Abstract

Prediction of football match outcome should follow approaches that are more generalized. Hence for our project we predict outcomes of English Premier League based on the historical data of the matches and using machine learning algorithms. We gathered data from past 10 seasons and extracted features like form, goals scored and conceded, shots ratio. The computation of form feature is different from has been prevalent till now. More focus is given to gain more insight and associate a deeper and better meaning to form of a team. Basic features like shots ratio and goals scored are combined to create feature of attacking quotient. We using Logistic Regression and implement voting algorithm between Random Forest and Naive Bayes classifier to achieve accuracy between 47-50% with mean absolute error of 0.37.

References

1. Douwe Buursma; Predicting sports events from past results, University of Twente, 2011.
2. Nivard, W. & Mei, R. D.Soccer analytics: Predicting the of soccer matches. (Master
thesis: UV University of Amsterdam), 2012.
3. Ben Ulmer and Matthew Fernandez; Predicting Soccer Match results in the English
   Premier League, cs299, 2014.

**Index Terms**

Computer Science  
Information Sciences

**Keywords**

Machine learning; Data mining; Prediction system; Football; Classifiers; Knowledge discovery
database system