

2023 PhD Projects

Project title	Machine learning methods for predicting soccer match outcomes
Principal supervisor	Daniel Berrar
Discipline	Data Science
Research area/keywords	sports outcome prediction; domain knowledge integration; supervised learning
Suitable for	Full time applicants, Part time applicants

Project background and description

“To what extent is it possible to predict the outcome of a soccer match with machine learning methods?” This is the fundamental question that underlies this research project. At least since the late 1960s [1], predicting the outcomes of soccer matches (also known as football) has been a topic of active research. Statistical models also play a key role for odds setting in the betting industry. Recently, methods from machine learning and advanced statistics have been employed for data analysis and modeling in soccer [2,3,4]. Despite the relatively simple rules of the game, the outcome is very difficult to reliably predict, since goals are relatively rare and the margin of victory is usually relatively low for most matches. Also, goals and other game-changing events (such as red cards, injuries, penalties, etc.) often do not occur as a result of superior or inferior play by one team, but are often due to poor refereeing, weather condition, etc. On the other hand, the outcome of a match is clearly not entirely due to chance, as the stronger team is more likely to win than the weaker team. Thus, it should be possible to develop forecasting models that perform better than trivial predictions, at least in the long run. This study will investigate the state of the art of machine learning for soccer outcome prediction, develop novel ideas for incorporating domain knowledge into the modeling process, and develop novel feature engineering methods. Novel learning algorithms will be designed, implemented, and evaluated. Excellent programming skills (in either R or Python) and a solid knowledge of the principles of machine learning are required.

Background reading/references

1. Reep, C., & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society, Series A (General)*, 131(4):581-585.
2. Berrar D., Lopes P., and Dubitzky, W. (2019) Incorporating domain knowledge in machine learning for soccer outcome prediction. *Machine Learning* 108(1):97-126.
3. Dubitzky, W., Lopes P., Davis J., and Berrar D. (2019) The Open International Soccer Database for machine learning. *Machine Learning* 108(1):9-28.
4. Hubáček, O., Šourek, G. & Železný, F. (2019) Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning* 108(1):29-47.