

A large, white soccer ball with black star-shaped patterns is positioned in the foreground on a vibrant green grass field. In the background, a large stadium with a curved, glass-paneled roof is visible. The sky is blue with some clouds. The stadium's seating area is filled with spectators, and various flags are hanging from the stands. A goalpost is visible in the distance on the right side of the field.

Perfect Form

An analysis for the football
Betting market

Patrick McGuinness

The football betting market is insanely big!

Estimated market sizes per annum 2016: \$billions

Football Betting: 700-1000 estimated

Global Defence: 1700

Clean water for whole of Africa: 535



Mission: How can we forecast game results based on information in league tables?

RANK		BARCLAYS PREMIER									
		Home					Away				
	NAME	P	W	D	L	F	A	W	D	L	F
1	 Man Utd	38	17	1	1	47	7	10	5	4	
2	 Chelsea	38	12	7	0	36	13	13	3	3	
3	 Arsenal	38	14	5	0	37	11	10	6	3	
4	 Liverpool	38	12	6	1	43	13	9	7		
5	 Everton	38	11	4	4	34	17	8	4		
6	 Aston Villa	38	10	3	6	34	22	6	9		
7	 Blackburn	38	8	7	4	26	19	7	6		

What are the important predictors?

Results:

Goal Difference proved the crucial measure

Making win/loss/draw percentages redundant.

“Tain’t what you do, it’s the way that you do it” Trummy Young & Sy Oliver, 1939

... that’s what predicts results

How far back should one go?

Current season: important!

Previous season: also important!

Season before that: much less important

What algorithm?

Pin-sticking / Lobster in tank - possibly better than an overconfident punter

Decision Tree: Ridiculously simple but surprisingly robust.

Random Forest: Disappointing. Worse than decision tree.

Logistic Regression: Pretty good!

Neural Network: Pretty good!

Decision Tree + Logistic Regression: Pretty good!



The winning model

Logistic Regression!

Simple, but an effective tool for this problem.

$P(\text{home win}) = \text{logistic}(\alpha)$ where $\alpha = -0.1396 + 0.9157 \cdot \text{GDrdiffLS} + 0.6622 \cdot a \cdot c + 0.1683 \cdot b$

$P(\text{away win}) = \text{logistic}(\alpha)$ where $\alpha = -1.1175 + 0.9751 \cdot \text{GDrdiffLS} - 0.6799 \cdot a \cdot c - 0.1419 \cdot b$

$P(\text{draw}) = 1 - P(\text{home win}) - P(\text{away win})$