
Machine Learning in Rugby Union: Predicting and Identifying Key Performance Indicators for Professional Rugby Union Players in Match Play Based Workload.

Ren Xiangyu^{*1}, Simon Boisbluche², Kilian Philippe³, Mathieu Demy², Sami Äyrämö⁴, Shuzhe Ding⁵, and Jacques Prioux¹

¹École normale supérieure - Rennes – ENS Rennes – France

²Rugby Club de Vannes – Rugby Club de Vannes – France

³Université de Pau et des Pays de l'Adour – Université de Pau et des Pays de l'Adour [UPPA] : EA3002 – France

⁴JAMK University of Applied Sciences [Jyväskylä] – Finlande

⁵East China Normal University [Shangai] – Chine

Résumé

Rugby union is an intermittent high-intensity contact sport requiring the analysis of various training and match metrics. Time motion analysis and video analysis have enhanced the understanding of the interplay between these two factors. However, limited studies have investigated the effect of workload on key performance indicators (KPIs) during matches. In this study, data collected from the global positioning system (GPS) was used to calculate cumulative workload values over 7, 14, and 21 days prior to each game. After dimensionality reduction through principal component analysis (PCA), these workload values were employed as features, with game KPIs as target variables. Modeling was conducted using linear regression (LR), support vector regression (SVR), random forest regression (RFR), and light gradient boosting machine (LightGBM) for regression tasks. The superiority of the model was assessed by coefficient of determination (R^2), root mean square error (RMSE), and correlation coefficient (r). The findings revealed that while individual GPS metrics exhibited weak correlations with KPIs, machine learning (ML) models particularly RFR, successfully captured complex interactions and nonlinear relationships. These models achieved significantly improved predictive performance, with values ranging from 0.40 to 0.72 for certain KPIs. Using SHapley Additive exPlanations (SHAP) analysis and partial dependence plots, this study enhanced the interpretability of ML models by identifying the influence of GPS features on KPIs and exploring their underlying mechanisms. These findings offer actionable insights for workload management, emphasizing critical factors that affect player performance. **Key Words:** team sports; time-motion analysis; workload monitoring; performance optimization; rugby training; video analysis

*Intervenant