Module 1. Fundamentals of data science

Welcome to module 1 of the Fundamentals of Data Science course. In this module, we will learn about the definitions and history of data science and how it applies to sports.

Unit 1. Background

Data is everywhere, in every facet of our lives, from the real-time sports statistics shown during TV broadcasts to the personalized product recommendations we receive online. In today's hyper-connected digital world, data generation, collection and analysis have become fundamental to how businesses, organizations and individuals make decisions.

The sports industry is an excellent laboratory for data science. From player performance analytics to fan engagement strategies, data-driven approaches have revolutionized how teams, leagues and media organizations operate. Computer vision and machine learning advancements have created new data sources rich in features and context.

These datasets have opened new analytics possibilities, allowing us to answer questions we couldn’t address before due to a lack of granular data.

In this course, you will explore and learn about data science through the lens of sport. The fundamental concepts, techniques, and applications define the ever-changing data science field. You will learn about the tools and methods of data analysis. Along the way, you will discover and appreciate how a data-driven approach has yielded positive results on and off the pitch and continues changing the sports landscape.

In this module and in the rest of the modules in the course, we will go through the evolution of data science in sports. From the early pioneers of sports analytics to the sophisticated applications built on context-rich data available today. We will examine how data has changed almost every sports operation, from talent identification and player development to in-game decision-making and fan engagement. Along the way, you'll build a robust foundation in the fundamental principles of data science that can be applied across various industries and domains.

On a personal level, we interact a lot with data and subconsciously do a lot of data analysis. For example, when we research buying a new car, we go through a ton of data, from the brand and the car's color to fuel economy, safety features, etc. We compare information on multiple car models and arrive at a decision. That is a textbook example of data analysis.

Jobs in data science, such as data scientists, analysts, engineers, etc., are seen as the most in-demand jobs in the 21st century. The theory, techniques, and tools you will learn in this course are applicable across every field and domain that uses data to solve problems.

We hope you find the rest of this module and the course valuable and engaging.

Unit 1.2 Definitions

We start the module by defining some terms used in the data world, which we will refer to throughout the course. Where applicable, examples from both sports and non-sports contexts are provided.

1.2.1 Data

Data can be defined in many ways, for example, we can say that it consists of a collection of unstructured facts, observations, perceptions, numbers, characters, symbols, and images that lack context or interpretation

This is the simplest and the most basic version of the definition.

We can describe data in many ways.

- Types
  - Quantitative: numerical information, e.g., a set of numbers - 2, 3, 5, 10. Or 1 kilo, 2 kilos, 3 lbs.
  - Qualitative: descriptive, non-numerical information, e.g., “roses, lilies, lavender”; “red, white, violet”.
- Formats
  - Structured: organized in a predefined format (e.g., databases, Excel spreadsheets).
  - Unstructured: lacks a specific structure (e.g., free text documents, images, video, audio files).
  - Semi-structured: a mix of both structured and unstructured data (e.g., JSON data, XML files).
- Sources
  - Primary: this is data collected directly from the source. It is usually done directly from a sensor, think of a weather thermometer

collecting the air temperature at your local weather station. In sports, data collectors collect live data from matches as the match happens.

  - Secondary: secondary data is derived from existing data. This data is created from data already collected from a primary source.

    Example: let us use the same temperature instance collected at a weather station. You can derive how much more (or less) the temperature was on the same day last year using the temperature collected on the same day the previous year.

    In sports, companies collect pitch-by-pitch data (in baseball or cricket), play-by-play data (in basketball), or capture every touch, pass and shot (in football). You can use that information to create derived statistics, like shot goals ratio, pass completion %, shooting percentage, etc.; these are secondary data derived from existing data.

- Characteristics
  - Volume: data can be distinguished by its size. Datasets can range from small (measured in kilobytes to a few megabytes) to big data (measured in gigabytes or larger), and anywhere in between.
  - Velocity: how fast the data is generated. Think of live data vs. data that takes a lot of post-processing.
- Uses
  - Information
  - Analysis and insights
  - Decision support
  - Predictions and forecasting
  - Machine learning and AI training
- Management
  - Collection
  - Storage

- Processing
- Analysis
- Visualization
- Security and privacy

Data plays a crucial role in various fields, including business, science and technology, forming the foundation for information and knowledge.

1.2.2 Data analysis

Data analysis is the process of exploring, organizing, cleaning, translating and transforming data via techniques like data modeling to draw conclusions, discover valuable information, answer questions and support decision-making processes.

Exploration. Understand the structure and characteristics of data through exploration via plotting and loading the data into an Excel spreadsheet (or a database).

- Understand the data format: structured, semi-structured or unstructured.
- Shape of the data: different columns and data types in the data.
- Size: number of rows.
- Distribution: how the data is laid out (we will discuss this in more detail in later modules).

Data cleaning. Involves correcting data for format errors and inconsistencies. One of the common issues with datasets is missing data. Determining how to handle missing data is also a part of the data cleaning process.

Examples: data format mismatches or filling up missing blank values with 0's or "N/A" are a couple of typical examples of data cleaning.

Transform. Transform the data into a format that makes further steps in the analysis easy. These could be as simple as pivoting data, aggregation, normalization, etc.

Modeling. Applying statistical, mathematical or computational techniques to extract patterns and insights from the data.

Conclusions. Interpret the modeling results and make logical inferences based on the output of the models and analyses. Conveying the context of the findings along with their limitations.

Decision making. Use the inferences and conclusions from the analysis to help make decisions or act.

1.2.3 Data analytics

Data analytics is the overall method of using data to draw insights, predictions and forecasts to help decision-making. Although the name sounds like data analysis, it is narrower in scope and is often a specific data analysis case. Analytics emphasizes the creation of actionable intelligence and the ultimate process of using the intel in making decisions.

Data analytics is a comprehensive approach to using data for insights, predictions and decision-making. It typically involves several key components.

- **Descriptive analytics:** this involves summarizing historical data to understand what has happened. An example is analyzing data from a finished NBA match, summarizing the point scorers for both teams and identifying the top scorers.

- **Diagnostic analytics**: this digs deeper to understand why something happened.

  Example: a football club fan website investigated a sudden decrease in page views by examining the recent changes they made to the content or design of the website pages.

- **Predictive analytics:** this uses historical data and statistical models to forecast future outcomes.
  Example: using data from past X seasons to build models that expected points a team would score in the next season (future). These models are widely used in sports betting to set up betting lines and prices.

- **Prescriptive analytics**: this suggests actions based on the insights gained.

  Example: a baseball pitcher receives an analysis from his analytics department about opponent batters, identifying which pitch types each batter is weakest at hitting. This insight will help the pitcher devise a pitch strategy to exploit the hitters' weaknesses.

- **Big data analytics:** this involves processing and analyzing massive datasets.

  Example: an NBA team analytics department uses optical tracking data from all the matches in the NBA from prior seasons to detect opponents' most used off-ball movement patterns. Optical tracking data collects the position of every player on the court at least 25 times a second. Each match file is a few GB. There are 1230 matches in the regular season of the NBA. So, a few Terabytes of data are generated every season. This is big data.

- **Real-time analytics:** real-time analytics use that is collected in real-time as a game is happening to generate immediate insights.

  Example: football teams receive up-to-the-minute match data (both numbers and video) from the bench. Teams use video clips of important moments in the game to adjust and adapt as soon as possible.

  Another example in sports is the team's fitness staff monitoring players live GPS data continuously during a match to make changes if they see a player is tired.

- **Visual analytics**: visual analytics uses visuals like maps, diagrams and interactive charts to aid data analysis.

  Example: using visuals like heat maps and touch maps to visualize the influence of every player on a football match

- **Text analytics:** text analytics extracts meaningful information from free and unstructured text data that is used to extract meaningful information.

  Example: a sports team's fan engagement department uses tweets and messages about its team on social media platforms to extract the prevailing sentiments of its fan base. This helps them address any negative feelings or feedback.

Data analytics uses advanced tools and techniques like Natural Language Processing (NLP) for text analysis, BI tools like Tableau and Power BI, data mining, machine learning and different statistical models.

Data analytics transforms raw data into actionable insights that drive strategic decision-making, improve operations, and create competitive advantages.

1.2.4 Big data

Big data refers to large, complex data sets that are difficult to process and manage. They need unique storage solutions and custom computing setups that involve a lot of parallelization. In many cases, this type of data is also collected rapidly.

Example: the non-sports world tends to have more extensive and more complex datasets than sports. Search engines collect a ton of information when users search on their platform. The data collected daily is in the magnitude of petabytes (1 PB = 1000 TB = 1000,000 GB).

1.2.5 DIKW Framework

Data – Information – Knowledge – Wisdom is sometimes called the "wisdom hierarchy" or the DIKW framework. Data is at the bottom and wisdom is at the top of this hierarchy. Each level builds on the other by adding value to the previous level.

- **Data**

  We have already discussed the data in detail. Data is an unorganized list of facts without any context.

- **Information**

  Information is processed, organized and contextualized data to make it meaningful. It typically answers "who, what, where, and when" questions. It adds relationships between data points.

- **Knowledge**

  Knowledge is information that is understood, applied and integrated into a broader context. It involves recognizing patterns and using information in specific situations.

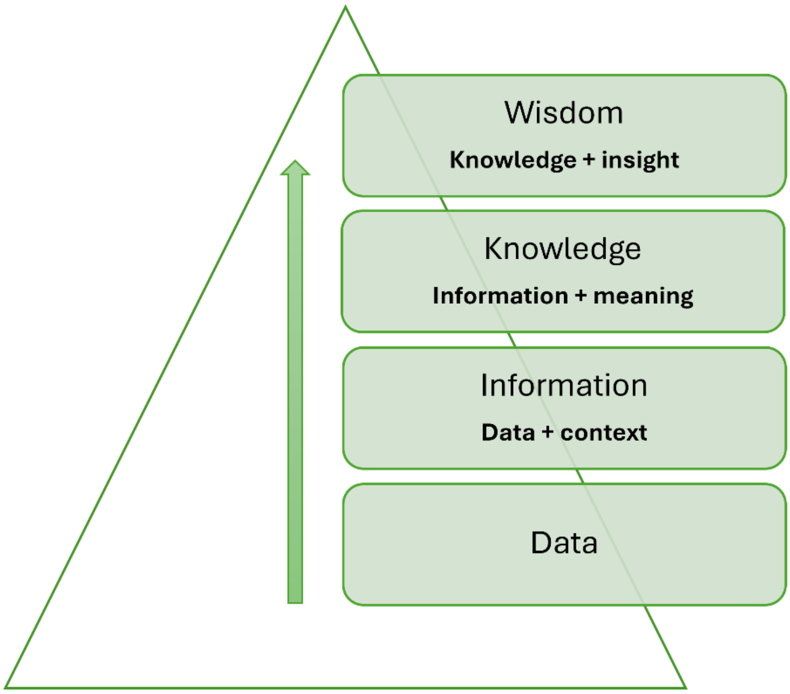
- **Wisdom**

  Wisdom is the top level of this DIKW hierarchy. It signifies the highest understanding. It answers the question "why" and involves judgment. It includes ethical considerations and long-term thinking. Data or knowledge alone is not sufficient; it requires experience.

  For example, knowing that a tomato is a fruit, is knowledge. But it is wise not to add tomato to a fruit salad (wisdom).

  The diagram below summarizes the DIKW framework.

Figure 1. DIKW framework



Source: own elaboration.

In each step of the hierarchy, there is additional context and understanding. After each step, the data is more usable and actionable.

Transferring data and information between people or systems is accessible at the lower levels of data and information. Still, it is much more complicated at the higher levels of knowledge and wisdom.

Technology and automation are more effective at the lower levels of data and information. Human intervention can be substituted at these levels without any issue. However, at the higher levels, more human intervention is required as there is an element of "judgment."

The DIKW framework provides a structure for the entire data science workflow, from raw data to actionable intelligence/wisdom.

While very useful in enhancing our understanding of data, this model has some limitations. The boundaries between different levels in the hierarchy can be vague at times. It doesn't account for the role of human experience and judgment in creating wisdom. Nevertheless, it is an excellent framework to understand the relationships between data, information, knowledge and wisdom.

Example

1. "22 °C" – This is data. A raw fact that represents temperature. Without context, this is not very useful, and it doesn't tell us anything.

2. “The temperature in Barcelona is 22 °C” - This is data with additional context that it is the temperature in Barcelona today.

3. “Knowing that 22 °C is a suitable temperature for outdoor activities” – This is knowledge. We know from experience that 22 °C is ideal for outdoor activities.

4. “It is wise to organize outdoor activities when the temperature is forecast to be 22 °C” – We are using knowledge and experience to make decisions.

1.2.6 Data science

Data science is the field that helps move data up the “wisdom hierarchy.”

Data science is the field in which we draw insights and knowledge from data using algorithms and scientific methods to transform data into actionable and applicable intelligence. It spans multiple disciplines, such as statistics, mathematics and computer science, along with expertise in the specific area where you use the data.

Mathematics and statistics are foundational to data science as they provide the theoretical basis for analyzing data. The algorithms to model and calculate + the methods to scientifically test the hypotheses. You can say they give the “what” and “how to”.

Computer science adds to the “how” by enabling the application of algorithms and methods at scale across vast amounts of data through computing power and the ability to organize, store and manipulate large amounts of data.

The importance of computer science in data science has increased manifold in recent years due to an explosion in the amount of data being generated every second.

Examples:

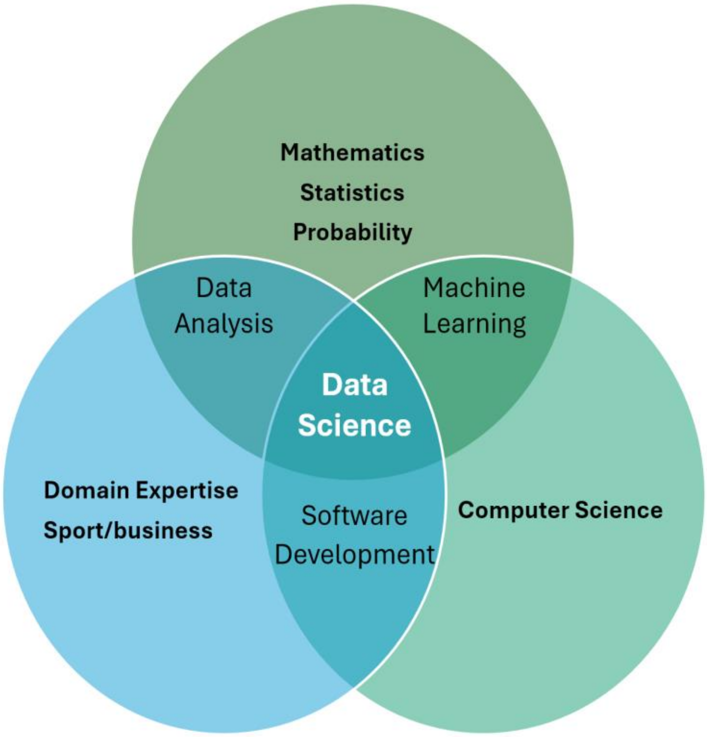
- Amazon sells millions of products daily. Each sale adds new data to their databases.

- 228 million search queries are made every hour worldwide on search engines like Google, Bing, etc. Each query generates search logs with 200 – 400 pieces of metadata stored on Google/Microsoft's data infrastructure.

- In sports, we went from keeping track of the score, goal scorers, starting lineup and substitute information, which is very small, to capturing the physical location of every player on the field 25 times a second (optical tracking data).

Data science has emerged as an indispensable field of expertise in the 21st century. No field in the world doesn’t have data science applications and a team of experts in data science doing meaningful work. In this course, we will limit data science and its applications in sports.

Figure 2. Data science



Source: own elaboration.

1.2.7 Relation between data analysis, data analytics and data science

The terms "data science," "data analysis" and "data analytics" are often used interchangeably, but there are some key differences and similarities between them.

Differences

1. **Scope and approach.**
    - **Data science**: encompasses the entire data lifecycle, from data collection and preprocessing to modeling, analysis and the communication of insights.
    - **Data analysis**: is a more focused task within the broader field of data science. It involves examining, cleaning, transforming and modeling data to uncover patterns, trends and insights.

- **Data analytics**: is a specific subset of data analysis, focusing more on applying analytical techniques and tools to solve business problems and make data-driven decisions.

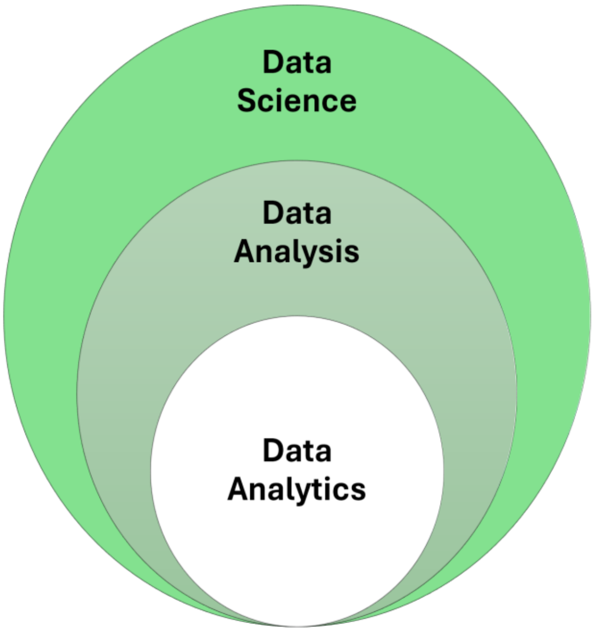
2. **Depth of expertise.**

    - **Data science**: data scientists typically have a deeper understanding of statistical modeling, machine learning and advanced analytical techniques, as well as the ability to apply them to complex, unstructured data.
    - **Data analysis**: data analysts often have a strong foundation in data manipulation, visualization and interpretation, but may not have the same depth of expertise in advanced statistical and machine learning methods as data scientists.
    - **Data analytics**: data analysts are typically more focused on using established analytical techniques and tools to address specific business challenges and inform decision-making.

3. **Outcomes**.

    - **Data science**: data science aims to develop new techniques, models and algorithms to generate novel insights and solve complex problems.
    - **Data analysis**: data analysis focuses on understanding the current state of a problem or situation and providing insights based on the available data.
    - **Data analytics**: data analytics is oriented towards making data-driven decisions and improving business performance by applying analytical techniques.

Figure 3. Relationship between data science, data analysis and data analytics



Source: own elaboration.

The simple Venn diagram captures the relationship between data analytics, data analysis and data science.

To illustrate the differences, let us go through an example of basketball.

Data analysis:

- Calculating a player's average points, rebounds and assists per game.
- Comparing the team's performance in different quarters.
- Analyzing free throw and three-point shooting percentages.

Data analytics:

- Identifying which player combinations have the highest winning percentage.
- Determining the optimal shot selection based on player and game situation.
- Analyzing opponent's play styles to develop effective defensive strategies.

Data science:

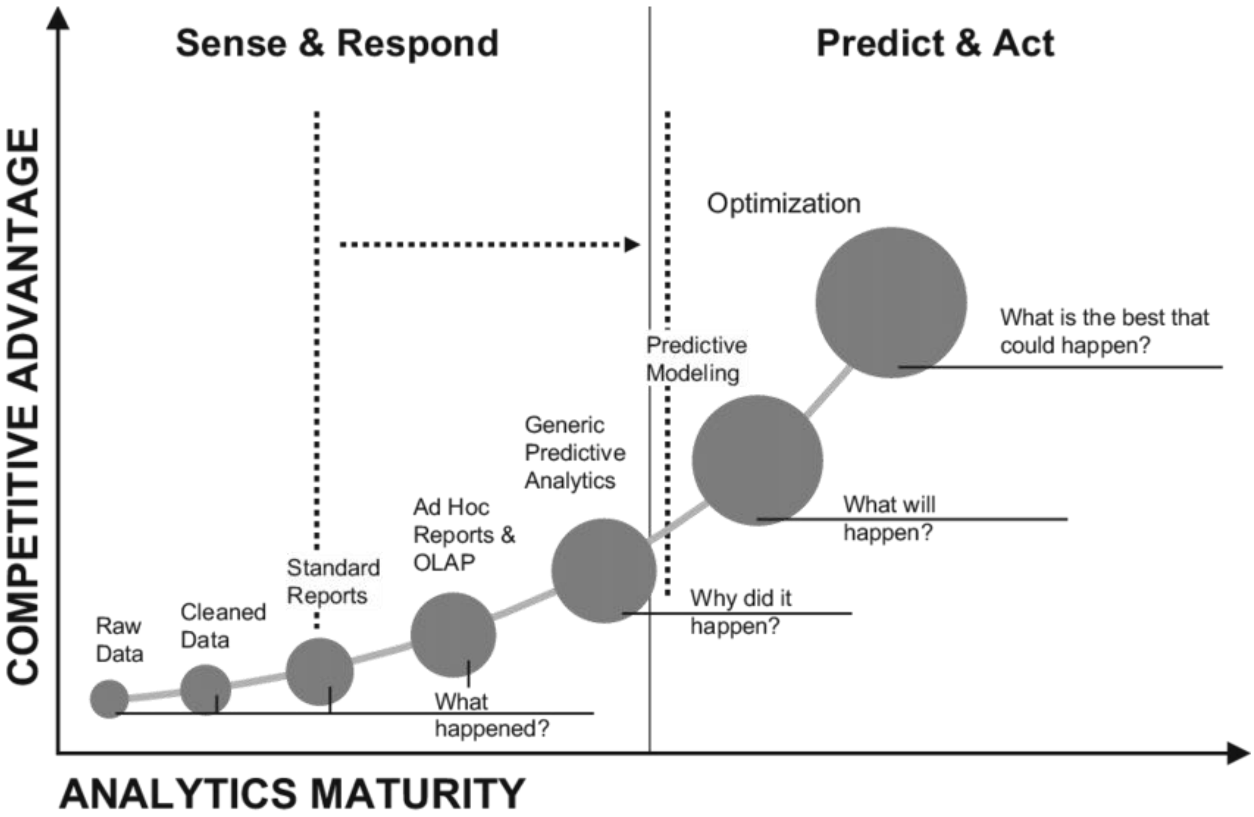
- Predicting player performance based on age, injuries and playing time.
- Developing algorithms to optimize player rotations and substitutions.

- Machine learning analyzes video footage to identify player movement patterns and potential weaknesses.

The boundaries between these three can be blurred in the real world, as it is typical for the same person to operate simultaneously in more than one of these areas.

Unit 1.3 Evolution of analytics capabilities

Figure 4. Evolution of analytics capabilities



Source: Brocke et al., 2017, https://lc.cx/hwwXBc

Unit 1.4 History of data science in sports

Before we proceed, it is instructive to examine the history of the data science revolution, including how it started, where we are now, and how we got here.

The use of data in sports has roots dating back centuries, though the analytical approaches have become increasingly sophisticated over time. In ancient Greece, records of the results and times of foot races during the Olympic Games were kept. Similarly, in ancient China, court historians meticulously tracked the performances of athletes and martial artists.

As organized sports emerged in the 19th and early 20th centuries, collecting and analyzing basic statistics became more common. In baseball, box scores recording batting averages and other vital metrics were kept from the sport's earliest professional days in the 1870s. Similarly, in the early 1900s, basketball began tracking points, rebounds, assists and other basic box score statistics.

However, the real revolution in sports analytics began in the 1950s with pioneering work by individuals like Charles Reep in soccer and Allan Roth in baseball. Reep's meticulous manual tracking of every pass and shot in soccer matches represented an early attempt to uncover strategic insights through data. Meanwhile, Roth used statistical analysis to challenge conventional baseball wisdom, laying the groundwork for the "sabermetric" movement.

The rise of computers in the 1960s and 1970s was a significant catalyst, allowing for storing and analyzing vastly more detailed sports data. Practitioners like Bill James in baseball and Jeff Sagarin in college basketball became famous for developing new analytical techniques and metrics. This led to a sea change in how teams and fans understood and evaluated athletic performance.

The 1990s and 2000s saw an explosion of new data sources, from video analysis to wearable technologies and optical tracking systems. Suddenly, sports organizations had access to a firehose of granular performance data, capturing every movement and action of athletes in unprecedented detail.

This dramatically expanded the potential applications of data science in sports, from talent identification and player development to in-game strategy and fan engagement. However, it also created new challenges as teams struggled to effectively organize, analyze and derive insights from these vast troves of data.

The response was the emergence of specialized sports analytics departments featuring interdisciplinary teams of statisticians, computer scientists and domain experts. These "sabermetricians" and "sports data scientists" have become indispensable assets for teams and leagues looking to gain a competitive edge.

Today, the influence of data science in sports continues to grow, transforming everything from scouting and recruitment to media coverage and fan experiences. As data collection and analysis capabilities continue to advance, the role of the sports data scientist has never been more crucial. The future holds exciting new frontiers as the sports world increasingly embraces the power of data-driven insights.

Unit 1.5 Data science in sports

The increasing adoption of data-driven approaches has transformed the world of sports, revolutionizing both the technical and business aspects of team and individual sports.

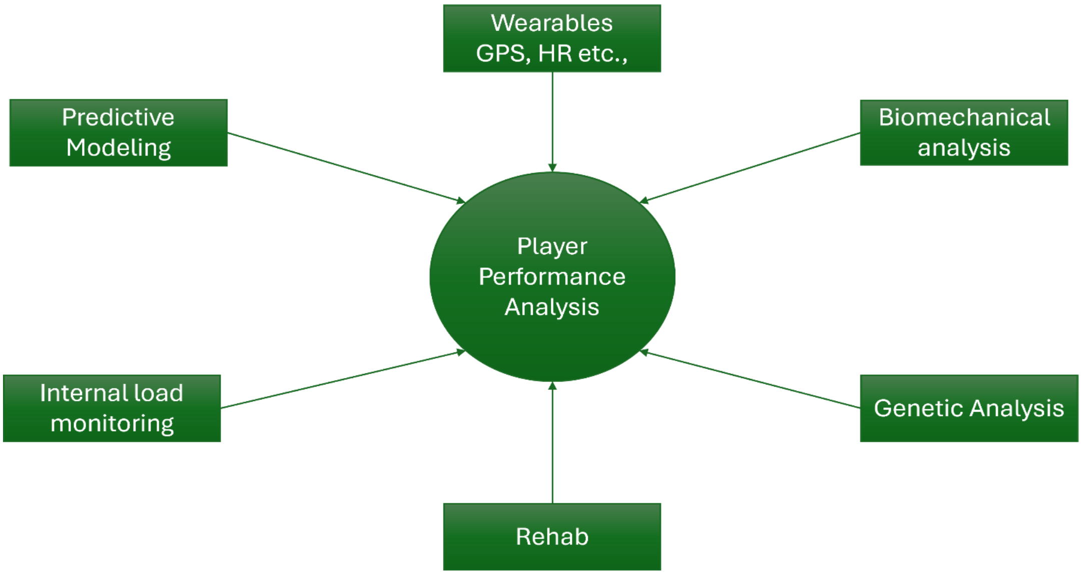
1.5.1 Sport/technical side – Player physical performance

- **Player health and fitness monitoring**
  - Data science is crucial in monitoring and optimizing player health and fitness. Sophisticated wearable technologies, such as GPS trackers and biometric sensors, capture granular data on every movement and exertion of players during training sessions and matches.
  - This wealth of data allows sports scientists and performance analysts to develop personalized training programs that maximize each player's potential while reducing the risk of injuries. By identifying patterns and insights from the data, teams can fine-tune their training methodologies to enhance overall player fitness and resilience.

Examples of player performance

- **Wearable technology**: devices like GPS trackers and heart rate monitors collect data on players' movements, physical exertion and biometrics. This data helps understand player fatigue, prevent injuries and optimize training.
- **Internal load monitoring:** this monitors sleep, strain, and recovery using heart rate data to optimize athlete performance.
- **Predictive modeling**: data on player workload, movement patterns and health metrics predict injury risks, allowing teams to adjust training loads or rest players proactively.
- **Rehabilitation monitoring**: data science tracks recovery progress, ensuring players return to peak conditions safely and efficiently.
- **Biomechanical analysis**: motion capture technology combined with data science analyzes athletes' movements in detail, identifying inefficiencies or incorrect techniques that could lead to injury or reduced performance.
- **Genetic analysis**: genetic data predicts an athlete's predisposition to specific injuries or recovery potential, allowing for more personalized training and medical care.

Figure 5. Player performance analysis



Source: own elaboration.

1.5.2 Player recruitment, team strategy and tactics

- **Team performance analysis and talent scouting**
  - Technical data, including metrics on passes, shots, hits, runs and other in-game actions, provides valuable insights into team and individual player performance. Advanced analytics enable teams to rigorously analyze their strengths and weaknesses and those of their opponents.
  - This data-driven approach to performance analysis helps teams refine their strategies and tactics and is a powerful tool for talent identification and recruitment. By analyzing data across a large player pool, teams can more effectively scout and acquire players with the desired skills and characteristics to fit their long-term vision.
- **Player development in academies**
  - Data science also transforms how sports clubs approach player development in youth academies. By closely monitoring the progression and trajectory of young players, coaches and analysts can identify areas for improvement, tailor training programs and maximize the potential of promising talent.
  - The use of data-driven insights in player development helps clubs cultivate a steady pipeline of homegrown talent, ensuring a consistent flow of high-quality players to the first team.
- **Examples of player recruitment and team strategy**

- **Scouting and recruitment**: talent identification and data analytics help scout and identify potential talent by analyzing vast amounts of data, including performance metrics, physical attributes and social media presence.

- **Game strategy optimization**: data science models analyze opponent tendencies, player matchups and historical game data to inform game plans and in-game decisions.

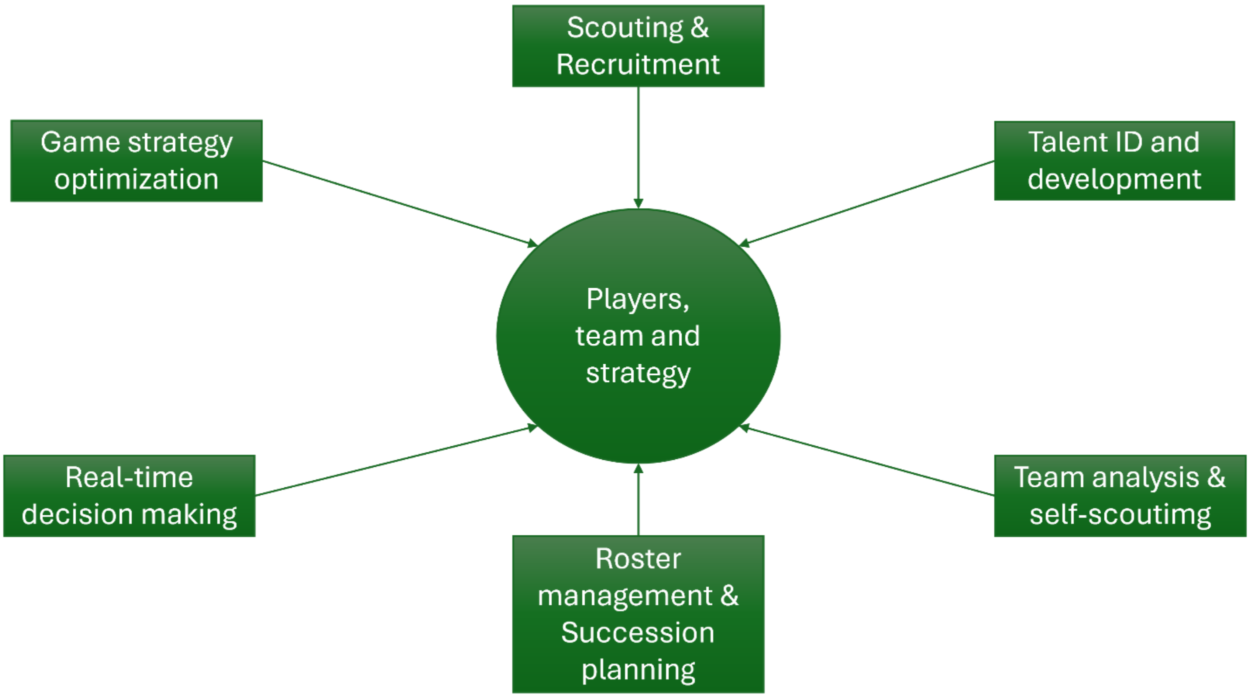
- **Real-time decision making**: advanced analytics give coaches insights during games, such as which lineup combinations work best or when to call timeouts.

- **Talent identification and development programs**: data analytics track the progress of young athletes over time, helping to identify and nurture talent from an early age. This can include analyzing growth patterns, skill development, and competitive performance.

- **Roster management and succession planning:** optimizing player salary budgets and contract extensions based on aging curve models.

- **Team analysis and self-scouting:** use data to analyze the team's performances, understand the team's strengths and identify areas for improvement.

Figure 6. Players, team and strategy



Source: own elaboration.

1.5.3 Business performance analysis

- **Fan engagement and experience optimization**
  - On the business side of sports, data plays a crucial role in understanding and optimizing the fan experience. By analyzing data on viewership patterns, fan engagement and spending habits, clubs can gain valuable insights into the preferences and behaviors of their supporters.
  - This data-driven approach enables clubs to enhance the fan experience, whether it's through targeted marketing campaigns, personalized content delivery, or optimized ticket pricing strategies. By better understanding and catering to their fans, clubs can make deeper connections and increase loyalty. This can also help them attract new fans.
- **Revenue generation and optimization**
  - In addition to improving the fan experience, data science also empowers sports organizations to optimize their revenue streams. By studying historical data on ticket sales, merchandising, sponsorships and other revenue sources, clubs can develop pricing models and strategies that maximize their financial returns.
  - This data-driven approach to revenue management allows clubs to make informed decisions about ticket pricing, merchandise offerings, sponsorship packages and other commercial activities, ultimately boosting their overall profitability and financial sustainability.

1.5.4 Data collection business

- The proliferation of sports data has led to a thriving ecosystem of data companies specializing in collecting, analyzing and providing insights to teams, leagues and media outlets. Prominent players in this space include Stats Perform, Statsbomb, Genius Sports (Second Spectrum), Impect, Wyscout/HUDL, ChyronHego/STI and Hawkeye (Hudl Statsbomb, 2022).
- These data companies leverage cutting-edge technologies, including artificial intelligence and computer vision, to automate and enhance the data collection and analysis processes. This shift from manual to automated data collection has significantly increased the volume, accuracy and

granularity of sports data available to industry stakeholders, enabling them to make more informed, data-driven decisions.

In the future, advances in AI and improved data quality and fidelity will result in more automated tasks and functions.

Unit 1.6 Ethical concerns in data and data science

With a lot of data comes some ethical concerns surrounding the usage of the data.

- Data privacy concerns: the collection, storage and use of large datasets can raise privacy concerns, mainly if the data includes personal or sensitive information about individuals. This puts a lot of emphasis on how the data is protected as well as the governance of who has access to the data.

- Bias in the algorithms: data science algorithms and models can inadvertently encode biases in the training data and the assumptions built into the algorithms. This can lead to unfair outcomes.

  Example: a 2018 study by the MIT Media Lab (Buolamwini, & Gebru, 2018), found that three commercial facial recognition systems had error rates of just 0.8 % for light-skinned men, but error rates of up to 34.7 % for dark-skinned women. This is because the training data over-represented lighter skin tones and male faces, making the algorithms more accurate at recognizing those groups.

  Football xG models were built using men's football data in sports. Women’s football is different. Using the xG model trained on men's football data to calculate xG in the women's game is fraught with issues and inaccuracies.

- Transparency and explainability: advanced data science techniques like deep learning produce accurate results. However, these models are opaque regarding what the model does to the inputs to generate the outputs. Explainability to decision-makers is very critical to successfully implementing a data-driven and evidence-based approach to decision-making.

- Data rights and consent to use: this applies primarily to user/individual data. There is a lot of ambiguity and gray around how someone's data can be used, who controls it, and whether proper consent has been obtained.

- Misuse of algorithms: data science algorithms used for beneficial purposes can also be misused for harmful endeavors like manipulation and exploitation.

As a data science professional, how can you ensure that you address the ethical concerns?

The best way to ensure you address the ethical concerns in data science listed above is through education, implementation, collaboration and communication.

- Educate yourself about the codes of conduct relevant to data science, like the IEEE ethically aligned design (IEEE Global Initiave. 2018).
- Prioritize data privacy. Follow and implement the latest data governance policies and security best practices.
- Audit and test models for possible situations of algorithmic bias.
- Document your work and the models as much as possible to ensure transparency and explainability.
- Be open to feedback and communication.
- Learn and improve continuously as new developments occur.

Suggested bibliography

Herbold, S. (2023). *Introduction to Data Science* (1st ed.). Springer.

Shang, L., Whitby, S., & Dando, M. (2021). Strengthening the biological and toxin weapons convention after COVID-19: Reaching agreement on a code of conduct and biological security education at the 2021 9th review conference. https://core.ac.uk/download/354515967.pdf

The Placement Team. (2023, October 6). 10 Reasons Why Data Science is Your Gateway to a Successful Future in Tech. *SkillUp Online*. https://skillup.online/blog/10-reasons-why-data-science-is-your-gateway-to-a-successful-future-in-tech/

References

Brocke, J. v. & Fay, M- & Böhm, M- & Haltenhof, V. (2017). Creating a Market Analytics Tool that Marketers LOVE to Use: A Case of Digital Transformation at Beiersdorf. In Oswald, G. & Kleinemeier, M. (Eds.), *Shaping the digital enterprise* (pp. 197-217). Springer. DOI: 10.1007/978-3-319-40967-2_10.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 8(1), 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Hudl Statsbomb. (2022, November 9). Analytics and Modelling in Women's Football. https://statsbomb.com/articles/soccer/analytics-and-modelling-in-womens-football/

IEEE Global Initiave. (2018). *Ethically Aligned Design*. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf