

Introducing logic into title-race predictions

Having spent 8 years in London I have had no choice but to develop an opinion on football... and since I am by no means an expert I often turned to statistics for help. Last year, when I started getting interested in machine learning algorithms, I decided to play around building a predictive model for the English Premier League and a couple of weeks ago I got involved in a little experiment via twitter where the aim was to provide a prediction for the remaining 13 games of the season. The prediction from the neural network model that I built can be found under the pseudonym "Teouch" in the article below. This article was written by Simon Gleave on February 12, 2016 for the website listed at the end of this briefing.

Thomas Loridan, Risk Frontiers

During my trip to London this week, I picked up a copy of The Times and discovered that the newspaper's resident stats man Bill Edgar had predicted the remaining 13 matches of the Premier League season for the current top-five using a simple model. From this process, he named Manchester City as this season's champions, ahead of Leicester City on goal difference. Comparing his projection to a number of other models, both simple and complex, reveals a number of interesting differences. Bill's approach is more logical than most media predictions but can he do better than these models?

Times numbers man Bill Edgar has, in his own words, "had a stab at introducing logic into a title-race prediction". He has assumed that the leading five clubs will lose away to a fellow top-five team, draw away to clubs who are sixth to 13th and win all of their other fixtures.



difference will be key
By Bill Edgar

Making sense of this season's Barclays Premier League has been tricky but we've had a stab at introducing logic to a title-race prediction. In assessing the run-ins of the leading five clubs, we have assumed they will lose any matches away to a fellow top-five side, draw away to clubs who are at present sixth to thirteenth and win all their other fixtures. Arsenal's tough schedule sees them slip to fourth in our model, while Manchester City benefit from having already played away to each of their main rivals. This forecasting system leaves City joint top on points with Leicester, given that Manuel Pellegrini's side would have fared the better of the two clubs over the final 13 rounds of matches, we will assume the former team's goal difference will remain superior and thus give them the title ahead of this season's upstarts.

| Leicester: | Tottenham |
|---|-----------------------------|
| Feb 14: Arsenal (a) LOSE | Feb 14: Man City (a) LOSE |
| Feb 27: Norwich (h) WIN | Feb 28: Swansea (h) WIN |
| Mar 1: West Brom (h) WIN | Mar 2: West Ham (a) DRAW |
| Mar 5: Watford (a) DRAW | Mar 5: Arsenal (h) WIN |
| Mar 14: Newcastle (h) WIN | Mar 12: Aston Villa (a) WIN |
| Mar 19: Crystal Palace (a) DRAW | Mar 19: Bournemouth (h) WIN |
| Apr 2: Southampton (h) WIN | Apr 2: Liverpool (a) DRAW |
| Apr 9: Sunderland (a) WIN | Apr 9: Man Utd (h) WIN |
| Apr 16: West Ham (h) WIN | Apr 16: Stoke (a) DRAW |
| Apr 23: Swansea (h) WIN | Apr 23: West Brom (h) WIN |
| Apr 30: Man Utd (a) LOSE | Apr 30: Chelsea (a) DRAW |
| May 7: Everton (h) WIN | May 7: Chelsea (a) DRAW |
| May 15: Chelsea (a) DRAW | May 7: Southampton (h) WIN |
| Now 53pts | May 15: Newcastle (a) WIN |
| Final total 80pts | Now 48pts |
| Finishing position 2nd (on goal difference) | Final total 76pts |
| | Finishing position 3rd |

| Arsenal: | Man City: | Man Utd: |
|--------------------------------|---|--------------------------------|
| Feb 14: Leicester (h) WIN | Feb 14: Tottenham (h) WIN | Feb 13: Sunderland (a) WIN |
| Feb 28: Man Utd (a) LOSE | Mar 2: Liverpool (a) DRAW | Feb 28: Arsenal (h) WIN |
| Mar 2: Swansea (h) WIN | Mar 5: Aston Villa (h) WIN | Mar 2: Watford (h) WIN |
| Mar 5: Tottenham (a) LOSE | Mar 12: Norwich (a) WIN | Mar 6: West Brom (a) WIN |
| Mar 12: West Brom (h) WIN | Mar 20: Man Utd (h) WIN | Mar 12: Crystal Palace (h) WIN |
| Mar 19: Everton (a) DRAW | Apr 2: Bournemouth (a) WIN | Mar 20: Man City (a) LOSE |
| Apr 2: Watford (h) WIN | Apr 9: West Brom (h) WIN | Apr 2: Everton (h) WIN |
| Apr 9: West Ham (a) DRAW | Apr 16: Chelsea (a) DRAW | Apr 9: Tottenham (a) LOSE |
| Apr 16: Crystal Palace (h) WIN | Apr 23: Stoke (h) WIN | Apr 16: Aston Villa (h) WIN |
| Apr 23: Sunderland (a) WIN | Apr 30: Southampton (a) DRAW | Apr 23: West Ham (a) DRAW |
| Apr 30: Norwich (h) WIN | May 7: Arsenal (h) WIN | Apr 30: Leicester (h) WIN |
| May 7: Man City (a) LOSE | May 15: Swansea (a) WIN | May 7: Norwich (h) WIN |
| May 15: Aston Villa (h) WIN | TBA: | May 15: Bournemouth (h) WIN |
| Now 48pts | Now 47pts | Now 41pts |
| Final total 74pts | Final total 80pts | Final total 72pts |
| Finishing position 4th | Finishing position 1st (on goal difference) | Finishing position 5th |



briefing note
311
Feb. 2016

In order to compare Bill's forecast to other methods, I put a request out on Twitter and quickly received projections from expected goals modeller [@colintrainor](#), non-expected goals modeller [@goalprojection](#) and a 56-variable neural network approach from Australia-based [@teouchanalytics](#). I have also added the results based projection from [@euroclubindex](#) and a basic benchmark of the same results as last season for the remainder of this one ([SCoRe](#)).

| Bill Edgar | Colin Trainor | GoalProjection | Teouch | ECI | SCoRe |
|--------------|---------------|----------------|--------------|--------------|--------------|
| Man City 80 | Leicester 74 | Leicester 74 | Leicester 74 | Leicester 73 | Arsenal 77 |
| Leicester 80 | Arsenal 72 | Man City 72 | Tottenham 72 | Arsenal 73 | Leicester 73 |
| Tottenham 76 | Tottenham 71 | Arsenal 72 | Man City 71 | Man City 72 | Man City 73 |
| Arsenal 74 | Man City 70 | Tottenham 71 | Arsenal 71 | Tottenham 70 | Tottenham 70 |
| Man Utd 72 | Man Utd 62 | Man Utd 62 | Man Utd 62 | Man Utd 64 | Man Utd 67 |

The accuracy of these predictions will be assessed at the end of the season using the mean absolute errors of both positions and points. It is already interesting to note some of the differences though.

Bill Edgar's approach is the only one to favour Manchester City as champions at this stage. The Times man's points totals are consistently above those of the other models, even the simple SCoRe which just assumes that the rest of the season will be the same as 2014/2015. The exception is Arsenal who Edgar's method expects to do worse than last season on the remaining fixtures. If Arsenal manage to take the same number of points in their remaining matches as they did last season, the 77 points they accrue will probably be enough to win the league.

The four more complex models all favour Leicester City to win the title now, albeit very marginally in the case of the Euro Club Index. They differ on the other three top four spots though with Arsenal, Manchester City and Tottenham Hotspur named in each position.

Across all six different approaches, Arsenal and Manchester City appear in every possible top-4 position, Leicester City are never outside the top two and Tottenham Hotspur are never named as title winners.

There is only position which everyone agrees on, a fifth place for Manchester United, so the current top four look set to play in next season's Champions League. Hardly surprising given that, according to the Euro Club Index, only Aston Villa (20th) and Sunderland (19th) are more likely to finish in their current positions than United.

Thanks to:

Colin Trainor – his model is not public but you can google Expected goals to find plenty of information from others.

Goal Projection – you can read about his method [here](#) where there is also a link to his rating system.

Teouch Analytics – the methodology is described [here](#).

Euro Club Index – the website is [here](#) and also contains a methodology page.

<https://scoreboardjournalism.wordpress.com/2016/02/12/the-times-bill-edgar-vs-other-models/>