uc3m | Universidad Carlos III de Madrid

Master Degree in...

Academic Year (e.g., 2018-2019)

Master Thesis

Data Analytics in Football: Pitch Control and Beyond

Author's complete name

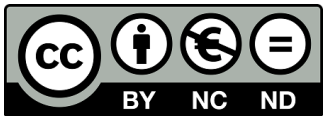
1st Tutor complete name

2nd Tutor complete name

Place and date

AVOID PLAGIARISM

The University uses the **Turnitin Feedback Studio** program within the Aula Global for the delivery of student work. This program compares the originality of the work delivered by each student with millions of electronic resources and detects those parts of the text that are copied and pasted. Plagiarizing in a TFM is considered a **Serious Misconduct**, and may result in permanent expulsion from the University.



[Include this code in case you want your Master Thesis published in Open Access University Repository]

CONTENTS

DEDICATION

LIST OF FIGURES

1. INTRODUCTION

2. DATA ANALYTICS IN FOOTBALL

The digital revolution is currently one of the most significant challenges of our time, altering numerous aspects of society. Football, in particular, has also been influenced by this transformation. Technological advancements and digitalization have resulted in a swift upsurge in the number of measuring devices, data collection and volumes of data. The leading data companies worldwide, including IBM, Intel, SAP and Microsoft, are vying for superior data analytics tools and leveraging sports as an example domain to showcase their products and brand power [Footballytics, 2021].

The practice of data analytics in football has a long history, dating back to the post-World War II era, when data collection and analysis was undertaken manually using pencil and paper [Footballytics, 2021]. It wasn't until Moneyball was published in 2003 that significant progress began to emerge: The book, "The Art of Winning an Unfair Game" introduced sports analytics to a broader audience. It illustrated the use of data analytics in identifying undervalued players and constructing a successful team. Since then, data analytics has become an integral component of sport, football inclusive [Footballytics, 2021].

One of the best examples of data analytics being applied to sports is basketball. Teams use data to analyze player performance, identify strengths and weaknesses, and develop strategies to win games [Sarlis and Tjortjis, 2020]. They use in-memory analytics, visualization, the cloud, mobility, camera footage, and sensors to transform their game. This performance analyses are of vital importance to a team, aiming to reduce expenditure, enhance team worth and refine processes across all levels and segments of operations. The German Football Association (DFB) and the National Basketball Association (NBA) are two unique cases of digital transformation from the sports world. Successful teams turn player performance data into action and gain a competitive advantage.

Over the last years, football analytics has gained significant popularity, aiming to delve deeper into the game by utilizing advanced data analysis techniques to optimize team and player performance. This chapter examines the various areas of football where data can be used for analysis, alongside the commonly found data types within this industry.

3. STATISTICAL ANALYSIS IN FOOTBALL

When discussing sports analytics in football, the first metric that often springs to mind is the Expected Goals (xG) ratio. This statistical indicator is a predictive Machine Learning (ML) model used to assess the likelihood of scoring for every shot made in the game. In the context of each shot, the xG model computes the scoring probability, leveraging a set of event parameters.

Wyscout xG model, for example, encompass the shot's spatial coordinates, the assisting player's position, the striking player's use of foot or head, the type of assist involved, the occurrence of a dribble by either a field player or the goalkeeper immediately preceding the shot, whether the shot arises from a set piece, whether it transpires during a counterattack or in a transitional phase of play, and the subjective assessment of shot danger as determined by a designated tagger. The amalgamation of these parameters serves as the foundation for training the xG model using historical Wyscout data, culminating in the prediction of the likelihood of a given shot resulting in a goal [Wyscout, 2023].

The probabilities range from 0 to 1. Thus, a shot with an xG value of 0.1 has a 10% chance of being scored. Penalties have a fixed xG value of 0.76.

Fig. 3.1 a visual representation of the cumulative development of expected goals (xG) during the Eibar - Malaga match, which took place on January 15th, 2023 in Spain's second division. Each data point on the graph corresponds to a shot made by both teams over the course of the game, offering a comprehensive overview of the evolving scoring opportunities and outcomes throughout the duration of the game.

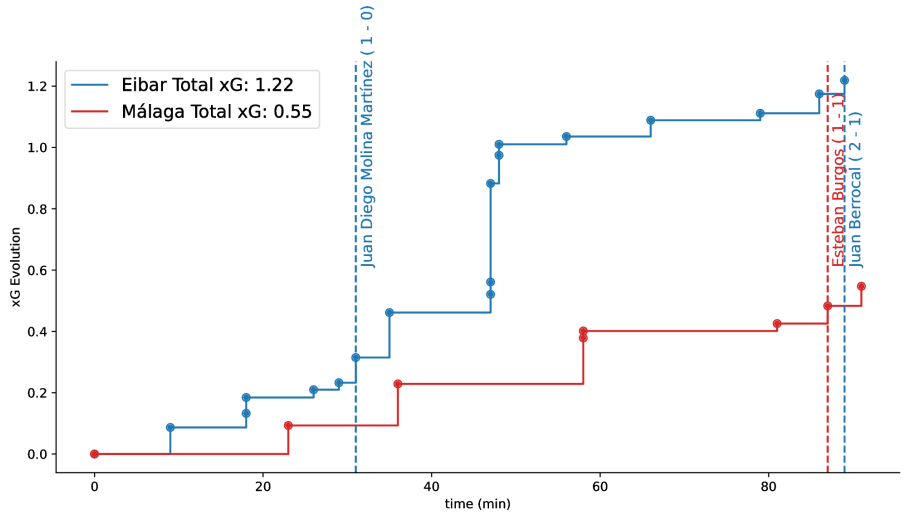


Fig. 3.1. Cumulative development of expected goals (xG) during the Eibar-Malaga match, held on January 15th in Spain's second division. Each point denotes a shot made by both teams throughout the game. Vertical dashed lines indicate the goal scored, displaying the player and the corresponding score at that specific moment of the match.

4. BIBLIOGRAPHY

Footballytics. Data analytics in football, 2021. URL https://www.footballytics.ch/post/data-analytics-in-football.

V. Sarlis and C. Tjortjis. Sports analytics — evaluation of basketball players and team performance. *Information Systems*, 93:101562, 2020. ISSN 0306-4379. DOI: https://doi.org/10.1016/j.is.2020.101562. URL https://www.sciencedirect.com/science/article/pii/S0306437920300557.

Wyscout. Wyscout data glossary, 2023. URL https://dataglossary.wyscout.com.