

Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football

Nicholas Bonello

2019

MSc Computer Science (Intelligent Systems)

Supervisors: Joeran Beel, Seamus Lawless

Assistant Supervisor: Jeremy Debattista

Fantasy Premier League performance predictors commonly base their datasets on purely historic statistical data. The main problem with this approach is that external factors such as injuries, managerial decisions and other tournament match statistics can never be factored into the final predictions. In this dissertation, we propose a novel approach for predicting future player performances that automatically incorporates human feedback into the predictions. By combining standard statistical measures with data from betting-markets, social media and web articles, we aim to make more informed predictions by expanding the number of factors that can be considered before making predictions. We explore the current state-of-the-art fantasy football recommender services and compare their prediction accuracy to the different algorithms we propose. When tested on the English Premier League 2018/19 season, we found that including multiple data-sources significantly improves the performance for recommender services. Our machine-learning models trained on historic statistics alone ranked within the top 800,000 players (top 13%) out of over 6.5 million FPL players. On the other hand, machine-learning models trained using multiple data-sources, namely; historic statistics, betting-markets, news-articles and blogs, achieved a rank of 20,000 (top 0.5%). This work shows the potential that multi-stream data analytics can have in any field where both statistics and external factors affect the final outcomes.