Chelsea group project reflection

Strength and weaknesses of our approach, Yash Karle

Mathematical Modelling of Football, Autumn 2022

1. Contribution

Given I've been a Chelsea fan since I was 10, naturally I was able to provide context to the group. I fed the group with references and ideas from time to time which acted as constraints for staying focused and better manage time and effort. I also took up the role of being the coordinator and reviewed work across groups to make sure that, within our respective groups, we are aligned and aiming towards a coherent analysis. Within the club performance subgroup, I acted as the tech lead owing to my programming and data science expertise and as a lead it was really reassuring to see progress made by our team. I built the baseline simulation model that again helped us refine our end objective. I organised regular follow-up meetings both within club performance and the whole team to make sure we're always bouncing off each other's ideas and staying aligned. Even though I was doing most of the modelling for the club performance team, I made sure everyone's opinions are heard and had tasks evenly distributed outside of modelling. I relied on my teammates to deliver detailed deep-dive on team KPIs and ensured that this was actioned upon by the set pieces and tracking data scouting teams. Given I am not based in Sweden/Uppsala, I was able to make the most of Slack for communication purposes. Overall, it was a rewarding feeling to help and review other's work and progress together towards a common goal.

2. Club performance review

Reflecting back on what our core objective was within the club performance group, in a typical football club setting, we set out to review team performances and variance in playing styles across the league from the last few seasons. This serves a two-fold purpose in allowing us to benchmark Chelsea against other opponents and model different attributes of team performance metrics to predict outcomes for the future seasons. Throughout, we kept thinking of a real-world situation wherein we were tasked to feed insights into other departments for improving the player recruitment as well as coaching. The main contribution of the group has to be the simulation model we built for predicting future outcomes. It served the purpose of setting realistic expectations for the upcoming season by factoring in alternative scenarios which is really a powerful way to set targets in a way that is consistent and easy to communicate to layman audiences throughout different departments within a football club. We were able to simulate expected gains based on simple and intuitive feature engineering within the model. Not only that the model can be used throughout the season to analyse current trends and even towards the latter stages of a season to act as a feedback loop for comparing the predicted versus actual outcomes.

3. Modelling for football

When we think of football and modelling we naturally think of data! Remembering history, the birth of modelling in football can be traced back to the betting industry where they had to develop metrics and models to predict footballing outcomes to gain an edge over the average crowds. They managed to factor in live market dynamics as "wisdom of crowds" into their models and even with the present day advances in machine learning and deep learning their models still manage to

beat the advanced models. Speaking of data, there are so many varied data sources available at varied level of granularity and context. Since the adaption of xG it has now become almost an universal metric. Identifying metrics that give you marginal gains over your opponents in identifying the signal has become football's biggest source of monetizing in recent years. Understanding of football is a complex task and subjective to for what purpose, what aspect of it and why we want to understand it. Applications of modelling and analytics within football across multiple industries from betting, media, football clubs across all levels to professional sports analytics organisations and research organisations. Not only that, even within football clubs data and modelling, done intuitively and practically, have the potential to drive competitive edge across departments like scouting, coaching, tactical analysis and even business operations. Football is such a dynamic and complex sport to analyse and model for, as modellers and analysts we need to stay modest and state caveats and limitations of our approaches. We must consume the outputs of these models with utmost caution, taking a "human in the loop" approach wherever possible and always stay open for challenges and rely on experimentation to mitigate risks. In that regard, I believe the industry has much to be gained from being transparent and publishing more work in the rapidly growing open source community.