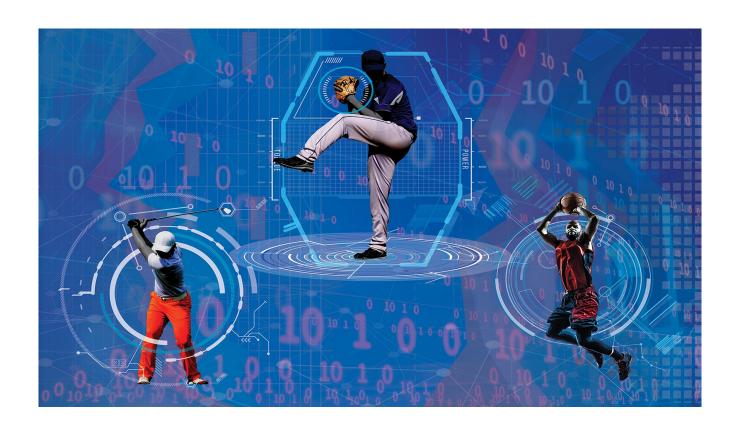# DATA-437
Sports Analytics
Lecture Notes

Oscar Aguilar

Grand View University

Fall 2021

Chapter 1

Introduction to Sports Analytics

1.1 Introduction

The sports industry's transformation is ongoing as clubs, leagues, broadcasters, venue operators, and professional players increasingly see the value of analytics to identify metrics and patterns that may not be obvious to the traditional scout/manager's eye.

According to a recent study by Grand View Research Inc, the global sports analytics market size expands at a CAGR of 31.2% and is expected to reach $4.6 billion by 2025. Increasing preference for sports as a career option and the growing demand for tracking and monitoring players' live data are expected to drive market growth. The use of analytics in sporting events helps various stakeholders, including sportsperson, associations, and fans, to gain in-depth insights on live in-game activity and past game events.

1.2 What is sports analytics?

Sports analytics are a collection of relevant, historical, statistics that can provide a competitive advantage to a team or individual players. Through the collection and analyzation of these data, sports analytics inform players, coaches and other staff in order to facilitate decision making both during and prior to sporting events. The term "sports analytics" was popularized in mainstream sports culture following the release of the 2011 film, *Moneyball*, in which Oakland Athletics General Manager Billy Beane (played by Brad Pitt) relies heavily on the use of analytics to build a competitive team on a minimal budget.

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 uses data to help rightsholders take decisions that would lead to higher growth and increased profitability.

1.3   Technology and Sports Analytics

As technology has advanced over the last decade, data collection has become more in-depth and can be conducted with relative ease. Advancements in data collection have allowed for sports analytics to grow as well, leading to the development of advanced statistics and machine learning, as well as sport specific technologies that allow for things like game simulations to be conducted by teams prior to play, improve fan acquisition and marketing strategies, and even understand the impact of sponsorship on each team as well as its fans.

Technology companies are making advancements in developing wearable devices for sports teams. Companies are transforming the scene with accurate measures of performance metrics, health, and safety of players allowing coaches and staff to make accurate decisions. With the increase in demand for high-intensity performance in sports, players are at high risk of attaining injuries. Wearable sports technologies are used to monitor training and in-game performance, reduce injury, illness, and keep track of recovery after an injury.

Wearable athletic devices are available in different shapes and sizes. The devices' seamless design is also incorporated into the fabric of sports apparel, built into sports equipment like balls and bats, and worn by athletes as small devices attached to the body in a waistband or skin patch and shoes. These devices are enabled with Bluetooth and GPS technology, providing real-time feed to coaches on their laptops or other electronic devices for analysis.

1.4   Advanced Analytics in Sports

In the section, we briefly review some of the areas in sports in which analytics have been successfully applied.

1.4.1   Injury Prediction

Many athletes and teams' most sought-after information is insights into when conditions may heighten the risk of injury. For teams, injuries can have a financial impact on revenue generation due to medical expenses, recovery, sponsorships, and ticket sales by losing competitions at early stages. Similarly, for players, having information on preventing injuries can help extend their careers, earnings and increase their valuation to the highest potential. Predicting injuries more effectively requires measures that help balance exertion and strain with the proper amount of recovery time, nutrition, and sleep.

Injury sustained due to excessive training, or game load can lead to an increase in injury rates. Logistic regression models using binomial distribution can help identify how players react to a particular training stimulus and determine the potential injury probability. The models can be categorized based on the stage (pre-season, early competition, late competition) of the season. The training workload can be adjusted accordingly to avoid the risk of injury.

Neuromuscular data is obtained by combining force platforms and motion analysis software to identify how each player takes advantage of different body muscles and their speed, reaction time, and weak points. Analyzing postures that cause a risk of injury can be corrected using

motion capture and high-speed cameras. Deep learning algorithms like Convolutional Neural Networks (CNNs) models can be built to understand better any deviation in an athlete's posture and technique.

1.4.2   Player Scouting

Teams making a financial investment in a player increasingly use automated video analysis and positional and tracking data in their scouting. Such insights give them confidence while assessing a player?s skills, biometrics, and medical information virtually. This process has helped immensely, especially during the wake of the pandemic.

Similarly, player scouting becomes a lot easier with the use of machine Learning techniques such as clustering and statistical analysis, which provide scouts a data-driven approach to identify the right players. Clustering can be performed after identifying the key metrics influential in selecting the right player. The k-means algorithm can effectively cluster the group of players into different clusters, which significantly simplifies the process of looking at the data. These clusters will help identify low financial value players who have similar impacts compared to celebrated players.

1.4.3   Strategy

Predicting the strengths, weaknesses, and tendencies of opposition teams and their personnel can help identify the right strategy for any game situation. The patterns of player movements are available using GPS tracking metrics, which opposition teams can leverage. Teams constantly develop themselves and no longer stick to a single formation throughout the game.

Formations are measured by calculating the vectors between each player and the rest of the teammates at successive instants during a match, averaging the vectors between each pair of players over a specified time interval to gain a precise measure of their designated relative positions. Teams can alter their strategies by identifying the defensive and offensive formation clusters most frequently paired together. Analytics/data science in sports can help maximize wins by offering trusted insights on what will likely happen after each decision to extract the best performance.

1.4.4   Season Ticket Churn

Retaining existing season ticket holders is cheaper than acquiring new ones. Churn prediction and identifying reasons behind churn become critical to sport organizations to predict their ROI. Factors influencing churn can be attributed to poor on-field performances, low game attendance, and low customer engagement. Churn prediction models using logistic regression can identify the season ticket holders that are likely to churn. Strategies to increase customer engagement through campaigns and promotions can help reduce the churn rates. Additionally, statistical techniques such as hypothesis testing using *paired t-tests* can be conducted to help us understand the impact of a campaign on a customer.

1.4.5   Player Valuation & Development

Building better rosters by understanding a player's actual value and the associated risks can save significant money for an organization. Financially smaller teams can now compete in bigger leagues by signing the right players with a data-driven approach. The advantage smaller teams hold is their ability to provide time for players to adapt to a system that helps the overall development of a player with potential. Training regimes and strategies can be created through analysis which can increase player value. Similarly, rapid feedback on a player's game or practice performance can be highlighted to analyze a player's strengths and weaknesses.

1.4.6   Pricing

Ticket sales generate the maximum revenue for any sporting organization. A ticket pricing model can help maximize revenue by evaluating the ticket price based on historical data. The data can help organizations identify the occupancy rates based on the opposition or the competition, thereby adjusting ticket rates based on its target revenue.

1.5   Conclusion

Recently, many sporting organizations are investing in sports analytics, and the returns are very evident. Recently, Manchester City hired Laurie Shaw, a former hedge fund veteran, to lead AI insights at City Football Group. The primary focus is on building machine-based models to manage player fatigue, injury, scouting, pre-match analysis, post-match analysis, and coach recruitment.

This is just the tip of the iceberg. The reliance on sports analytics will increase manifolds with the emergence of advanced tracking devices and data collection set-up. Some of the emerging areas include wearable devices industry, medical industry, insurance, betting, and the gaming industry. It is high time for sporting organizations to invest in sports analytics or seek support from advanced analytics companies to stay resilient in this modern era.