A Modern Hybrid Machine Learning Approach for the Prediction of International Football Matches

Andreas Groll[1], Gunther Schauberger[2]

[1] Department of Statistics, TU Dortmund University, Vogelpothweg 87, 44227 Dortmund, Germany
(E-mail: groll@statistik.tu-dortmund.de)

[2] Department of Sport and Health Sciences, Chair of Epidemiology, Technical University of Munich, Munich, Germany
(E-mail: gunther.schauberger@tum.de)

Abstract. Conventional approaches that analyze and predict the results of international matches in football are mostly based on the framework of Generalized Linear Models. The most frequently used type of regression models in the literature is the *Poisson model*. It has been shown that the predictive performance of such models can be improved by combining them with different regularization methods such as penalization (see, e.g., Groll and Abedieh, 2013; Groll et al., 2015).
More recently, also methods from the machine learning field such as *boosting* (Groll et al., 2018) and *random forests* (Groll et al., 2019) turned out to be very powerful in the prediction football match outcomes. Here, we propose a hybrid boosting extension based on *extreme gradient boosting* for modeling football matches. The model is fitted to match data from previous UEFA European Championships (EUROs) and based on the corresponding estimates all match outcomes of the EURO 2020 are repeatedly simulated (100,000 times), resulting in winning probabilities for all participating national teams.
Keywords: Football, UEFA European Championships, Poisson regression, Boosting, Hybrid Machine Learning.

References

1. A. Groll and J. Abedieh. Spain retains its title and sets a new record – generalized linear mixed models on European football championships. *Journal of Quantitative Analysis in Sports*, 9, 1, 51-66, 2013.
2. A. Groll, G. Schauberger, and G. Tutz. Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: an application to the FIFA World Cup 2014. *Journal of Quantitative Analysis in Sports*, 11, 2, 97-115, 2015.
3. A. Groll, T. Kneib, A. Mayr and G. Schauberger. On the Dependency of Soccer Scores – A Sparse Bivariate Poisson Model for the UEFA European Football Championship 2016, *Journal of Quantitative Analysis in Sports*, 14, 2, 65-79, 2018.
4. A. Groll, C. Ley, H. Van Eetvelde and G. Schauberger. A hybrid random forest to predict soccer matches in international tournaments, *Journal of Quantitative Analysis in Sports*, 15, 4, 271-287, 2019.