

MIS40970 Assignment 2: Sports Analytics

Prof Michael O'Neill
School of Business
University College Dublin
Ireland
m.oneill@ucd.ie

1 Assignment

The assignment is in the domain of sports analytics, more specifically Premier League Football. The 2016 Premier League is currently well underway, and there are many weeks of matches remaining with the winners of the League as yet to be determined.

Your task is to analyse Premier League match data provided since 2010 up to and including last weekend (6 March 2016). You should attempt to uncover interesting and potentially useful patterns and rules which might be used to determine the outcome of the remaining matches in the 2016 league.

1. Apply what you have learned about Data Mining to date (Weeks 1-7 inclusive) to undertake this analysis. In particular, we expect to see evidence of visual inspection and initial exploration of the data, classification (decision tree learning) and clustering.
2. Include plain English interpretation of your analysis.
3. You should conduct your analysis in **either R or SAS Enterprise Miner**, making sure to show evidence of your work.
4. As the matches taking place on March 19, 20, and April 2 and 3 will all have been completed by the submission deadline, in your submission,

please consider if the results of these matches are consistent with your analysis? **Treat the results of these matches (see below) as out-of-sample (or test) data.** The results will be available @ e.g., <http://www.bbc.com/sport/football/premier-league/results>

Figure 1: Premier League Fixtures (end March, early April)

Saturday 19th March 2016					
BARCLAYS PREMIER LEAGUE					
>	Everton	v	Arsenal	12:45	
>	Chelsea	v	West Ham	15:00	
>	Crystal Palace	v	Leicester	15:00	
>	Watford	v	Stoke	15:00	
>	West Brom	v	Norwich	15:00	
>	Swansea	v	Aston Villa	17:30	

Sunday 20th March 2016					
BARCLAYS PREMIER LEAGUE					
>	Newcastle	v	Sunderland	13:30	
>	Southampton	v	Liverpool	13:30	
>	Man City	v	Man Utd	16:00	
>	Tottenham	v	Bournemouth	16:00	

Saturday 2nd April 2016					
BARCLAYS PREMIER LEAGUE					
>	Aston Villa	v	Chelsea	12:45	
>	Arsenal	v	Watford	15:00	
>	Bournemouth	v	Man City	15:00	
>	Norwich	v	Newcastle	15:00	
>	Stoke	v	Swansea	15:00	
>	Sunderland	v	West Brom	15:00	
>	West Ham	v	Crystal Palace	15:00	
>	Liverpool	v	Tottenham	17:30	

Sunday 3rd April 2016					
BARCLAYS PREMIER LEAGUE					
>	Leicester	v	Southampton	13:30	
>	Man Utd	v	Everton	16:00	

2 Submission Details & Deadline

Please submit a report (pdf) through Blackboard by 17h00 Wednesday 6 April. This assignment is worth 20% of the total credit for MIS40970. **In your submission, please make sure that in each case you include any code or images of screenshots which were used to address each item below. Also, include the output (if any) generated, i.e., Show Your Work!**

3 Dataset

The datasets (one for each year) are available from Blackboard, and contains the results of Premier League matches from 2010 to 2016. The attributes are described in an associated file also available from Blackboard.