 13 May 2019
42. Torre PS, Haberstoh T (2014) New biometric tests invade the NBA. Espn. com. [http://espn.go.com/nba/story/\\_/id/11629773/newnba-biometric-testing-less-michael-lewis-more-george-orwell](http://espn.go.com/nba/story/_/id/11629773/newnba-biometric-testing-less-michael-lewis-more-george-orwell). Accessed 4 Jan 2016
43. Evans JR (2015) Modern analytics and the future of quality and performance excellence. *Qual Manag J* 22(4):6–17
44. Michael K, Miller KW (2013) Big data: new opportunities: and new challenges. *Computer* 46:22–24
45. Rose N (1999) Powers of freedom: reframing political thought. Cambridge Press, Cambridge
46. Rauter S, Fister I, Fister I Jr (2015) How to deal with sports activity datasets for data mining and analysis: some tips and future challenges. *Int J Adv Pervasive Ubiquitous Comput* 7(2):27–37

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.