

Progress report

Title

Predictive Model for English Premier League Match Outcomes

Progress scope update

Original data source 1: match odds is substituted with team recent performance data (wins in recent 5 games etc.). The performance data can reflect whether a team is in good shape in an objective way, while the odds provided by sports books are somehow subjective as sometimes the change in odds might only aim to balance the market etc.

Data sources

understat.com: web scrape

football-data.co.uk: CSV

Issues

Incomplete sections: readme, results.ipynb, test section.

The accuracy rate is a little over 50%, which is acceptable because it is significantly better than randomly guess (33.33%). However, this number might still be improved since the 50% accuracy rate must have included a large portion of correct predictions of matches between two teams with huge gaps.

Next steps:

1. Include ROI (Return over investment) to better understand how well the model works.
2. Include an ELO system to better reflect the strength of each team. (might need much more data to build the system)
3. Predict the xG difference instead of winning, draw or loss.
(xG is a more reliable and stable calculation of goals in the long term, might eliminate the component of luck in each game)