

The Data Analytics Controversy in European Soccer

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Nicholas A Martin

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

STS Research Paper

The Data Analytics Controversy in European Soccer

Introduction:

In today's marketplace the only way to stay competitive is increasing market value. In this case market value is defined as the overall net worth of the organization. This can be found from aircraft design to professional sports management. It is imperative for companies to find ways to increase their market values. Digital methods allow companies to streamline the process of finding ways to increase market values. They offer diverse applications in problems of optimization. As digital methods grow their application grows within the market. Digital methods are even making appearances in the sporting world, most notably in the sport of European Soccer. Teams have incorporated digital methods to gain advantages over their opponents. Soccer teams, in order to stay competitive, employ digital methods such as data analytics to find solutions that will help them win on and off the pitch. Organizations around the sport have also begun to employ digital methods to shape the sport. The growing use of digital methods is causing the sport to change prompting controversy on its application.

Sociotechnical Research:

In professional soccer winning earns teams millions of dollars. To succeed teams must utilize every advantage they can find. In soccer, observational data has been collected since the 1950s; data analysis was introduced in 1999. Charles Reep was the first to lead observational data collection. He was hired and helped a struggling team avoid, "relegation after winning 13 out of 14 matches and doubling their goals scored" (Luzum & Model, 2021). Teams witnessing the success began using their own data collectors. As data collection needs increased, collectors

formed businesses. These businesses paved the way for modern analytics. The modern analytics revolution began in the early 2000's when OPTA began collecting individual player statistics using digital methods. Digital methods provide diverse data sets that could not be tracked without computers. Data analysts have the job of interpreting the data and producing it into useful information. What started out as observational player analysis has expanded exponentially. Today analytics are used not only in player analysis but player potential, team analysis and even refereeing decisions. Some, however, contend that analytics has gone too far, and that many human judgments cannot be automated.

Methods:

Research for this paper included a variety of different sources and documentation. The goal of this paper is not to argue for or against the use of digital methods. The aim of the research is to show how digital methods have changed the game of soccer, while also revealing the differing options from participants within the sport. Most of the information gathered was taken from various research papers on the subject of data analytics and how it has changed the game. These were chosen in order to build an accurate picture of what changes there have been over time. The primary example of this was the sports incorporation of digital methods in the player recruitment process. Research also includes analysis of individual participant groups within the sport. Each participant group was chosen because of their significant role within the sport. The information gathered reveals each of the group's identity, objectives, and opinions on the increasing usage of digital methods. These participants range from the fans to analytics firms collecting the data. Each of these groups had various ways of releasing information and

communicating their options on digital methods. Fans and fan groups were much less organized than companies and associations being affected by digital methods. Most of the research gathered from fan groups came from social media platforms and interviews. Whereas data firm's and other more organized group's research came from press releases, new reports, and articles. Despite, inconsistencies within the sources of research between the participant groups all their objectives and opinions were accurately found.

Analysis and Research:

Researchers have studied analytics in soccer since its inception, and the research done within the sport is all encompassing. Originally the research was limited because many believed that soccer could not incorporate statistics and digital methods. This is because soccer is a low scoring game with few statistical markers (Roura, 2019). It was also believed that digital methods could not predict decisive moments because there are so few. Soccer is a sport with eleven players per team playing ninety minutes. The only previously reliable team data points for a match were goals, assists, and clean sheets. Clean sheets are when a team does not concede a goal. As digital methods have modernized many new individualized statistics have been created. These individualized statistics have direct impacts on matches. With these statistics, teams can easily gather information about their own players and opponents as well as overall tactics (Roura, 2019). These stats are allowing teams to gain competitive advantages. Overall, the methods have helped level the playing field between teams, while bringing new life into the game (Roura, 2019). Going beyond the scope of individual teams, other significant participants within the sport have incorporated digital methods. Governing bodies, broadcasting agencies,

leagues and even referees are using these digital methods changing the game. This is a stark contrast to the original perception of analytics in the role of soccer. All of this would not have been possible without the modernization of digital methods propelled by analytics organizations collecting the data and marketing it.

Analytics Organizations:

Digital method firms market their services to soccer leagues and clubs. Oracle presents “Research shows that fans are interested in the integration of match data and analysis.” (Austin 2021). Match data promotes more understanding of the game and increases the number of fans watching. Oracle also promises that teaming with the Premier League will increase excitement around the sport (Austin 2021). By offering leagues more statistics the firms increase revenue. The firm’s market different data packages with different amounts and sets of data similarly to how Apple markets different versions of the iPad. Through supporting leagues, they can market to the clubs within the leagues as well. Data firms collect stats about individual players simultaneously offering them to the league and clubs.

Digital methods organizations have begun to appear within individual club structures as well. As clubs seek to keep their competitive advantages, they employ workers from organizations such as Oracle as contractors. This has allowed clubs to build their own analytics teams. To stay competitive these companies, advocate for more data science within the sport. Data firms look to offer data packages better than the data packages clubs and leagues can produce themselves. This makes for an interesting relationship between the companies and clubs. While offering data to clubs and leagues they also look to out compete them in order so clubs

and leagues will still rely on their products. Firms accomplish this by looking for new ways to collect data using more powerful digital methods, along with searching for more key match metrics that determine the outcomes of games.

The most recent breakthrough in key match metrics is the idea of the expected goal. Since its introduction in 2012 by OPTA the “expected goal” has become the most widespread metric in soccer (Whitmore, 2022). This metric calculates the quality of a scoring chance based off the likelihood of it resulting in a goal (Whitmore, 2022). The metric only calculates the chances of a specific shot attempt on goal. Expected goals do not include missed opportunities that could have led to a shot on goal. Expected goals use the position of the shot, angle, and defenders positioning to calculate the odds of a goal using thousands of previous data points from previous matches. OPTA have directly given teams a chance of understanding how to best score matches. Data firms have also produced methods that go beyond the teams and are being incorporated by referees.

The Referees Association:

The main point of contention with analytics is Video Assisted Refereeing (VAR). Tamir and Bar-eli (2021) observe that VAR is an important addition to the game even despite the criticism from various groups. The Referees Association favors analytics. Referees administer the tool, and according to the association: “Referees are happy to trust in the advice of a video assistant” (Gardner 2019). Referees who have access to the tool have the ability to take a second look at events that occur in the game. VAR also allows a second set of eyes to watch plays as well.

The support for VAR was not always high. At the beginning of VAR's release referees were concerned about its usage. Before late 2010's VAR had never been used in the sport. The main concern with incorporating the technology was reducing the fluidity of the game. This concern was based off the slow nature of VAR in other sports. In sports such as American football and baseball referees can spend several minutes reviewing events. These stoppages would go against the non-stop play that distinguishes soccer, and is a huge draw for the fans. In soccer the clock never stops and additional time is added to ends of the halves to make up for any breaks in play.

Despite the stoppages, the hope for VAR was having enhanced officiation. The Referees Association believes that VAR, "will improve the accuracy of Key Match Incidents to 87 percent" while also protecting referees (Premier League, 2019). The accuracy is compared to the current accuracy of refereeing decisions without the use of VAR. As VAR has been incorporated, overall accuracy has increased. Referees are also becoming more comfortable with using the tool and doing so quickly. Each season VAR decisions have increased in speed. This is because referees are becoming more comfortable with relying on the referees off the field rather than running to monitor to review a play.

Despite all the benefits of VAR games are still not refereed perfectly, because all plays are still judged differently by various referees. VAR is still administered differently across the leagues as well so decisions and rules vary in different leagues. This is all to be expected because VAR was never intended to make officiating perfect. It is designed to assist and produce more accurate decisions. Even in the new age of VAR, referees still remain the main target of abuse during soccer matches. VAR offers protection for on-the-pitch referees from making crucial mistakes, while increasing the total amount of referees. This all contributes to a fairer match with

less bias. As VAR increasingly protects referees it invites more critics of the technology itself. This is prompting fan groups and some players speaking out against its usage claiming VAR is hurting soccer.

Fans Against VAR:

The other side of the controversy includes many fan groups and players standing against VAR. Fans Against VAR is a social media forum. The two main groups use Twitter and Facebook. Members are petitioning the sport's governing bodies to end digital methods in officiating, claiming: "VAR has stolen the heart and soul from football" (Willmott, 2019). This claim comes from the belief that the VAR destroys the fluidity of the sport. VAR allows referees to retract or make calls using video in real time, contrary to before VAR's incorporation. Using the video can stop a match for minutes at a time. This stoppage takes away from non-stop play. Some members within the groups have vowed not to participate within the sport if VAR is not abolished. Fans Against VAR also advocates that the VAR process is too precise forcing referees to hesitate in making decisions.

Clubs:

Clubs have started to incorporate digital methods into strategy on and off the pitch. Clubs like Chelsea use analytics to offer fans services by bringing extra detail and match analysis (Cech, 2020). Chelsea's goal is to bring fans closer to the game, and to encourage greater participation. Participation allows support to become collectivized. Encouraging more

participation leads to communities and cultures being formed around the club. Clubs investing into their fans hope to generate greater support, revenue, and better performances. By generating better fan support home matches become increasingly harder for the away teams. Clubs with massive home stadiums and supportive fans help generate a more competitive atmosphere for the home team. The revenue gained can be put towards player recruitment and training.

Data provided by digital methods can unearth hidden gem players that can be purchased for low prices and sold for high prices. Analytics guides teams in finding and attracting the best talent (Kalenderoglu, 2019). Data also provides the ability for clubs to buy these “hidden” and undervalued players to better their chances of winning. Teams operating this way are often called “money ball teams” inspired by the movie *Money Ball*. This method of operation has been steadily increasing across Europe. This has created massive signing races for specific youth prospects and under valued talents. Clubs use digital methods to judge potential players and current players. These methods include on the pitch performance and training metrics. Digital methods allow clubs to choose the best players for matches. The methods are based in how players have performed in matches and training. As clubs incorporate more uses for digital methods, most have started to build their own data departments. Clubs are trying to become more self-sufficient from the data firms. Often times clubs hire independent contractors from organizations and OPTA to build these departments.

Broadcasters:

Broadcasting agencies have incorporated digital methods to provide extensive coverage of matches. These agencies want to bring fans more in depth analysis to increase participation.

Just like clubs and leagues the broadcasters want more participation to increase revenue. Broadcasters accomplish this by offering key match stats and key moments allow fans to understand the game better without having to watch the full game. Broadcasters also use digital methods to provide information about games in shorter amounts of time. They can turn ninety-minute matches into thirty minute or less programs for fans making matches more accessible. They do this by using digital methods to analyze only the most impactful moments of the game such as goals, important chances, and fouls that affect the match. Broadcasters also incorporate stats within match analysis allows the agencies to provide accurate information and to support their opinions of the match. All of these methods are similar to FIFA's original intent with using data analytics back during the 2014 world cup. Broadcasters also work together to provide information about rule changes coming from the governing bodies of the sport.

Governing Bodies:

Governing bodies of the sport have had to create laws protecting players from the consequences of digital methods. UEFA have had to identify how data applies to players and how collection groups can use the data. Data firms try to collect medical data and personal information to build accurate player profiles. Governing bodies have taken steps to protect personal and medical information from data agencies. Medical information has been deemed too personal for data collection agencies. Under governing bodies and league players have specific privacy rights. UEFA, the Union of European Football Associations, protects players from aggressive data collections stating, that they provide adequate security measures in compliance with data protection laws (UEFA, 2020). Most of the data protected is specific medical information both physically and mentally. Bodies have also begun to expand the resources

available for soccer players and coaches to protect themselves. They have also been working towards providing other necessary means of protection like therapy and personal finance resources. Younger players in their teens and youth players have further protections provided by UEFA and leagues. Other bodies such as FIFA, the Fédération Internationale de Football Association and the FA, the Football association have implemented similar standards. These organizations work to ensure that players are not being targeted and belittled by data collecting agencies.

Conclusion:

Data collection and its subsequent analysis has been used throughout history. Data has been used to make improvements to previous systems. Organizations are always trying to improve to stay competitive and produce more effectively. It is only recently that digital methods have been introduced as a viable method of data collection. Some argue that the digital methods are more efficient than human collection. The need for human interpretation will always be needed, however, as data collection increases. It is important to analyze the effects of digital methods in order to keep the integrity of the industries employing the methods. Managing collection and its usage within the sports industry protects players and fans. The research of digital methods is a way to keep integrity within the sporting world, whether that is protecting, players, fans or referees. Understanding the dynamics between those using and those being affected by digital methods is key to comprehending how digital methods should be employed.

References

- Austin. (2021, May 6). Premier League selects Oracle Cloud Infrastructure to power new Advanced Football Analytics. Oracle.
<https://www.oracle.com/news/announcement/premier-league-selects-oracle-cloud-infrastructure-2021-05-06/>.
- Bojanova, I. (2014, July-Aug.). IT Enhances Football at World Cup 2014. *IT Professional 16* (4): 12-17. IEEE Computer Society.
- Cech, P. (2020, July 8). Key statistics from Crystal Palace vs Chelsea: Official site: Chelsea Football Club. ChelseaFC. <https://www.chelseafc.com/en/news/2020/07/08/the-data-day--crystal-palace-v-chelsea---dominant-zouma>.
- Gardner, J. (2008, Nov. 19). RA Chair Has His Say on VAR. The Referees' Association.
- insideBIGDATA. (2019, August 9). How data analytics improve business decisions. insideBIGDATA. <https://insidebigdata.com/2019/08/09/how-data-analytics-improve-business-decisions/>.
- Kalenderoğlu, U. (1970, January 1). *Football player profiling using OPTA MATCH EVENT DATA: Hierarchical clustering*. DSpace@MEF.
<http://openaccess.mef.edu.tr/xmlui/handle/20.500.11779/1214>.
- League, P. (2019, May 12). 'var will help referees to make better decisions'. The Referees' Association - 'VAR will help referees to make better decisions'. <https://www.the-ra.org/news/var-will-help-referees-to-make-better-decisions>.
- Luzum, N., & Model, M. (2021). History and background. The Soccer Analytics Revolution. <https://sites.duke.edu/socceranalyticsrevolution/history-and-background/>.
- Petroski, H. (1985). From Slide Rule to Computer. In *To Engineer Is Human* (New York: St. Martins Press).
- Rainie, R., and Anderson, J. (2017). Humanity and human judgment are lost when data and predictive modeling become paramount. Pew Research Center.
- Gabriela Colon Roura Feb 3, 2019. "Stats in Soccer." *The Tartan*, 3 Feb. 2019, <https://thetartan.org/2019/2/4/sports/soccer-stats>.
- Tamir, I., & Bar-eli, M. (2021, Jan. 12). The Moral Gatekeeper: Soccer and Technology, the Case of Video Assistant Referee (VAR). *Frontiers in Psychology 11*.
- UEFA. Union of European Football Associations. Player Privacy Notice.
- Whitmore, J. (2022, March 14). *Evolving expected goals (XG)*. The Analyst. Retrieved April 15, 2022, from <https://theanalyst.com/na/2022/03/evolving-expected-goals-xg/>