Soccer Predictive Analytics
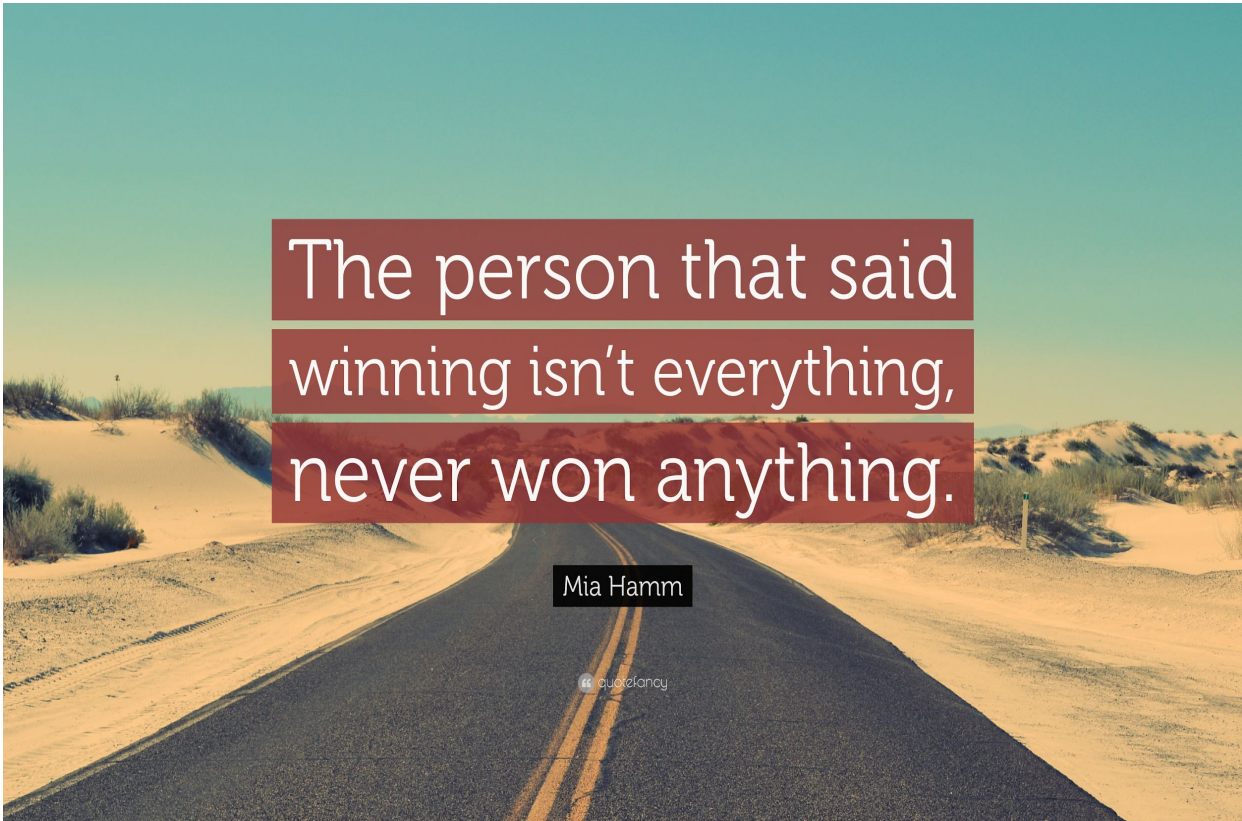
By Chizoba Obasi

Motivation

The person that said winning isn't everything, never won anything.

Mia Hamm

Motivation

- **English Premier League (EPL)**

    - Most watched soccer league

    - Most profitable soccer league

    - 212 territories

    - 643 million homes

    - Potential TV audience of 4.7 billion people

    - Revenue of € 2.2 billion/yr in domestic and international TV rights

- **High level of uncertainty, unpredictability and variability in soccer**

- **Sports betting industry - a major global financial industry**
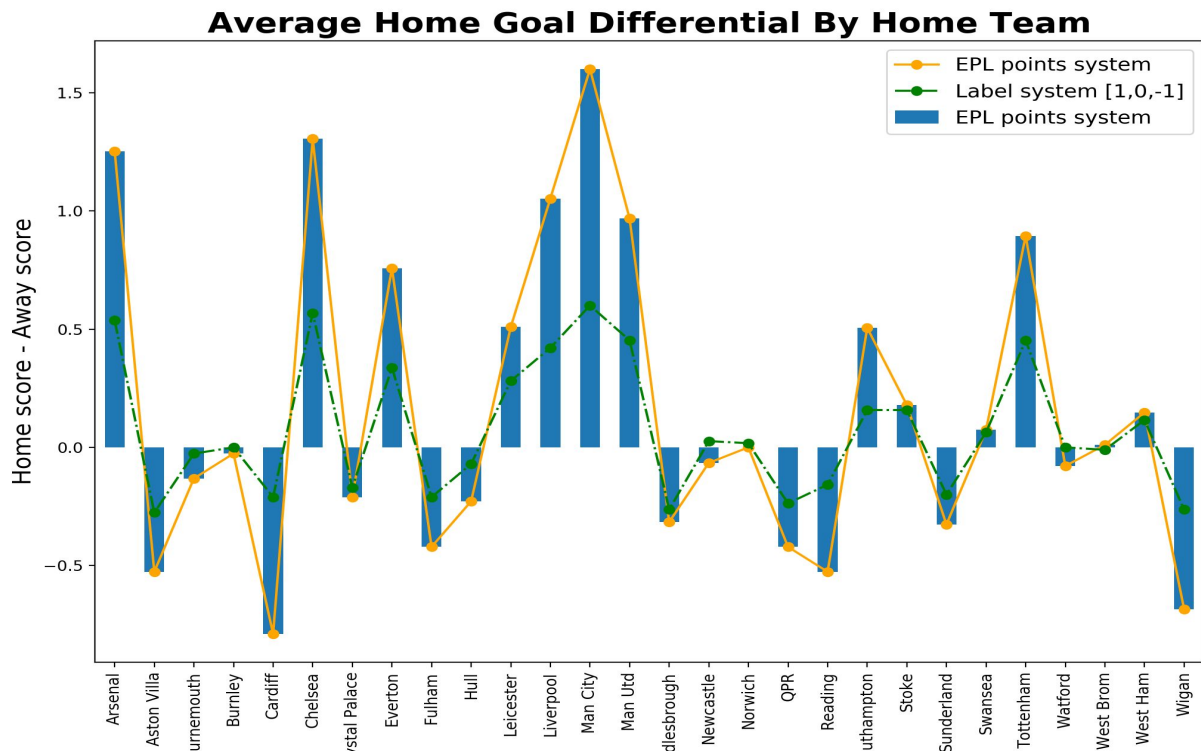
Observations - "BIG" team consistency

Point Tally per Team from 2012/13 to 2016/2017
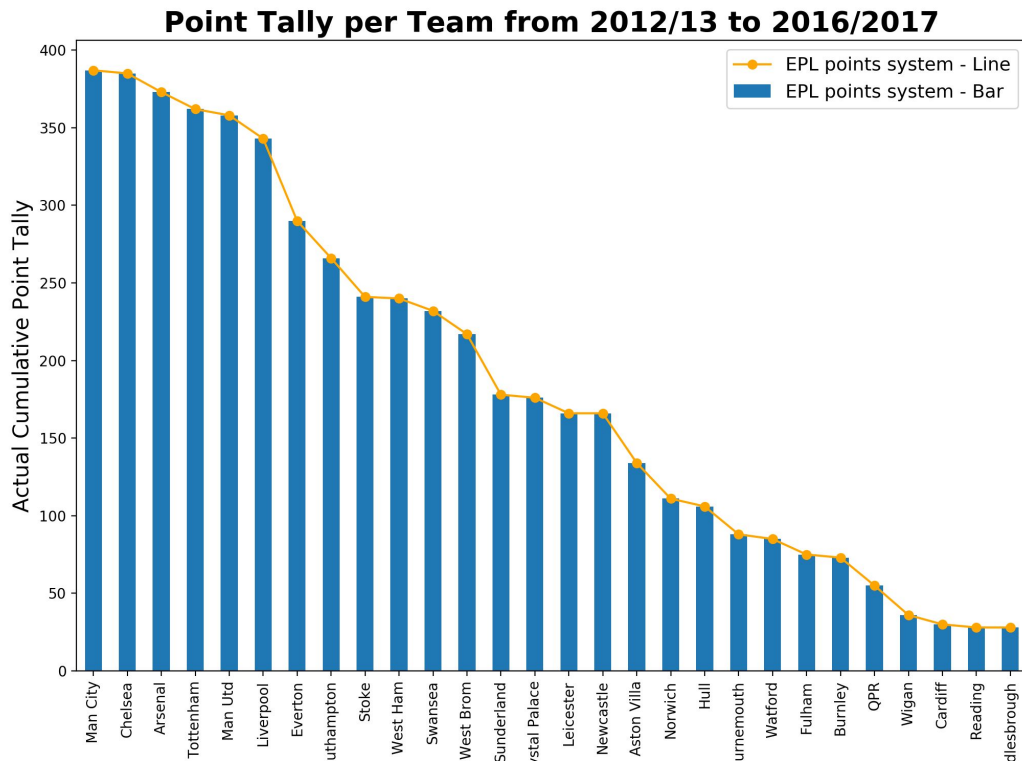
- Manchester City
- Chelsea
- Arsenal
- Tottenham
- Manchester United
- Liverpool

❖ **EPL point system**
➢ Win - 3 points
➢ Draw - 1 point
➢ Loss - 0 points
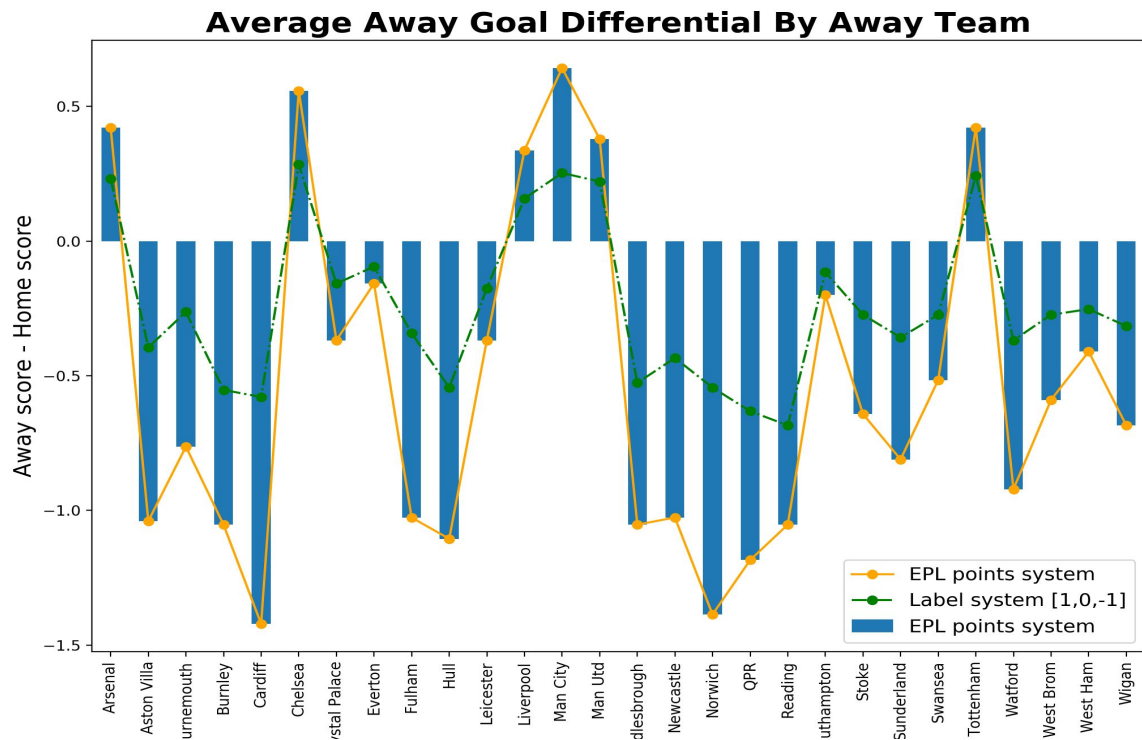
Observations - Home Advantage

Average Home Goal Differential By Home Team

❖ **EPL point system**
  ➢ Win: 3 points
  ➢ Draw: 1 point
  ➢ Loss: 0 points

❖ **Label Class system**
  ➢ Win: 1
  ➢ Draw: 0
  ➢ Loss: -1

Observations - "BIG" team effect

Average Away Goal Differential By Away Team
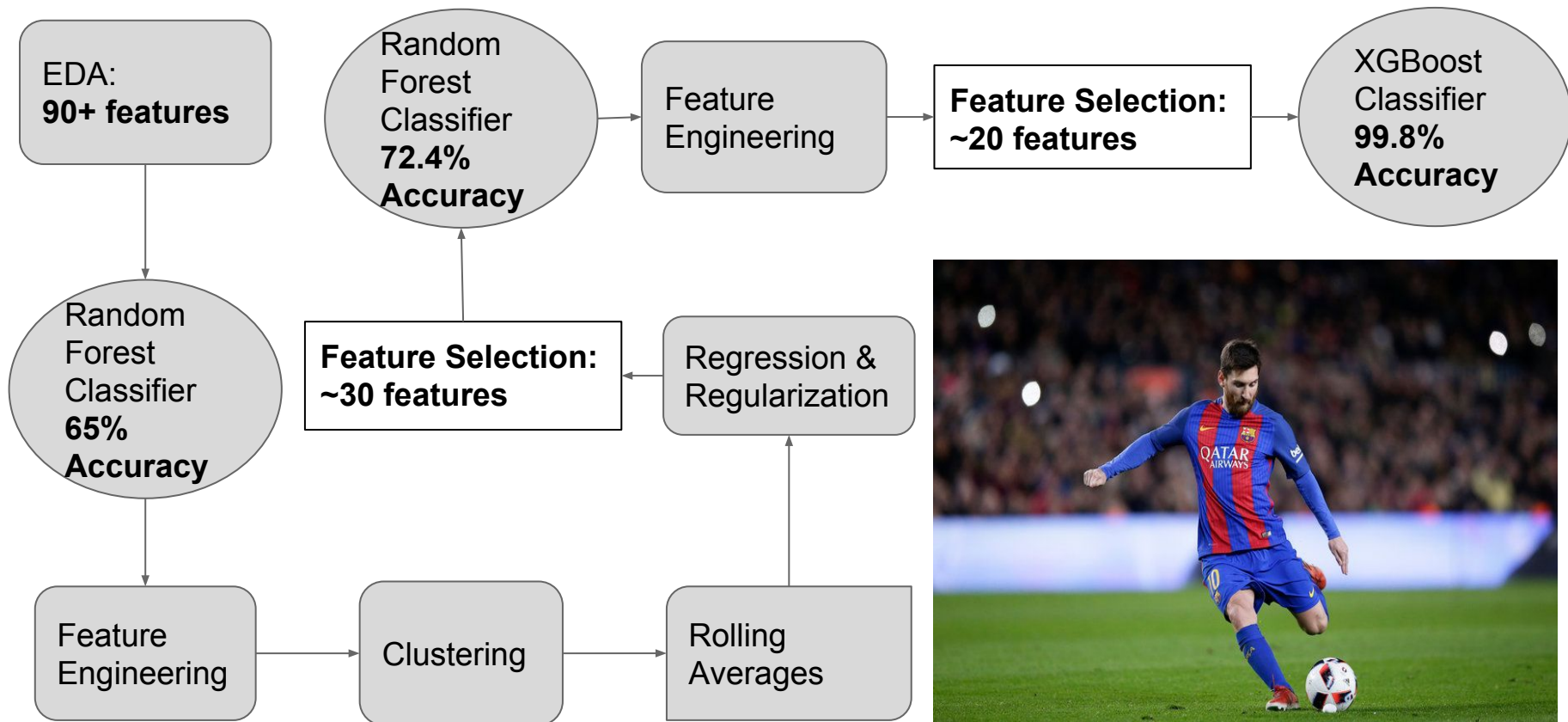
❖ **EPL point system**
  ➤ Win: 3 points
  ➤ Draw: 1 point
  ➤ Loss: 0 points

❖ **Label Class system**
  ➤ Win: 1
  ➤ Draw: 0
  ➤ Loss: -1

Model : win/loss/draw classification

Model Results

Random Forest Classification Prediction Accuracy Per Class

72.40%	Predicted		
TRUE	Loss	Draw	Win
Loss	**80.70**	14.04	5.26
Draw	21.28	**36.17**	42.55
Win	5.23	8.14	**86.63**

98.70%	Predicted		
TRUE	Loss	Draw	Win
Loss	**98.18**	0.84	0.98
Draw	0.64	**97.45**	1.91
Win	0.00	0.00	**100.00**

XGBoost Classification Prediction Accuracy Per Class

99.80%	Predicted		
TRUE	Loss	Draw	Win
Loss	**99.72**	0.00	0.28
Draw	0.00	**99.58**	0.42
Win	0.00	0.00	**100.00**

Technologies
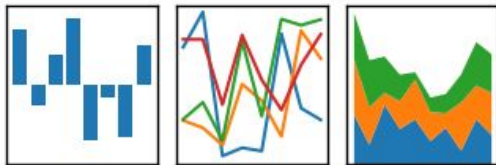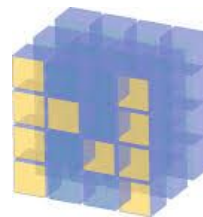
Conclusion

Recommendations

★ **Feature selection** is very important for modeling improvement
★ More **informed feature engineering** based on target variable definition
★ **Larger dataset** - more concrete and justifiable model selection
★ **Individual player statistics** and impact on team tactics, and results
  ○ "Star" player effect
  ○ Player injuries, tiredness/rest
★ **Coaching tactics'** impacts on results
★ Accuracy evaluation based on **high win/loss chance** (>= +/- 2 goal differentials)

Questions ?

linkedin.com/in/chizoba-obasi

co14@utexas.edu