Match outcomes prediction of six top English Premier League clubs via machine learning technique

Rabiu Muazu Musa, Anwar P.P. Abdul Majeed, Mohd Azraai Mohd Razman, and Mohd Ali Hanafiah Shaharudin

Innovative Manufacturing, Mechatronics and Sports, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, 26600 Pekan, Malaysia
amajeed@ump.edu.my

Abstract.
The English Premier League (EPL) is one of the most widely covered league in the world. The prediction of football matches, particularly EPL has received due attention over the past two decades by means of both conventional statistical and machine learning approaches. More often than not, the predictions reported in the literature have rather been dissatisfactory in forecasting the outcome of the matches. This work offers a unique approach in predicting EPL match outcomes, i.e., win, lose or draw by considering top six teams in the league namely Manchester United, Manchester City, Liverpool, Arsenal, Chelsea and Tottenham Hotspur over the span of four consecutive seasons from 2013 to 2016. Fifteen features were selected based on their relevance to the game. Six different Support Vector Machine (SVM) model variations viz. linear, quadratic, cubic, fine radial basis function (RBF), medium RBF, as well as course RBF were developed to predict the match outcomes. A five-fold cross-validation technique was employed whilst, a separate fresh data was supplied to the best model developed in evaluating the predictive efficacy of the model. It was demonstrated from the study that the linear SVM model provided an excellent prediction accuracy of 100 % on both the trained as well as untrained data. Therefore, it could be concluded that the selection of the relevant features, as well as the methodology employed, could yield a reliable prediction of top six EPL clubs match outcomes.

Keywords: Support Vector Machine, Football, Match Outcome, Feature Selection.