# Importance of Data Analysis In The Field of Physical Education and Sports

Kishore Mukhopadhyay

Associate Professor In Physical Education
Union Christian Training College, Berhampore, Murshidabad, West Bengal, India

**Abstract**

Nowadays, sport represents the most important marginal thing in the world. For instance, if you turn the TV on, broadcasts of various sport competitions, interviews, news, and even talk shows about sport, are presented every day. Magazines, newspapers, health organizations, universities are promoting a healthy lifestyle and the benefits of practicing in sports. Approximately 15 years ago, athletes and trainers were not aware about the opportunities of using their collected data, since computers were not as available as they are nowadays. The situation has now been slightly different, although data obtained during sports training sessions are produced by a multitude of athletes around the world daily, these data are typically hidden or kept secret by trainers or the athletes themselves.The present study described the various aspects of data and their collection, preservation and analyzing for the betterment of sports and the individual athlete. It describes the potential for paradigm shifts in training, monitoring, rehabilitation, talent acquisition and even urban planning that data mining can provide. Furthermore, it highlights some risk concerns pertaining to big data in sport such as user privacy, data accuracy, interpretation of information and athlete autonomy.

**Key word :** Data, Big Data, Tools, Analytic Technique and Sports

**Introduction**

"Information is power. But like all power, there are those who want to keep it for themselves." Technology in sports today is a fast-growing field. The desire to move faster, recover quicker and last longer, have led to the increased integration of technology into sports. The prevalent fields of today's research and high tech products are textiles, material science and electronics. Through better nutrition and training, the athletes of today are becoming faster and stronger. Old records are constantly being broken, and new ones set. While the vast majority of these achievements are likely due to the atheletes themselves, improvements in sports technology have also played a notable role (1).

The increased popularity of smartphones and watches allows athletes to train smarter. More and more cyclists today use a different sports application, that run on mobile devices, such as Strava, Endomondo, Garmin Connect and many others. The primary tasks of such applications are tracking and analyzing their workouts. These tracked workouts can also be exported as XML files and analyzed later. In other words, they offer an immense of different tasks of data analysis and data mining. Some ideas about using these methods were discussed Iztok et.al (2). The mentioned data should be also used for planning the sport training sessions (3). In line with this, I have made some literature reviews about these research topic and realized that there is a lack of researches tackling the data analysis and data mining of sport activities created by sport trackers (4). This was the primary fact, for deciding an analyst to collect training data of different sports.

The use of technology in the analysis of sports performance has become a usual and necessary practice among professional teams. Monitoring the athlete's physical activity allows clubs to know their current physical condition and their playing style, as well as also know the players' fatigue or the load they are subject to during training sessions or matches. Thanks to all kinds of sensors and cameras we have real-time information about our player's position, speed and heart rate data that can also be

processed to analyze trends or possible improvements in the game. Two key questions arise at the moment of implanting this kind of innovative solutions: on the one hand, what kind of technology allows better monitoring our players in terms of efficiency; on the other hand, how to process all this data in a useful manner and being efficient with the resources at hand.

Outside the professional world, we are used to seeing all kinds of technological solutions created to monitor our advances. *Runners* can choose between carrying the mobile stuck to the arm to know the distance covered, wearing the latest watch or *smartband* with a GPS or wearing a tape around the chest with the same purpose. Gymnasts count the repetitions in each machine used or the count of their push-ups and crunches with different *apps*. However, monitoring elite athletes are where a real technological revolution has taken place. Since FIFA approved in 2015 the use of portable systems for monitoring performance during the matches, the public has been getting used to seeing these characteristic GPS vests under the players' shirts. Inside and outside the pitch, state-of-the-art technologies are used to know their movements, and the trend is only growing. Different future analyses of the *Real Time Location Systems in Sports* (RTLS) demonstrate that this is a growing market and it will generate countless business opportunities over the next few years. The comparison is so impressive as easy to understand: a market with a business volume of 280 million dollars in 2018 will turn to 2050 million in 2024. A substantial economic growth that will bring about a more than probable innovation in available technologies (5).

An overview of how team sports are recognized is vital in better understanding the different parameters that should be accounted for to extract meaning from data. They are dynamic systems of interaction, where individual and collective patterns of behaviour emerge from converging multiple constraints on the players.

Games falling under invasion games category comprise of two teams that compete for possession of a ball in a constrained playing area. Each team has the objective of scoring by putting the ball into the opposition's goal and also of defending its own goal (6) . The team that scores the greatest number of goals by the end of the game is the winner. Examples include American football, basketball, soccer, hockey, etc.The data from invasion games, mostly consists of ball and player trajectories throughout the game. It also includes events such as goals, assists, shots, etc. at any given time. Advanced object tracking systems produce spatio-temporal traces of these player trajectories with high definition and high frequency (7).

**Background**

In the last two decades, there has been a gradual change in how we refer to the observation, recording and analysis of performance in sport.

This Site explores some of the changes that have occurred. We tend to hear and read less about notational analysis now and talk more about **analytics**. This indicates an important change in the community of practice that analyses performance in sport. Bill Gerard (8) has provided an overview of these changes.

In 2005, the Journal of Quantitative Analysis in Sports appeared "as the first academic journal dedicated to statistical analysis in sports" (9). Jim Albert (10) noted in an editorial in the Journal "In 2005, there were 32 manuscripts submitted, and submissions have steadily increased with 106 manuscripts submitted in 2011". The Journal is an official journal of the American Statistical Association. The first volume of the Journal of Sports Analytics appeared in 2015.

The first MIT Sloan Sports Analytics Conference was held in 2007. The goal of the annual conference is to provide a forum for industry professionals and students to discuss the role of sport analytics in the global sports industry. There were 175 attendeesat the inaugural conference convened by Daryl Morey and hosted in classrooms on the MIT campus. In 2013 there were 2,700 attendees and it took place in the Boston Convention and Exhibition Centre.

An MIT report of the 2014 Conference asserted:

The sports analytics revolution has happened. The question is no longer whether to use analytics to measure and monitor team performance and value. The competition from here on out—on the court, on the field, or in the front office—is a matter of how those analytics are used.

There was a Sports Analytics Innovation Summit in London in March, 2014. Since then there has been an explosion of events in this growing field of study (11).

**Data**

Data are characteristics or information, usually numerical, that are collected through observation(12).

In a more technical sense, data are a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum (singular of data) is a single value of a single variable (13). Although the terms "data" and "information" are often used interchangeably, these terms have distinct meanings. In some popular publications, data are sometimes said to be transformed into information when they are viewed in context or in post-analysis(14). In academic treatments of the subject, however, data are simply units of information. Data are employed in scientific research, businesses management (e.g., sales data, revenue, profits, stock price), finance, governance (e.g., crime rates, unemployment rates, literacy rates), and in virtually every other form of human organizational activity (e.g., censuses of the number of homeless people by non-profit organizations).

Data are measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools. Data as a general concept refers to the fact that some existing information or knowledge is *represented* or *coded* in some form suitable for better usage or processing. *Raw data* ("unprocessed data") is a collection of numbers or characters before it has been "cleaned" and corrected by researchers. Raw data needs to be corrected to remove outliers or obvious instrument or data entry errors (e.g., a thermometer reading from an outdoor Arctic location recording a tropical temperature). Data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next stage. Field data is raw data that are collected in an uncontrolled "in situ" environment. Experimental data is data that are generated within the context of a scientific investigation by observation and recording.The data has been described as the new oil of the digital economy(15,16).

**Data documents**

Whenever data needs to be registered, data exists in the form of a data documents. Kinds of data documents include:

- data repository
- data study
- data set
- software
- data paper
- database
- data handbook
- data journal

Some of these data documents (data repositories, data studies, data sets and software) are indexed in Data Citation Indexes, while data papers are indexed in traditional bibliographic databases, e.g., Science Citation Index. (17).

**Types of Data In Sports**

The process of sports training is the most important task of the coach. The process consists of four stages:

(1) Planning,

(2) Implementation,

(3) Monitoring and

(4) Evaluation.

Although all four phases are equally important, the majority of coaches have been focused primarily on the first two phases in the past. With the rapid development of new technologies, which allows tracking of an athlete's indicator load during the training process, the third and fourth stages in this process (i.e., monitoring and evaluation) became manageable. However,

both stages cannot be considered without data that can arise from various sources. In this study, we focus on the four data sources, as follows (17) :

- **Sports tracker data:** Data from sport trackers (18) and smart watches are essential in endurance sports. Loosely speaking, sports trackers appeared with the advent of smartphones. Most of the smartphones also contain a GPS receiver that is capable of recording the current position of a person. In line with this, a lot of applications (19,20) devoted for tracking sport activities were developed for smartphones. In 2017, there is a bunch of such applications, where the more popular ones are: Strava, Endomondo, Runkeeper, Runtastic. Additionally, there are also special sports watches (e.g., Garmin, Suunto, Polar) that also consists of a GPS receiver. The main advance of these sports watches over the smartphones is their convenience. Watches can be mounted easily on a hand and even a bike, and are easy to transport. They can also be used in the water. For that reason, professional athletes prefer to use watches instead of taking a smartphone on training. Roughly speaking, watches and smartphones, data encompass GPS data, data from heart rate monitors, cadence, power-meters and even some data that is available via an Internet connection (e.g. current weather). Data from GPS receivers show an exact map where athletes are performing a sports training session, while information like total ascent, descent are used for later analysis. The average and maximal speed can also be obtained from GPS data. Other sensors (21, 22), like power-meters, heart rate monitors, cadence sensors, also produce more parameters about the performed training that can be used as load indicators, and are important for performance analysis (20 ).

- **Laboratory measured performance data :** Monitoring and evaluation of sports training are very important parts for an athlete in the training process. Usually, measurements are pillars of an athlete's performance. Hence, they are conducted on various time intervals in dynamic conditions. For example, during the winter, cyclists conduct sports training on cycling treadmills. These treadmills are always situated in laboratory conditions (e.g., a gym), while athletes can set some parameters on the treadmill. Parameters are intended to change the difficulty of a workout, simulate mountain conditions, etc. Treadmills are very special, because trainers can see an easy comparison with previous sessions of an athlete on the same settings.

- **Laboratory measured health data :** Health data are very important because, on the one hand, they contribute by evaluating the athlete's current form while, on the other hand, they can be an indicator of several health symptoms. An example of the kind of health data involved: An athlete's anthropometry, body composition, blood parameters etc (17).

- **Sport results data:** At the end, the most important evaluation process of the sports training is the sport results of the athletes achieved during the season. Nowadays, most of the serious races can be tracked live online (e.g. IRONMAN tracker), while all race results are usually published online within the competition day. Analyzing the athlete's results is of the utmost importance, since someone can get a lot of insight about his/her performance when comparing these with the results of the opponents (17).

**Data analytics**

Analytics is a rapidly evolving technology that involves the use of data and statistical analysis to make well-thought and researched decisions. A technique known as statistical methodology plays a major role in analytics, which extracts relevant information about measurements and their relationships. Analytic methods use data to draw conclusions and determine the best course of action. Analytics are cutting across just about every sector, including health care, business, law, security, sports, social media, politics etc (23).

- Data analysis means a process of cleaning, transforming and modeling data to discover useful information for business decision-making
- Types of Data Analysis are Text, Statistical, Diagnostic, Predictive, Prescriptive Analysis
- Data Analysis consists of Data Requirement Gathering, Data Collection, Data Cleaning,
- Data Analysis, Data Interpretation, Data Visualization (24).

**Sports analytics**

Data + statistical analysis = sports analytics. That's the gist of it in simple terms.

It's about using the hard numbers combined with a complex data analysis methodology, not 'gut feelings'. Let's put this in Moneyball terms. The 'old school' way of recruiting ball players looks at speed, strength, mental toughness, how good the swing and release 'looks' (25).

Sports analytics refers to the use of data and advanced statistics to measure performance and make informed decisions, in order to gain a competitive sports advantage. Sports analytics, when properly applied, can yield tremendous competitive advantages to a team or an individual.

As technology continues to advance, data collection becomes more complex and the amount of available data grows, so does our ability to measure and improve performance. This data expansion has led to the development of advanced statistics and performance optimization.

There are two key aspects of sports analytics — on-field and off-field analytics.

On-field analytics deals with improving the on-field performance of teams and players. It digs deep into aspects such as game tactics and player fitness. Off-field analytics deals with the business side of sports. Off-field analytics focuses on helping a sport organization or body surface patterns and insights through data that would help increase ticket and merchandise sales, improve fan engagement, etc. Off-field analytics essentially use  the data to help rights holders take decisions that would lead to higher growth and increased profitability (26).

**Importance of sport analytics**

- **Improves players' and game performance**

The use of sports analytics helps teams make useful decisions to enhance team performance. Decisions involving player acquisition, fielding the best possible rotation on a given game, which areas the team must focus on to win games are a few of the processes analytics plays an important role in.

- **Enhances organization's business performance**

Another major category of analytics in sports revolves around business-oriented applications. These include the use of analytics in ticket price**.** Ticket pricing is a reliable method for teams to optimize revenue through analytics.

A commonly used ticket pricing approach is to offer dynamic pricing, in which variable prices change throughout the season based on factors such as how well the team is doing and the success of the team's opponent in the league standings.

Fans' loyalty analysis, promotions, social media, and fans' engagements are ways social media analytics are leading the practice in measuring fans' sentiments and engagements.

Also, analyzing social media data and interacting with fans on social platforms help teams hear and respond to the voice of the customers, thereby aiding their next line of action.

- **Analyzes player health and injury probability**

Lastly, sports analytics can be used in players health and injury analytics. That is, predicting the situations when players will be more vulnerable to injuries, so that time off or substitutions can be planned in advance, thereby enhancing team effectiveness.

In conclusion, all areas of sports organizations are leveraging analytics in sports to help teams gain a competitive edge .

- **Talent acquisition and Identification**

Coaches and scientists generally accept that elite endurance athletes have been through many years of training to enter and ride at the top level of the sport. However, the progression of success from junior to senior level is still unclear. For identification of talents of a beginner and possibilities of sports career the data analysis is able to show the right direction.


**Types of Analysis**

There are several **types of Data Analysis** techniques that exist based on business, Sports  and technology. However, the major types of data analysis are (27) :

- Text Analysis
- Statistical Analysis

- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

**Tools of data analysis**



Figure-1 Data analysis tools soure https://www.guru99.com/what-is-data-analysis.html

Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretations.The complete list of tools are presented in figure -1 ( (28).

**Analytics Technology Partners**

Foster a *close relationship between analytics staff and key technology partners* (video coordinators, IT, vendors, etc.). Analytics in all organizations depend on technology, and they do so especially in sports. Modern technologies such as video, location devices, and biometric sensors generate data in massive amounts, and require substantial computing horsepower to analyze. An analytics staff can't work effectively without close coordination with IT and sports technology staff and leadership.

A final frontier approach to business analytics is taking place at the Phoenix Suns, which has partnered with Verizon Wireless to learn about various activities of fans before and after Suns games. Since Verizon knows the locations of its mobile phone customers, it can tell organizations like the Suns (through a business unit called Precision Market Insights) where customers are at different times. Verizon was able to tell the Suns, for example:

• Where fans attending the team's games live (and hence where people live who don't attend games) in order to target team advertising;

• How many fans at games are from out of town (about 22% in one month studied);

• The attributes of fans who attend games (most likely between 25 and 54 in age, with household incomes of more than $50,000, and parents with children at home);

• How often game attendees also attend baseball spring training games in the Phoenix area (13%);

• Increases in traffic at a fast food chain within 24 hours of a game that are linked to a Suns promotion (8.4%) (29).

Of course, full advantage from player and team performance analytics would seem to come when all the coaches and players on a team embrace analytics and use them to enhance their performance. That appears to be the approach taken by the English Premier League soccer team West Ham United, whose manager, Sam Allardyce, has been an early adopter of analytics in that sport (see sidebar, "Analytical Leadership Profile—Sam Allardyce") (30).

West Ham has a meeting after games in which Allardyce and David Woodfine, the team's head of performance analysis, spends 15 minutes going over the numbers that characterize the team's performance during the game. The primary data source is the Prozone player and team performance data of which Allardyce was an early user at a previous club. After the team

overview, individual profiles of players and their specific metrics are reviewed and discussed. Evidence-based coaching points are presented to players, including analysis of GPS captured training data, wearable device data, such as sleep and heart monitors, video clips and Prozone numbers (31).

Table-1. Table stake analysis and frontier analysis for sports (29)

| Table Stakes Analytics | Frontier Analytics |
|---|---|
| Use of external data sources | Analytics on video data |
| Descriptive analytics on players | Analytics on locational/biometric data |
| Optimal lineup analytics (basketball) | Open data analysis by fans |
| Player scoring for draft analysis | Engaging players in analytics |
| Player salary optimization | Gathering and using proprietary data |
| Simulation of games | |
| Analysis of game tactics | |

**Big Data In Sports**

Sports produce large amounts of data related to the players, team performance, and audience. With Big Data, it is possible to analyze large amounts of this data and utilize it. Since there is a higher demand for the sports statistics, big data is the ideal technology for sports. Big Data is creating more excitement in the field of sports. It is taking sports to a higher echelon. Michael Lewis's book "Moneyball"(2013) introduced the concept big data in sports long before it was actually used. It showed that data can be used to make important decisions.

**Applications of big data in sports**

Big data are advancing sports by taking it to the next level. It has firmly established itself in various areas.Some of the applications of big data in sports are mentioned below:

*Prediction of match results*: Big Data could help to interpret better match results. The founder of Advanced NFL Stats website, Brian Burke (2015) states that using big data coaches and players can predict the results of the match and take proper decisions.

*Analyzing team performance*: Big data, analyzes the various factors of the players that give an insight to the players to analyze their own strengths and weaknesses as well as that of the competitors thereby improving the team's performance.

*Understanding the viewer's preferences:* Viewers play one of the most significant roles in sports and hence, concentrating on the viewer's preferences is a priority. Big data sustains the viewer's interest in the game by providing them real-time statistics, more data sharing capacities and so on.

*Improved decisions*: Umpires have a very important decision to make during the matches and one wrong decision by the umpire can change the entire game. Big data is very convenient to use during such times. It provides real-time data to the umpires which help them in making correct decisions.

*Giving out statistics*: Deeper understanding of the game is possible than before because of the enormous amounts of information available today. Big Data products such as IBM's Slam Tracker provide point-by-point analytics of tennis matches from every stroke and point (32).

**Conclusion**

This paper describes briefly creating and exploring the archiving of valuable sports data that are generated during the process of sport training. Some years ago, nobody had given any special attention to these data. With the rise of data mining and artificial intelligence methods, data are also becoming very attractive for research purposes (e.g., performance analysis of realizing sports training sessions). In the Present study, after reviewing various resources the following conclusion may be drawn ;

- Data is required for talent identification in sports
- Proper coaching and preparation of an athlete
- Sports performance tracking, especially in endurance activities using GPR

- Various laboratory measurements and monitoring
- Performance analysis
- Correction of sports skills
- Development of strategy
- Analysis results
- Proper officiating by giving correct decision
- Health and fitness development
- Injury prevention and rehabilitation
- Promote business and customer satisfaction
- Individual rating  of sports person
- Game analysis
- Individualised equipment and technique
- Sports statistics
- Media analysis

**References**

1. Exploratorium, 2012, "The Science of Cycling: Aerodynamics & Drafting." Available at http://www. exploratorium.edu/cycling/aerodynamics1.html (15 October 2012).

2. Iztok Jr. Fister, Dušan Fister, Iztok Fister, and Simon Fong, (2013) Data mining in sporting activities created by sports trackers. In Computational and Business Intelligence (ISCBI), 2013 International Symposium on, pages 88–91. IEEE, 2013.

3. Iztok Fister, Samo Rauter, Xin-She Yang, Karin Ljubiˇc, and Iztok Jr. Fister, (2015)  Planning the sports training sessions with the bat algorithm. Neurocomputing, 149:993–1002.

4. Samo Rauter, Mojca Doupona Topiˇc, and Iztok Jr. Fister, (2014), Mobile sport applications can make our cycling more sociable. In Youth sport: abstract book, Ljubljana: University of Ljubljana, Faculty of Sport, page 41.

5. https://barcainnovationhub.com/monitoring-technologies-for-sports-analysis/

6. Pearson, P., Webb, P. (2008) An Integrated Approach to Teaching Games for Understanding (TGFU). Available: http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1053&context=edupapers
   [Accessed: 16- Apr- 2017].

7. Batra Sarthak , (2017),The Role of Data Analytics in Modern Day Sports, https://www.researchgate.net/publication/330578956_The_Role_of_Data_Analytics_in_Modern_Day_Sports.

8. Bill Gerrard (2015). Analytics, Technology and High Performance Sport. In N. Schulenkorf & S. Frawley (Eds.) Critical Issues in Global Sport Management. Routledge: London.

9. Benjamin Alamar (2013). Sports Analytics: A Guide for Coaches, Managers and Other Decision Makers. Columbia University Press: New York.

10. Jim Albert & Stephanie Kovalchik (2017). The Do's and Don'ts of Sports Metrics: The Tennis ATP Leaderboard. Chance, 30(1), 26-34.

11. https://sites.google.com/site/ucsportinformaticsandanalytics/analytics

12. OECD Glossary of Statistical Terms. OECD.( 2008). p. 119. ISBN 978-92-64-025561

13. "Statistical Language - What are Data?". Australian Bureau of Statistics. 2013-07-13. Archived from the original on 2019-04-19. Retrieved 2020-03-09.

14.  "Data is the new oil". July 16, 2018. Archived from the original on 2018-07-16.

15. "Data vs Information - Difference and Comparison, Diffen". www.diffen.com. Retrieved 2018-12-11.

16. Yonego, Joris Toonders (July 23, 2014). "Data Is the New Oil of the Digital Economy" – via www.wired.com

17. Fister et. al. (2016), the importance of monitoring and maintaining data in sports Training process, Proceedings of the 8th Conference for youth sport https://www.researchgate.net/publication/315891894

18. Fister Jr, I., Fister, I., Fister, D., & Fong, S. (2013). Data mining in sporting activities created by sports trackers. In Computational and Business Intelligence (ISCBI), 2013 International Symposium on, 88–91.

19. Ferrari, L., & Mamei M. (2013). "Identifying and understanding urban sport areas using nokia sports tracker." Pervasive and Mobile Computing 9(5), 616-628.

20. Mutijarsa, K., Ichwan, M., & Utami, D. B. (2016). Heart rate prediction based on cycling cadence using feedforward neural network. In Computer, Control, Informatics and its Applications (IC3INA), 2016 International Conference on (pp. 72-76). IEEE

21. Novatchkov, H., & Baca, A. (2013). Artificial intelligence in sports on the example of weighttraining. Journal of sports science & medicine, 12(1), 27

22. Novatchkov, H., & Baca, A. (2012). Machine learning methods for the automatic evaluation of exercises on sensor-equipped weight training machines. Procedia Engineering, 34, 562-567.

23. https://hackernoon.com/why-is-sports-analytics-so-important-ec373yt5

24. https://en.wikipedia.org/wiki/Data_analysis

25. https://www.agilesportsanalytics.com/what-is-sports-analytics/

26. Ray, Sugato (June 22, 2017). "The Evolution and Future of Analytics in Sport". Proem Sports, Sports Analytics , Singapore & India. Retrieved August 5, 2018.

27. https://www.agilesportsanalytics.com/what-is-sports-analytics/

28. https://www.guru99.com/what-is-data-analysis.html

29. Davenport Thomas H. (2014), Analytics in Sports: The New Science of Winning, International Institute for Analitics, Copyright ©2014 International Institute for Analytics.

30. Information about Allardyce comes from regular in-person meetings during 2012-13 and 2013-14 season, an interview for this study by Al Sim (Feb 2014), and the following articles: Sam Allardyce – Barclays Premier League Profile http://espnfc.com/manager/_/id/29/sam-allardyce?cc=5739

31. Matt Lawless, "Big Sam is 100 not out.. and we reveal how life at West Ham is a numbers game under Allardyce," Daily Mail, Sept. 24, 2013, http://www.dailymail.co.uk/sport/football/article-2430369/Sam-Allardyce-100th-West-Ham-match-Sportsmail-reveals-life-Hammers-numbers-game-Big-Sam.html

32. BIG DATA Made Simple, 2016 (http://bigdatamadesimple.com/big-data-an-ultimate-weapon-in the-field-ofsports/)

.