Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 10, Issue 2; April-June 2023, pp. 72-77

http://www.krishisanskriti.org/Publication.html

The Impact of Sports Analytics on Player Selection, Team Performance and Fan Engagement

Yuvraj Krishan Adya

Pathways World School, Gurgaon
E-mail: adyayuvraj@gmail.com

Introduction

Sports analytics, which refers to the method of utilizing data and statistical analysis to inform decision-making in sports (Gómez, Lago-Peñas, & Pollard, 2018), has emerged as a transformative paradigm within the world of sports, furnishing teams with novel tools to scrutinize player performance, make informed decisions regarding player selection, and connect with fans (Kannan & Narasimhan, 2020). Technological innovation and data analytics have revolutionized the way sports teams approach decision-making, propelling sports analytics to the forefront of the sports industry which involves leveraging technology to amass data on player performance, team strategy, and other key factors that influence performance. While this methodology itself is not fairly recent, having been in practice in various forms since the 1960s, its sudden advancement with the increasingly large amounts of available data, technological tools, and statistical methods have made it far more accessible, and consequently influential, in the world of sports. For example, sports teams traditionally relied on the subjective opinions of coaches and scouts to assess player potential and identify top talent. However, the subjectivity of these methods often led to inconsistencies and errors in player evaluation and selection, resulting in sub-optimal team performance. This realization led to the need for more objective and data-driven approaches to player selection and performance evaluation which was primarily led by the rise of technology and the availability of data. This has led to the development of more sophisticated methods of analysis, such as machine learning and predictive modeling. In addition to these factors, the need for sports analytics could also be said to have emerged from the ever-increasing competition in the sports industry, which has prompted teams and organizations to seek out innovative and effective methods of gaining a competitive advantage over their rivals. These methods are being used to analyze vast amounts of data, including player performance statistics, game footage, and even social media activity.

While sports analytics has the potential to revolutionize the way that teams approach decision-making, several challenges must be overcome to fully realize its potential. The primary challenge lies in the complexity and volume of data generated by different sports, which requires sophisticated analytical techniques to extract meaningful insights. Another challenge is the lack of standardization in data collection and storage, making it difficult to compare and analyze data across different teams and leagues. Additionally, the success of sports analytics depends heavily on the quality and accuracy of data, which may be influenced by factors such as technology limitations, data biases, and data privacy concerns. Furthermore, the interpretation and application of analytical insights in decision-making require expertise and collaboration between analysts, coaches, and players, which may pose organizational and cultural challenges. Finally, there is a risk of overreliance on data-driven decision-making, which may neglect the human element of sports and lead to unforeseen consequences. These challenges highlight the need for continued research and development in sports analytics to overcome these obstacles and fully realize its potential in enhancing player performance, team success, and fan engagement.

Therefore, there remains a significant gap in existing research regarding its social and economic impact and what consequences it could project (Kannan & Narasimhan, 2020). Hence, this research aims to explore the impact of such sports analytics on the various social realities connected with its mathematical model, particularly player selection, team performance, and fan engagement and address this gap by exploring the impact of such sports analytics on the various social realities connected with its mathematical model, particularly player selection, team performance, and fan engagement. Therefore, the objectives of this study are to identify the benefits and limitations of sports analytics, to analyze the effectiveness of sports analytics in improving player selection and team performance, and to explore the impact of sports analytics on fan engagement.

Literature Review

As sports become increasingly data-driven, the role of analytics and the use of strategic mathematical and statistical models in decision-making processes has become more prominent. In recent years, sports analytics has gained attention as a tool to improve player performance, enhance team strategy, and optimize fan engagement which can be actively monetized and used to generate material profits within and outside of the stadium. With the emergence of new technologies and data sources, sports organizations are now able to gather vast amounts of information on various aspects of the game. While the earliest recorded use of such techniques can be dated back to Henry Chadwick's Box Score in 1858, which used tabular representations of player performance in baseball to help analysts better assess the functioning capabilities of the team as well as individual players, this certain area of knowledge wouldn't gain significant traction till the 20th century as the changing capitalist economy of the world revolutionized how sports was viewed across the globe. One of the pioneers of sports analytics is Bill James, a baseball writer who used statistical analysis to develop new metrics that could better explain player performance. James' work was initially dismissed by the baseball establishment but eventually gained traction. By the 1990s, many Major League Baseball teams were using advanced analytics to make decisions about player personnel and game strategy.

More importantly, James is credited with having coined the term "sabermetrics" which is based on the idea that traditional measures of player performance, such as batting average or runs batted in, do not fully capture a player's contribution to the team. It seeks to identify new metrics that more accurately reflect a player's value (Baumer & Zimbalist, 2015) which was famously popularized in Michael Lewis's book, *Moneyball*, which detailed how the Oakland A's used data to select undervalued players and build a competitive team. Since then, sports analytics has become a vital tool for selecting players. Beane and his team's application of sabermetrics fundamentally changed how baseball players were evaluated, shifting the focus from traditional scouting methods to statistical analysis. *Moneyball* highlighted the potential of data analytics in sports and sparked a widespread interest in its application, leading to its adoption by several other teams. One of the significant methods used in the book was the analysis of on-base percentage (OBP) and slugging percentage (SLG). These metrics were found to be better predictors of a player's offensive contribution than batting average, a more traditional measure used in baseball. Beane used this knowledge to build a team of undervalued players with high OBP and SLG, allowing him to construct a competitive team despite a limited budget. This approach showcased how data analysis could be used to identify inefficiencies in the market and find undervalued players.

A Survey of Statistical Methods and Computational Tools

Regression analysis, a statistical technique used to identify relationships between variables, involves fitting a line or curve to a set of data points, and then using that line or curve to make predictions about future data points. The most common form of regression analysis is linear regression, which involves fitting a straight line to the data. However, other forms of regression analysis, such as polynomial regression and logistic regression, can also be used depending on the nature of the data. Similarly in sports analytics, regression analysis can be used in a variety of ways including simple linear regression, multiple linear regression, and logistic regression. For example, regression analysis can be used to predict the outcome of games based on various factors such as team performance, weather conditions, and home-field advantage. Researchers have used regression analysis to predict the outcomes of baseball games, football games, and basketball games, among others (Cervone, 2016; Zou, Chen, & Chen, 2020). For example, a study by Lewis and Burke (2015) used regression analysis to estimate the impact of different types of pitching on a team's winning percentage. They found that strikeout rate, walk rate, and ground ball rate were the most important predictors of team success and that pitchers who excelled in these areas were more valuable to their team. Similarly in professional fields, regression analysis may be used for the development of the Player Impact Plus-Minus (PIPM) metric for evaluating NBA players. PIPM is based on a regression model that estimates a player's impact on his team's performance, taking into account various factors such as their scoring, rebounding, and defensive ability (Goldsberry, 2019). PIPM has become a popular metric for evaluating NBA players because it provides a comprehensive measure of their overall impact on the game.

Bayesian analysis, another statistical methodology that involves the use of Bayes' theorem to update prior beliefs or probabilities in light of new data or evidence, is a powerful tool for data analysis that allows for the incorporation of uncertainty and prior knowledge into statistical models (Gelman et al., 2013). In sports analysis, not only can Bayesian analysis be used to model the performance of individual players or teams and make predictions about future outcomes but it can also be used to estimate the impact of different factors on game outcomes, such as weather conditions or player injuries (Liu et al., 2021). Bayesian networks, which are graphical models that represent the relationships between different variables in a probabilistic framework, have been particularly useful for modelling complex systems such as the interactions between players on a team or the effects of different training regimens on performance (Gudmundsson et al., 2018). Therefore, Coaches and analysts can use Bayesian methods to make informed decisions about game strategies, player selection, and other critical aspects of the game. By incorporating prior knowledge and uncertainty into their decision-making processes, they can make more accurate predictions and reduce the risk of making costly mistakes.

Time series analysis is another statistical methodology used to analyze data that varies over time and involves studying patterns and trends in the data to make predictions about future values or to identify underlying factors that influence the data. Time series analysis has numerous applications in many fields, including finance, economics, and engineering. In recent years, it has also become increasingly popular in sports analytics (Chen et al., 2018). It's extensively used to model and predict game outcomes. For example, researchers have used time series analysis to predict the results of basketball games (Gharanfoli et al., 2020) and soccer matches (Naseem et al., 2021) by analyzing historical data on game performance. By identifying patterns and trends in the data, these models can help to inform game strategies and improve team performance. Researchers have also used this technique to analyze data on baseball players and identify factors that influence their performance, such as changes in batting stance or injuries (Albert & Bennett, 2013). In soccer, time series analysis has been used to model the performance of individual players by analyzing their movement patterns on the field (Linkenauger et al., 2012). By analyzing data on player performance over time, researchers can identify patterns that may indicate an increased risk of injury. For example, researchers have used time series analysis to study the impact of training regimens on injury rates in soccer players (Bittencourt et al., 2016).

Machine learning, on the other hand, is a subfield of artificial intelligence which involves the development of algorithms that can learn from data and make predictions or decisions without being explicitly programmed and has become increasingly important in sports analytics with the rising computational technologies available as a means of analyzing large datasets and making accurate predictions about player and team performance. Parallel to other methods, it is often used to predict game outcomes. Researchers have used machine learning algorithms to analyze various factors that can influence game outcomes, such as team performance metrics, weather conditions, and player injuries, and make predictions about which team is likely to win (Liu et al., 2018). These predictions can be valuable for sports fans and bettors, as well as for coaches and analysts looking to make informed decisions about game strategy and player selection. Further, some studies have used machine learning algorithms to analyze vast amounts of player data and identify patterns and trends that may not be apparent to human analysts. For example, machine learning algorithms can be used to identify players who are likely to have a breakout season based on their performance metrics (Grund & Skalski, 2019). They can also be used to predict player injuries based on past injury data and other relevant factors (Aggarwal et al., 2019). What sets apart machine learning from other traditional statistical methods is its ability to formulate complex models and data networks like player and team performance models. These models can be further used to identify strengths and weaknesses in a team's performance and provide insights into how they can improve. For example, machine learning models can be used to identify the most effective play calls for a particular team based on their performance metrics (Owens et al., 2018).

Similarly, Decision trees are also a type of machine-learning algorithm that can be used to model complex decision-making processes. They involve recursively partitioning data into subsets based on the values of different predictor variables, and constructing a tree structure that represents the resulting sequence of decisions that can be used to predict game outcomes, identify key performance indicators for players, and optimise game strategies (Wang et al., 2020). For example, a decision tree model could be trained to predict the outcome of a basketball game based on the statistics of each team, such as field goal percentage, turnovers, and rebounds. What privileges them over traditional machine-learning algorithms is that they are easy to interpret and visualize, which makes them useful for communicating insights to coaches and players. Decision trees can also handle both categorical and continuous predictor variables and can handle interactions between different predictors (Fernández-Delgado et al., 2014). In the work of Gudmundsson et al. (2018), for example, decision trees were used to identify the most important features for predicting goals in soccer matches. They found that the most important features were the number of shots on goal, the location of shots on goal, and the distance to the goal. Similarly for American football, Wang et al. (2020) used decision trees to predict the success rate of different types of offensive plays. They found that the most successful plays tended to involve short passes or run up the middle and that the success rate of plays was strongly influenced by the down-and-distance situation.

Moral and Legal Concerns

While sports analysis has distinctly revolutionized the way we approach athletics by providing a data-driven approach to performance optimization, the source of such vast data accumulation and other aspects of the methodology puts forth multilayered ethical and legal dilemmas, which must be carefully navigated to ensure fair and transparent practices.

Athletes and teams generate vast quantities of data during training and competition, which can be gathered and scrutinized to gain insights and enhance performance. Yet, this process can result in the violation of athletes' privacy. Tracking devices that monitor an athlete's location and movements, for instance, can reveal sensitive information that can compromise their training habits and personal life. This breach of privacy can be seen as a violation of an athlete's fundamental rights. As Claudio Tamburrini (2018) points out, "The privacy of athletes must be respected, and any data collected must be used ethically and with consent." In other words, athletes must be fully informed of the data that is being gathered, how it will be used, and who will have access to it.

Furthermore, the anonymization of data should be considered whenever possible to safeguard the privacy of athletes. By doing so, sports analysts can navigate the ethical dilemmas associated with data collection and analysis, ensuring the responsible and respectful use of athlete data.

As discussed earlier, machine learning algorithms used in sports analysis rely on historical data to predict future events. However, these algorithms may perpetuate biases that exist in the historical data, leading to ethical concerns about fairness and discrimination in sports. For instance, if historical data is biased against female athletes, the predictions made by these algorithms may be less accurate for female athletes than for male athletes. This highlights the need for algorithms that are free from biases to ensure fair and transparent practices in sports analysis. Michael A. Lindsay and Paul M. Pedersen (2018) argue that it is crucial to develop algorithms and models that minimize the impact of biases in sports analysis. They suggest that sports analysts must take proactive measures to ensure that their algorithms are free from biases and that their predictions are fair to all athletes. By doing so, sports analysts can contribute to a more equitable and transparent sports industry.

Such data collection, which may be noted as the intellectual property of the athletes and their teams, can also be highly valuable to others in the sports industry (Chadwick & Nolan, 2018). For instance, a team's playbook or a player's training regimen may qualify as intellectual property, which is protected by copyright or trade secret laws (Hoffman & Rapp, 2018). Therefore, the unauthorized use of this data can result in legal action against the infringing party. According to Vacca (2019), sports analysts must exercise caution to ensure that they are not infringing on the rights of others. Specifically, analysts should seek permission from athletes and teams before collecting and using their data, and they should be well-versed in intellectual property laws to ensure compliance. It is, therefore, essential for sports analysts to navigate this complex landscape carefully, taking necessary precautions to ensure that their practices are ethical and legally compliant.

Case Study One: Data and Discussion

The Boston Red Sox, a professional baseball team in the United States, has long been known for its innovative use of sports analytics to gain a competitive edge. In recent years, the team has focused on using data analysis to identify undervalued players that traditional scouting methods may overlook.

One example of this is the signing of J.D. Martinez before the 2018 season. Martinez, a talented hitter, had struggled with injuries and inconsistent performance in previous seasons. However, the Red Sox used data analysis to identify that Martinez had made significant changes to his swing mechanics, which had led to improved performance in the previous year. Using data from Statcast, a technology that tracks player performance metrics, the Red Sox were able to identify that Martinez had made changes to his launch angle and exit velocity, which had resulted in a higher rate of hard-hit balls. This information allowed the Red Sox to project that Martinez would be an impactful hitter for their team, even though his previous performance may have suggested otherwise. Based on this analysis, the Red Sox signed Martinez to a five-year contract worth $110 million. The decision paid off, as Martinez had a phenomenal season in 2018. He hit 43 home runs and had a batting average of .330, leading the Red Sox to a World Series championship. Martinez was also named an All-Star that year and finished fourth in the voting for the American League Most Valuable Player award.

This example highlights how sports analytics can help identify undervalued players and make informed decisions about player selection. By using data to identify patterns and trends in player performance, teams can make better decisions about which players to sign and how to allocate their resources. The Red Sox is not the only team to use this approach, as many other professional sports teams have also invested heavily in sports analytics in recent years. According to a survey by the Wall Street Journal, over 90% of Major League Baseball teams now have a dedicated analytics department, and many have multiple analysts on staff. This trend is not limited to baseball, as other sports such as basketball and soccer have also seen a significant increase in the use of sports analytics in recent years.

Case Study Two: Data and Discussion

In 2014, the Sacramento Kings began a partnership with a startup company called YinzCam to create a mobile app that would allow fans to interact with the team in new and innovative ways (Kerwin, 2018). The app, called "Call the Shot," was designed to engage fans during live games by asking them to predict the outcome of in-game events. Fans could earn points for making accurate predictions, and those points could be redeemed for prizes such as autographed memorabilia or game tickets. According to the team, over 70% of fans who attended games during the 2014-2015 season downloaded the app and participated in the Call the Shot game (Brauer, 2016). In addition, the team reported that fans who used the app were more likely to return for future games, buy team merchandise, and engage with the team on social media.

Fig. 1. Call the Shot by Sacramento Kings. NBA.Com.

But the Call the Shot app was more than just a fun game for fans. It also provided valuable data for the Kings front office to better understand their fans and their preferences. By analyzing the data collected from the app, the team was able to identify which in-game events were most popular among fans and which types of prizes were most effective in incentivizing participation (Brauer, 2016). The team also used the data to create new opportunities for fan engagement. For example, the Kings began offering special in-game experiences to fans who earned a certain number of points in the Call the Shot game, such as the opportunity to shoot a half-court shot for a prize or to participate in a post-game press conference with a Kings player (Kerwin, 2018). The success of the Call the Shot app inspired the Kings to continue exploring new ways to use sports analytics to engage with their fans. In 2016, the team introduced a new program called "Loyalty Coin," which rewarded fans for attending games and engaging with the team on social media. Fans could earn points by checking in at games, sharing team content on social media, and attending team events. Those points could then be redeemed for prizes such as game tickets, autographed memorabilia, and even the opportunity to travel with the team to an away game (Kerwin, 2018). The Kings have also used sports analytics to improve the in-game experience for fans. In 2018, the team introduced a new feature on their mobile app called "Kings + Golden 1 Center," which provides fans with real-time information about wait times at concession stands, restroom availability, and parking information. The team also uses data from the app to better understand which areas of the arena are most popular among fans and which types of promotions are most effective in driving ticket sales (Kerwin, 2018).

Case Study Three: Data and Discussion

Southampton FC used a combination of data analysis and scouting to identify players who were undervalued by other teams. The team looked at various factors such as player statistics, age, and injury history to identify players who were likely to be good value for money (Shakir, 2017). One such player who was identified through this process was Sadio Mane, who was signed by Southampton FC for £11.8 million in 2014. At the time, Mane was relatively unknown and had only played one season in the Austrian Bundesliga. However, Southampton FC's analysis showed that Mane had impressive underlying statistics, such as his high expected goals and assists (xG and xA) per 90 minutes, indicating that he had the potential to be a valuable addition to the team (Shakir, 2017). The following table represents some of the players that Southampton FC identified as undervalued through data analysis and went on to have successful seasons with the team, thus highlighting the value that Southampton FC was able to get by using sports analytics.

Mane went on to have a successful season with Southampton FC, scoring 10 goals and providing 3 assists in 30 appearances. His strong performances caught the attention of Liverpool FC, who signed him for a fee of £34 million in 2016 (Shakir, 2017). Mane has since become one of the best players in the English Premier League and has helped Liverpool FC win numerous trophies, including the Premier League title in the 2019-2020 season. Southampton FC's success in using sports analytics to identify undervalued players has inspired other teams in the Premier League to adopt similar strategies. For example, Brentford FC, a club that was recently promoted to the Premier League, has gained a reputation for using advanced data analysis to identify undervalued players and has been successful in recent years (Taylor, 2021). This highlights the importance of data analysis in professional sports. By using data to identify undervalued players, teams can gain a competitive edge over their rivals, ultimately leading to better performances on the field.

Player	Signed for	Appearances	Goals	Assists
Sadio Mane	£11.8 million	30	10	3
Virgil van Dijk	£13 million	68	7	2
Dusan Tadic	£10.9 million	162	23	30
Graziano Pelle	£8 million	81	30	12
Charlie Austin	£4 million	71	20	5

Fig. 2: Undervalued Players Signed by Southampton FC Using Sports Analytics. Soutamptonfc.com

Conclusion

In conclusion, sports analytics has had a significant impact on player selection, team performance, and fan engagement. The use of advanced statistical techniques has allowed teams to identify undervalued players and make more informed decisions about player acquisition. Teams have also been able to optimize their strategies and tactics based on insights gleaned from data analysis. As a result, teams that embrace sports analytics have a competitive advantage over those that do not. The case studies presented in this paper demonstrate the versatility of sports analytics and how it can be applied in different ways to achieve various goals. From engaging fans to winning championships, sports analytics can provide valuable insights and drive success for teams. Furthermore, the increasing adoption of sports analytics in professional sports highlights the importance of data-driven decision-making in today's competitive landscape. As technology and data collection methods continue to evolve, we can expect sports analytics to play an even greater role in the future of sports.

References

[1]. Albert, J. (2003). Bayesian Computation with R. Springer Science & Business Media.

[2]. Albert, J. (2016). Basketball analytics: Objective and efficient strategies for understanding how teams win. CRC Press.

[3]. Albert, R. (2013). The challenges of sports analytics. Communications of the ACM, 56(4), 32-34.

[4]. Burke, M., Joyce, D., & Greenberg, A. (2014). The growing importance of analytics in the NBA. Journal of Quantitative Analysis in Sports, 10(2), 121-127.

[5]. Deutscher, J. C., & Anderson, M. B. (2016). Basketball on paper: Rules and tools for performance analysis. CRC Press.

[6]. ESPN Staff. (2018, February 26). Why the Red Sox's J.D. Martinez might be a bargain even at $110 million. ESPN. https://www.espn.com/mlb/story/_/id/22502188/why-boston-red-sox-jd-martinez-might-bargain-even-110-million

[7]. Gardner, C. J., & Singer, J. (2012). The ethics of using medical data from Nazi experiments. The American journal of bioethics, 12(9), 31-37.

[8]. Gershman, J. (2017, March 27). How the Red Sox are winning over fans of 'Moneyball.' Wall Street Journal. https://www.wsj.com/articles/how-the-red-sox-are-winning-over-fans-of-moneyball-1490654273

[9]. Gómez, M. A., Lago-Peñas, C., & Pollard, R. (2018). The evolving role of performance analysis within sports coaching. International Journal of Performance Analysis in Sport, 18(6), 999-1014.

[10]. Hill, M. P., & Cao, J. (2021). Big Data Analytics in Sports: Ethical and Legal Considerations. Journal of Business Ethics, 168(1), 1-15.

[11]. Kannan, V. R., & Narasimhan, R. (2020). Application of analytics in sports management: A review. Journal of Business Research, 109, 555-568.

[12]. Lewis, M., & Burke, B. J. (2015). The impact of pitching on team success. Journal of Quantitative Analysis in Sports, 11(4), 195-212.

[13]. Li, X., & Yang, Y. (2020). The challenges and opportunities of sports analytics. Journal of Business Research, 109, 541-554.

[14]. MLB.com. (2018, November 15). JD Martinez stats, fantasy & news. MLB.com. https://www.mlb.com/player/jd-martinez-502110

[15]. Schatz, A. G. (2013). Football outsiders Almanac 2013: The essential guide to the 2013 NFL and college football seasons. CreateSpace Independent Publishing Platform.

[16]. Statcast. (n.d.). MLB.com. https://www.mlb.com/news/statcast

[17]. Tepfer, B. (2018, August 16). Red Sox's J.D. Martinez turning into one of game's best hitters. NBC Sports. https://www.nbcsports.com/boston/red-sox/red-soxs-jd-martinez-turning-one-games-best-hitters

[18]. Thaler, R., & Sunstein, C. R. (2009). Nudge: Improving decisions about health, wealth, and happiness. Penguin.

[19]. Wertheim, L. J. (2019). The Ringer guide to the 2019-20 NBA season. Hachette UK.

[20]. Xu, Y., & Hwang, S. (2015). Predicting outcomes of soccer matches using data mining. Journal of Computer Science and Technology, 30(3), 537-548.