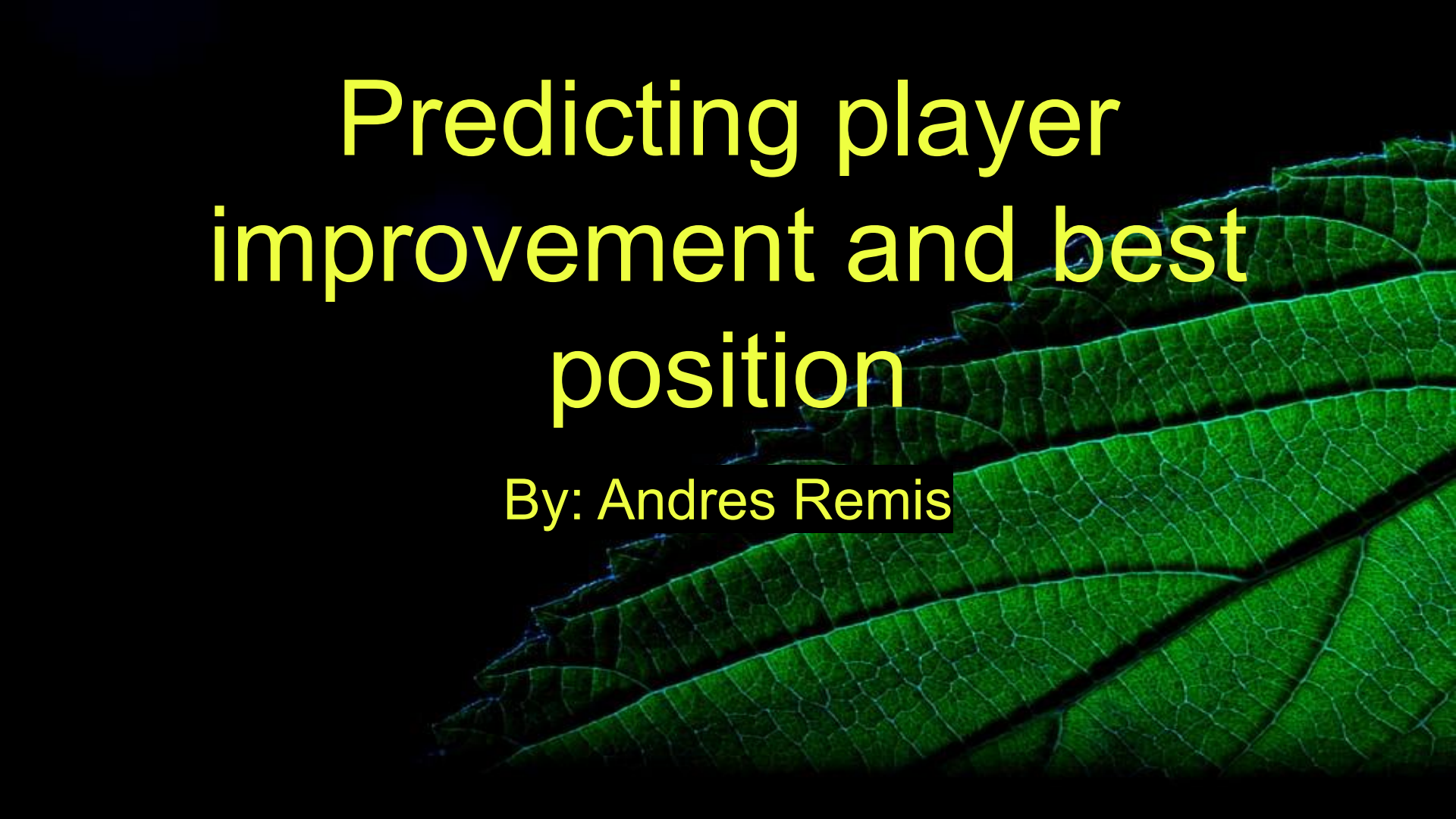
Predicting player improvement and best position

By: Andres Remis



Predicting player improvement is valuable for NBA teams

Generally, players are valued by their past performances. Therefore, players who improve a lot bring both competitive and economic advantages to teams.

- Predicting player improvement help team management.
  - Target players to acquire/release
  - Plan for performance changes of players already on the team
- Fans have interest as well (fantasy soccer)

Data acquisition and cleaning

- Player age, team and performance data (1980-2017) from Kaggle dataset, 2018 data scraped from soccer-reference.com
- Player position data (1978-2015) from Kaggle dataset, 1965-1977, 2016-2017 data scraped from soccer-reference.com
- In total, 13,378 rows and 49 features in the raw dataset.
- Duplicate, highly similar or highly correlated features were dropped.
- Cleaned data contains 24 features

Using ΔWS (win shares) as improvement measure

- Young players improve, old players decline
- Worse players have more room for improvement|
- Players who missed more games are more likely to improve
- Players are more likely to “regress to the mean” than continuously improve/decline

Regression models: dealing with unbalanced dataset

- Players with small changes are overrepresented.
- Unweighted model prioritize the error of those players.
- Resulting in narrower predicted range.
- Assigning more weights to underrepresented players help with this problem.

Regression models performance

Weighted RMSE:

- Benchmark (1 feature): 3.84
- Linear regression: 2.98
- SVM: 2.86
- Random Forest: 2.93
- Gradient Boost: 2.96

Classification models

- Log loss:
  - 0.603-0.613 between 5 models
- Accuracy:
  - 0.672-0.675 between 5 models
- SVM performed best among single algorithms, but the differences were small.

Conclusion and future directions

- Built useful models to predict whether and how much a player will improve.
- Accuracy of the models has room for improvement.
- Capture more of players' individual traits.
- Ideas include:
  - Physical data (speed, jump, etc.)
  - Financial data (contract year, amount of pay, etc.)
  - Team interaction data (strengths of players of the same position on the team)