SFU

Predicting Soccer Games

Sagar Parikh, Chirayu Shah, Abhi Savaliya

Premier League

Objectives

Football is one of the most unpredictable sport in the history. However, Machine Learning can be used to predict these matches with favorable outcomes. We aim to use different machine learning models to predict the matches with intuitive feature engineering.

Feature Engineering

Team Ratings - To neutralize the strong effect of any one influential player, we take the average of year-wise FIFA ratings of the squad of both teams.

Winning streak (5 Matches) - The winning momentum can be an important predictor in deciding the outcome of their next game.

Winning streak (Home) – Outcomes against the same club plays an important predictor for the next game.

From above, **FIFA ratings are very crucial because of high co-relation with the result**!

Challenges

- Filling the missing data such as squads for each match for recent years was quite challenging since there was no API to directly fetch the squads provided the match details. We had to manually fill in the irregularities to make data from different sources consistent.
- Creating the new features from the existing ones which are important in predicting the outcome and tuning parameters for each models was also quite challenging.

Exploratory Data Analysis

Data Collection & Integration

- **Sources:** *EPL Soccer Games 2008-16, FIFA Ratings, Kaggle, Football-Data*
- **Integration:** *Merging inconsistent datasets*

Data Explorations & Insights

- **EDA:** *Match Outcome Dependency On Team Rating, Analyzing The Home Team Advantage, Head-to-head, Player Statistics*
- **Insights:** *46% Of The Time Home Team Wins*

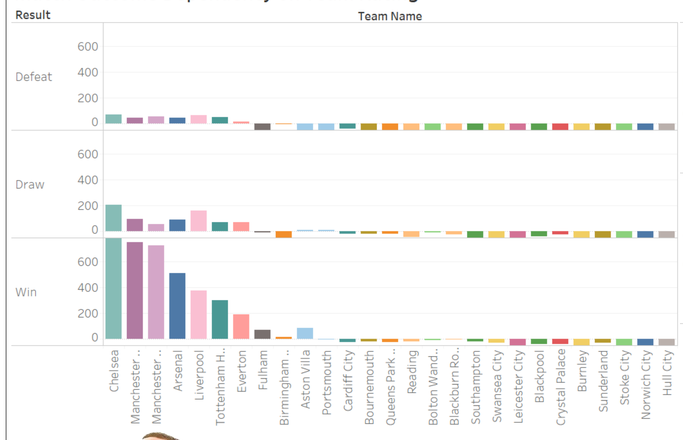
Predictive Modeling

- **Feature Engineering:** *Team Ratings, Winning Streaks Home and Away*
- **Model Selection:** *KNN-Classifier, RandomForest, Logistic Regression and Naive Bayes*

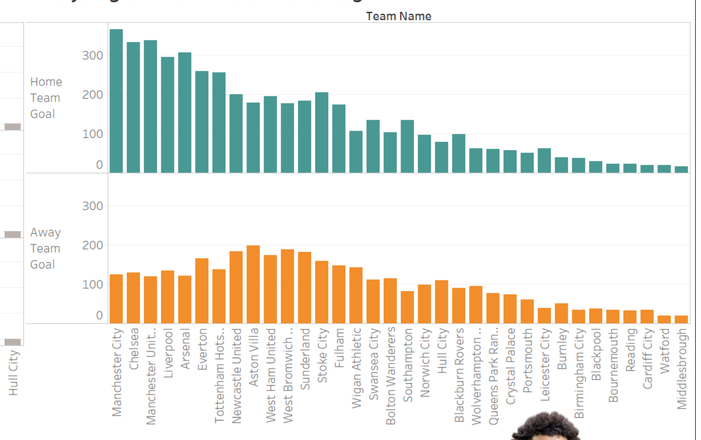
Evaluation & Visualizations

- **Evaluation:** *Maximum accuracy (61%) achieved by SVM Classifier and 54% with Naïve Bayes Classifier*

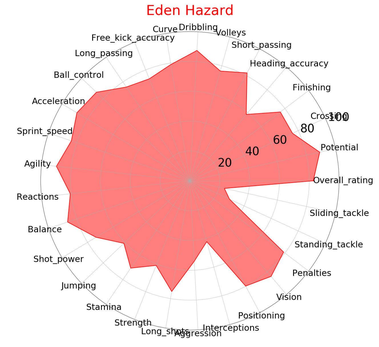
Match Outcome Dependency on Team Rating



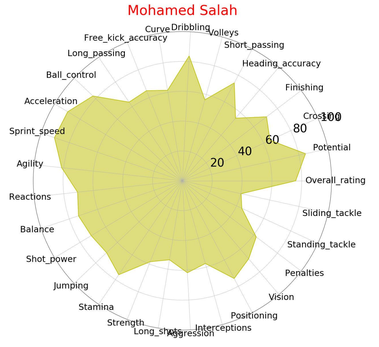
Analysing the Home Team Advantage



Eden Hazard



Mohamed Salah

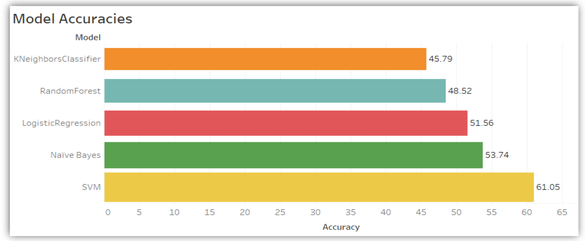


Results

We predict the outcome of the matches into Win, Draw or Defeat using different Classification models.

Classification algorithms:

KNN	**46%**	**RandomForest**	**49%**
Logistic Regression	**52%**	**Naïve Bayes**	**54%**
SVM	**61%**		



Conclusion

Predicting the matches with accuracy higher than 70% is nearly impossible. Soccer is a very dynamic sport because of un-forecasted transfers, players injuries and underdog winners. Also, it defies the common ML perception that having a huge dataset is a necessity to get adequate predictions. Team kinetics and emotions change continuously which are difficult to be transformed into features. After much exploration and feature engineering we were able to achieve an accuracy of 61%.

Future Scope

Currently, we have predicted the Football Match outcomes. However, betting is the next challenge as it involves predicting matches with higher accuracy as well predict the dynamic odds in real-time. We plan to recommend the best platform to maximize the betting profits.