Football Team Strength Models

Mathieu Tischer and Maximilian Prinz

UNIVERSITÄT DES SAARLANDES

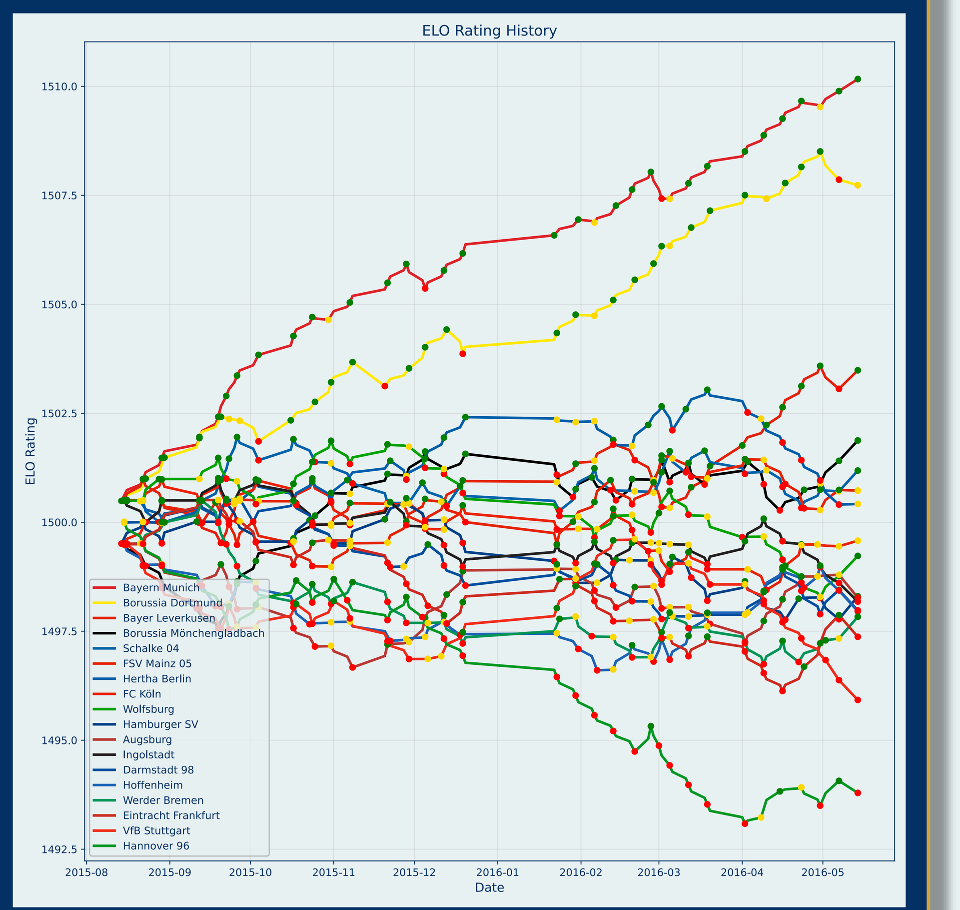
Data Science and Artificial Intelligence
Chair for Sports Analytics

Introduction and Motivation

Our aim is to compare traditional team strength metrics with more advanced Machine Learning models for the purpose of Football match outcome predictions. We use a Dataset consisting of match data with detailed description of every event in the match.

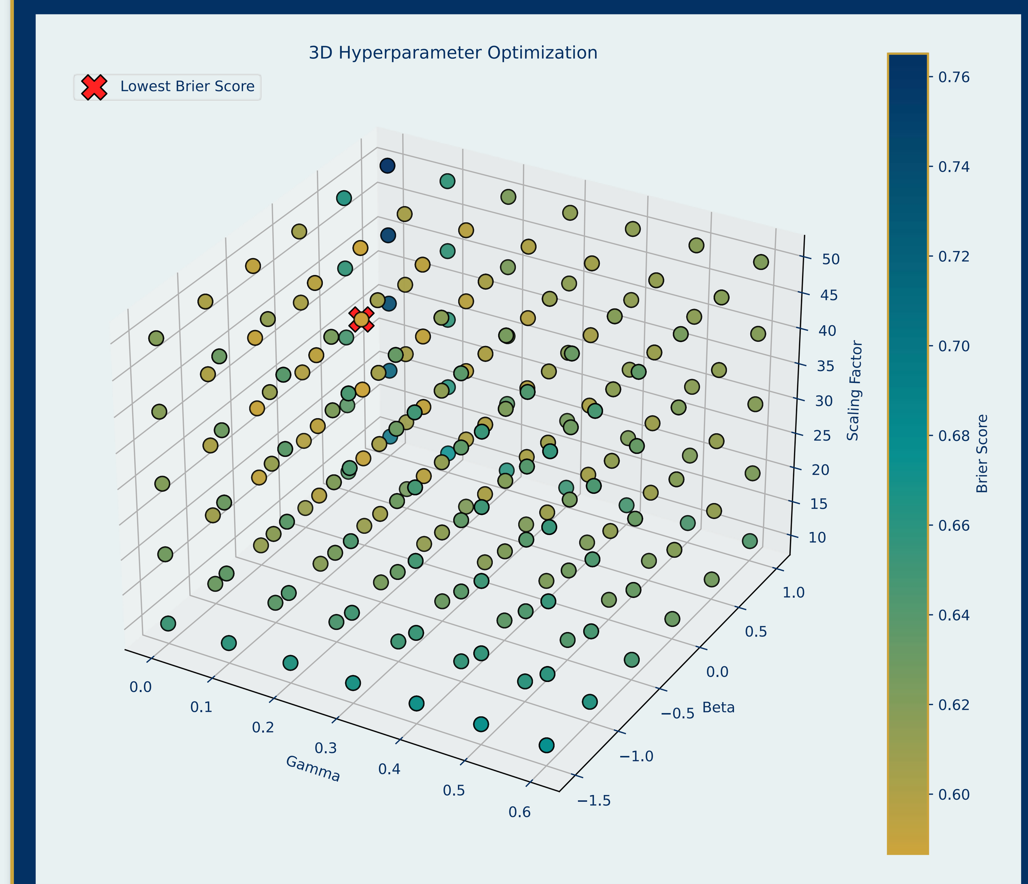
ELO Model

- We apply the ELO Formula once to the matches of the 15/16 Season (seen to the right)
- Final ELO standings are used as Starting Point for the real ELO Model
- Real ELO Model uses Three-Way Probability model, and runs through the 15/16 Season again while making predictions and keeping track of the Brier Loss
- Brier Score takes value between 0 (perfect prediction) and 2



We then used a Grid Seach technique to find the optimal Parameters
$SF = 40.0, R_0 = 1500, \beta = 0.0, \gamma = 0.1$

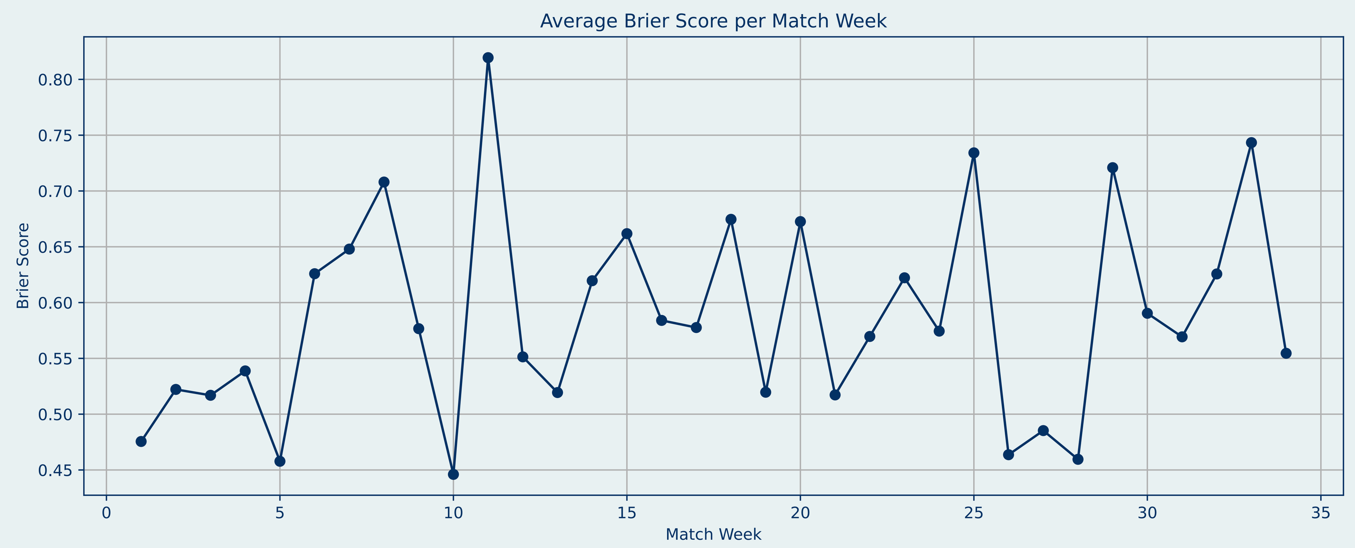
$$E_H = e^{\alpha(R_H - R_A)} + e^{\beta - \gamma|R_H - R_A|}$$
$$R_H^{n+1} = R_H^n + K(S - E_H)$$



- ❖ K determines how quickly ELO ratings change
- ❖ R_0 is the Starting ELO
- ❖ SF translates ELO differences into outcome probabilities (directly related to α)
- ❖ β controls probability mass given to draws
- ❖ γ controls how fast draw probabilities decrease
- ❖ S is the actual outcome of the match

Intermezzo

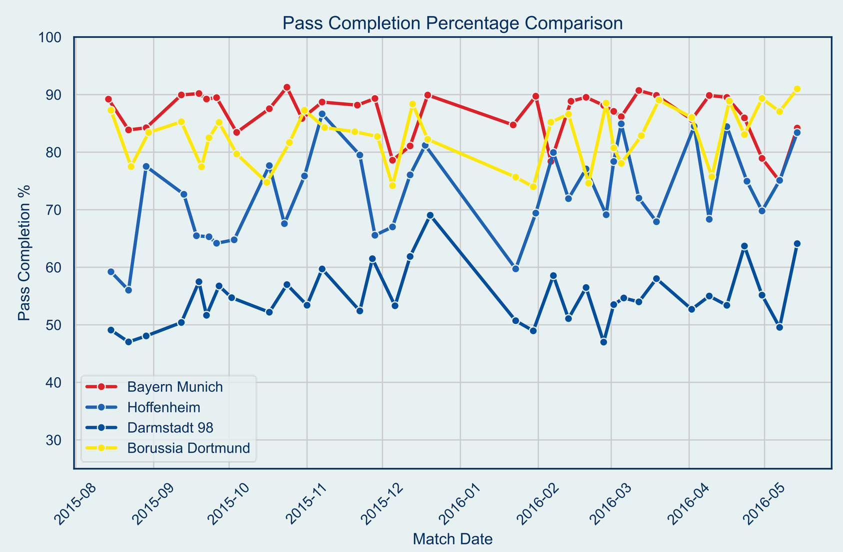
The ELO Model gives us an optimal Brier Score of 0.58 . Unfortunately, this score is still too high, and we conclude that ELO is not good for making predictions. Its' simplicity causes it to suffer from the events of draws and surprising match outcomes despite hyperparameter tuning.



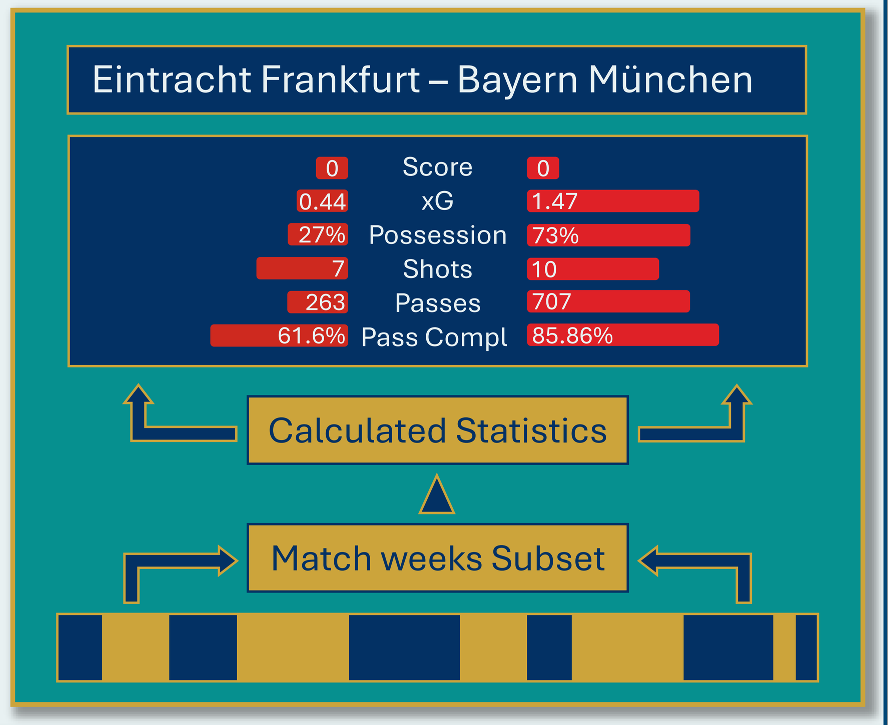
So, our next approach is to find important predictors for match outcomes and to apply them to a real Machine Learning Model.

Extracting predictors & Creating Data

- We store the predictors for each match and Team in a Dataframe
- Goal is to find a correlation between match statistics and outcomes
- Predictors include:
  - ❖ Avg. Possession
  - ❖ xG
  - ❖ Goals Scored/Conceded
  - ❖ Etc...

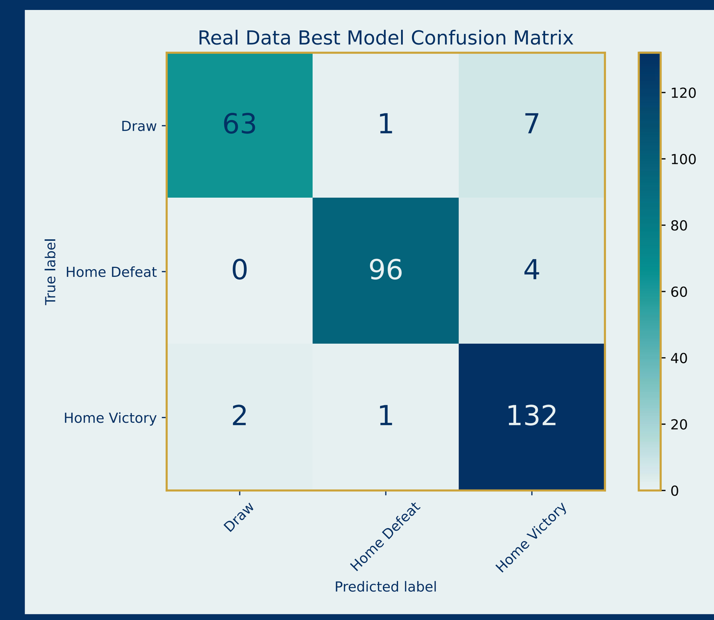


- We are working with Statsbomb's free Football Dataset
- Relatively small Dataset (306 Matches for 15/16 Bundesliga)
- We need to create Fake Data to train models based on the 2015/2016 Season
- Fake Match Label is actual outcome of the real match

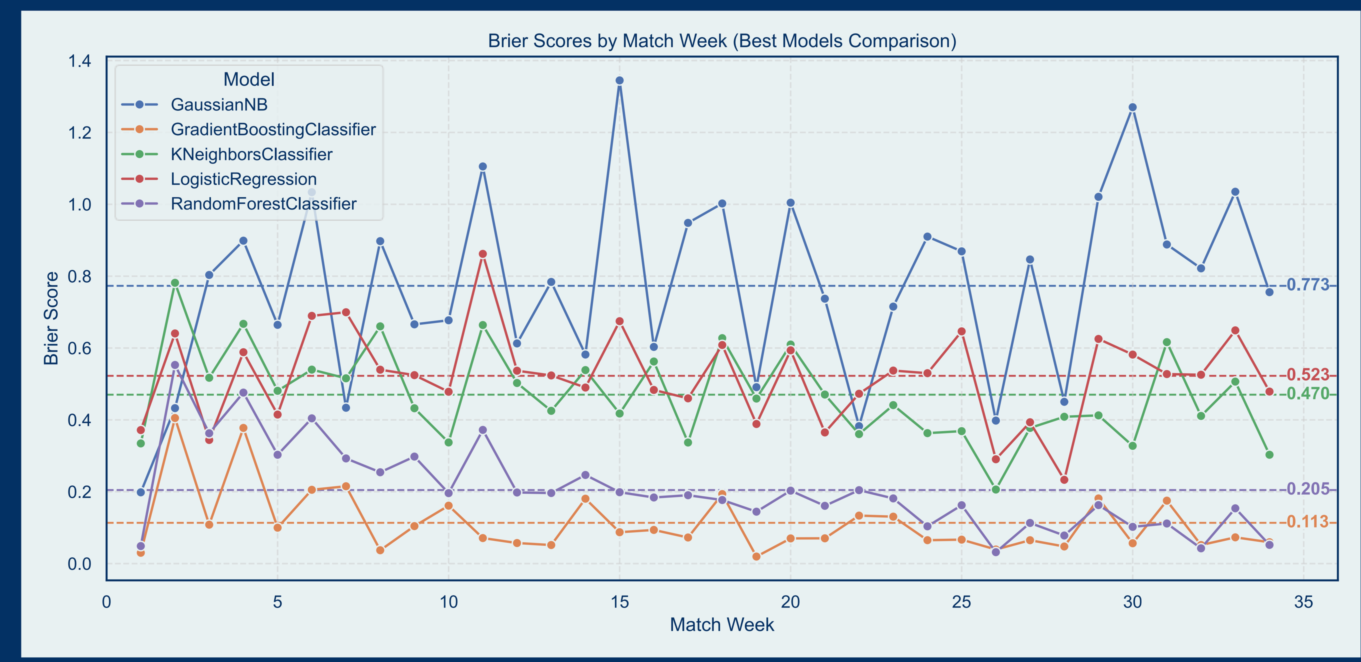


Using Machine Learning models

- We now train advanced Machine Learning models with the real and fake data
- We keep the Brier Score as a metric to quantify how good our predictions are
- The plot on the right shows the Best Model's Confusion Matrix on the real data
- We trained various Models with great variation in results



After simulating enough Bundesliga match weeks to properly train the models, we see that some models excelled while most of them got average or even poor results. The Gradient Boosting Classifier yields the best score of 0.113 .



Results and Outlook

Given that the Dataset is quite small, we had to simulate a lot of match weeks to get good results. If we only train the Gradient Boosting Classifier on one Season like we did with the ELO model, it just barely outperforms it.
On the other hand, after training for 510 simulated match weeks (i.e., 15 years of Bundesliga matches), we were able to accomplish decent results.

While good, our predictors are nowhere near perfect as some of them share a rather high correlation. So, in a future approach we plan on tweaking all our predictors, while also adding normally distributed error and other useful features to make the data generation more robust.