

PREDICTING INJURY RISK IN ELITE FOOTBALL

SPORTS ANALYTICS PROJECT

THE ANALYTICS IV:

- Thomas Renwick
- Santiago Ruiz
- Santiago Botero
- Sebastian de Wind



INTRODUCTION

- Data analytics in sports has transformed player management and training optimization.
- However, predicting injuries using integrated performance data remains underexplored.
- The project aims to use traditional machine learning to predict the likelihood of injury in football players.



MOTIVATION & SIGNIFICANCE

Player injuries in elite football lead to significant financial losses and disrupted team performance.

Developing an injury prediction model could:

- Help teams to personalize training loads.
- Reduce injury incidence through proactive interventions.
- Optimize player availability and performance over the season.





DATA SOURCES AND PROPOSED METHODS

Data Description

- **GPS Data:** Daily player movement and workload metrics.
- **Recovery Status Data:** Composite scores for recovery across multiple domains.
- **Physical Capability Data:** Measure baseline physical qualities.

Accessibility/Limitations

- All datasets are provided through the publicly available Chelsea FC Performance Insights GitHub repository.
- The data pertains to a single player.
- No injury labels.

Machine Learning Model

- Data Cleaning/Feature Engineering.
- Logistic Regression, Random Forest and XG Boost.
- Recall, Precision, AUC-ROC, F1 Score.

EXPECTED OUTCOMES AND CHALLENGES

ANTICIPATED RESULTS

- To be able to predict the probability of a player sustaining an injury on any given day.
- Identify critical risk factors.
- Deployed Streamlit application.

SCALABILITY

- Flexible and scalable framework.
- Application across entire football squad.

CHALLENGES

- Class imbalance.
- Biased model unless properly addressed.
- Datasets represent a single player.

