ORIE 4741 Project Proposal

Tomas Alvarez (ta352), Benjamin Yeh (by253), Rafael Chaves (rvc29)

1 Introduction

Our project is concerned with the market making of sports betting. Many companies attempt to quantify the odds of certain sporting events and we'd like to see if we can create useful predictions using available sports data. We may venture into multiple sports, but since soccer seems to be the one with the most data available, it will be our inital focus.

2 Questions

- Can we develop a model that accurately predicts the outcome of sports games?
- Can we use such a model to create a sport betting portfolio that outperforms a random portfolio?

3 Why do we care?

From the perspective of the company taking bets it is important that odds are calculated correctly to minimize risk. If odds of certain events are calculated inaccurately, it could lead to major losses for the company. If a model can be made that accurately beats the house, then a company should reconsider the methods they use to calculate odds.

4 Why are we likely to succeed?

The collected data sets offer deep quantitative insight into matches and teams, such as the number of matches played, wins, losses, and league points scored by one team. Further, team information spans over about a decade, providing a large repository of data to develop and train over. At the very least, a model trained on large amounts of historical data should be able to perform better than a random portfolio, or than the intuition of an average sports bettor.

5 Datasets

We aggregate soccer data from the following sources:

- https://fbref.com/en: Provides data on every game and their outcomes. Also provides a variety of team-based and player-based statistics.
- https://www.football-data.co.uk/data.php: Games and the odds assigned to each game by various betting companies. With historical odds we can test how well a suggested portfolio would have performed.
- http://clubelo.com/: An API that gives an ELO rating for a club at any point in time. This could be used a feature to predict the probability of a certain team winning.