

Call for Papers for a Special Issue in the Springer Journal

[Machine Learning](#)

Machine Learning for Soccer

Special Issue Editors:

Werner Dubitzky¹, Daniel Berrar², Philippe Lopes³, Jesse Davis⁴

¹ Freelance scientist, Germany, werner@dubitzky.com

² College of Engineering, Shibaura Institute of Technology, Japan, dberrar@shibaura-it.ac.jp

³ Sport and Exercise Science Department, University of Evry-Val d'Essonne, and INSERM UMRS 1124, Paris Descartes University, France, philippe.lopes@univ-evry.fr

⁴ Department of Computer Science, KU Leuven, Belgium, jesse.davis@kuleuven.be

The [Machine Learning](#) journal invites submissions of original contributions to machine learning research for soccer analytics. Soccer¹ is the biggest global sport and is a fast-growing multibillion dollar industry. The annual revenue of European football clubs alone is estimated at \$27bn. Data science and analytics are being more frequently employed on both the club and national levels to improve performance, equipment, marketing, scouting, etc. In conjunction with this special issue, we will offer a machine learning challenge task where the goal is to predict the outcomes of future matches based on a data set of over 200,000 soccer matches from soccer leagues around world. This special issue solicits papers about machine learning approaches for all aspects of soccer, including (but not limited to) topics such as:

- Predicting the outcome of individual soccer matches, entire competitions and tournaments;
- Predicting the performance of individual soccer players, entire teams, and team elements;
- Aiding the design, planning and selection of competitive soccer strategies, tactics and teams;
- Developing and improving the performance of soccer players and teams;
- Evaluating and profiling young talented soccer players (scouting);
- Improving the development of young soccer players and integration of new players;
- Analyzing complex data (video, sensors, texts, etc.) from soccer players and matches;
- Designing soccer training programs that help avoid injuries and improve the recovery of injured players;
- Predicting the outcome of future soccer matches as part of this special issue's *prediction challenge* (see text below).

¹ We use term *soccer* (instead of football) as in an international context it is less ambiguous than the term football (which also relates to American football).

Prediction Challenge

This special issue includes a machine learning research challenge task based on a training data set of over 200,000 soccer matches. Challenge participants should use this data set to construct a model that predicts the outcome of a defined set of future soccer matches. This challenge presents a unique real-world machine learning prediction problem and it involves solving various machine learning tasks: data integration/fusion, feature modeling/learning, and outcome prediction. People interested in participating in the challenge should contact the guest editors immediately to receive the *training data set* and a description of the challenge. The challenge participants will receive an *updated training set* and a *prediction data set* on 22 March 2017 (which includes the fixtures of future matches with unknown outcome) and are required to submit their predictions by midnight on 30 March 2017 Central European Time (CET). Details about the data and evaluation criteria will be provided upon expression of interest (see Important Dates below).

Important Dates

The schedule for this special issue has been designed around the timing of the FIFA Soccer World Cup 2018 in Russia (starting in June 2018). The World Cup is the biggest globally televised sports event. The idea is that the special issue will be published at the beginning of 2018, thus maximizing the interest in and impact of its contributions. The overall schedule of this special issue is as follows:

Overall Schedule of Special Issue

Date	Activity/Milestone
15 May 2017	Deadline for manuscript submissions
15 August 2017	Review results (round 1)
15 October 2017	Revised papers due
15 November 2017	Final review results / final manuscript selection
Beginning 2018	Publication of special issue

If you are intending to submit a manuscript, it would be useful if you could indicate this by sending a short email to the guest editors with a tentative title and the contact details of the corresponding author.

All manuscripts are subject to the journal's submission, formatting and reviewing requirements and are to be submitted via the online submission system.

Prediction Challenge Schedule

Participants of the prediction challenge must also adhere to the following schedule. It has been designed around the weekend of 24/25 March 2017 when many leagues suspend play due to World Cup qualification matches.

Date	Activity/Milestone
Ongoing	Participants contact the special issue editors to express/register their interest in the prediction challenge. Challenge participants get access to the <i>training data set</i> including a description of the data and prediction challenge. The training data set consists of the soccer results (league, season, date of match, home team name, away team name, home score, away score) of over 200,000 past soccer matches (all from regular league play only, not from tournaments, friendlies, or international matches).
22 March 2017	Participants receive an <i>updated training set</i> and the <i>prediction data set</i> . The <i>updated training set</i> includes the results of approximately 4000 additional matches from league play in the ongoing season. The <i>prediction set</i> contains ca. 400 matches that will be played <i>after 30 March 2017</i> and the participants should predict the outcomes of these matches. Thus, at the time when the predictions are made and submitted to the special issue editors, the outcome of these matches is unknown. This is the ultimate test for predictive machine learning models.
Midnight CET 30 March 2017	Strict deadline for challenge participants to submit their predictions for the future matches in the prediction data set.
15 April 2017	Challenge participants are notified about their performance on the task and how they fared compared to other participants. Authors of the top-ranked predictions are invited to submit a full manuscript on their machine approach in line with the overall schedule of the special issue (see above). Those invited to submit their manuscript, are also requested to apply their model to all the matches in the updated training set.
