

Institut für Statistik

Vortrag

Montag, 17. Juni 2024, 17:00 Uhr

SR für Statistik (NT03098), Kopernikussgasse 24, 3. OG.

A combined machine learning approach for modeling and prediction of international football matches

ANDREAS GROLL

(Department of Statistics, TU Dortmund)

Abstract: Many approaches that analyze and predict the results of international matches in football/soccer are based on statistical models incorporating several potentially influential features with respect to a national team's sportive success, such as the bookmakers' ratings or the FIFA ranking. Based on all matches from the five previous UEFA EUROs 2004-2020, we combine a LASSO-penalized Poisson regression model with two alternative modeling classes, so-called random forests and extreme gradient boosting, which can be seen as mixture between machine learning and statistical modeling and are known for their high predictive power.

Moreover, we incorporate so-called hybrid predictors, i.e. features which were obtained by a separate statistical model. For different (weighted) combinations of the three modeling techniques from above, the predictive performance with regard to several goodness-of-fit measures is compared. Based on the estimates of the best performing method all match outcomes of the UEFA EUROs 2024 in Germany are repeatedly simulated (1,000,000 times), resulting in winning probabilities for all participating national teams.

Keywords: Football, UEFA EUROs, LASSO regression, Random forests, XGBoost, hybrid modeling.

H. Friedl