Predicting Early NBA Shooting Talent

By: Austin Coffelt

Data Collected From Basketball Reference

Problem:

Who were the Best Shooters of the Pre-Three Era (before 1980) in the NBA?

- Less data in the early years of the NBA, making comparison difficult
- Only one stat available that directly relates to shooting ability (FT%)

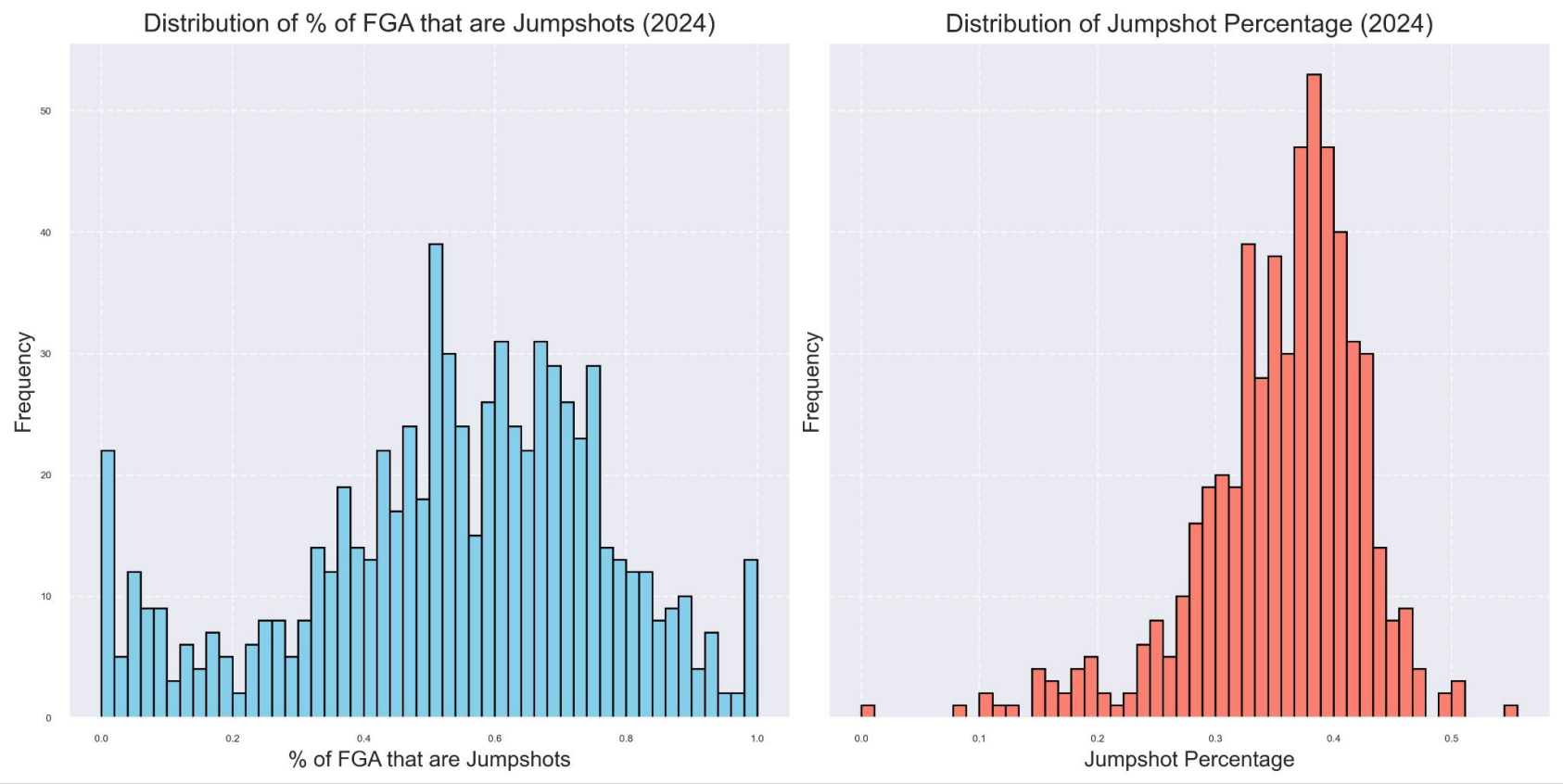
Solution:

Build two models, one that predicts % of FGA that are Jumpshots and one that predicts shooting percentage on those shots using data that would be available for the pre-1980 NBA.

- Predictive Models along with FT% would give us a good proxy to judge shooting talent of the early NBA

Data Collected From Basketball Reference

Data Exploration



Data Collected From Basketball Reference

Data Pre-Processing

- Centered and Scaled every stat (both outcome and predictor) to their year

Missing Values

- Jumpshot Percentage was only stat missing values (happens when a player does not shoot any Jumpers)
- Kept observations missing values when predicting % of FGA that are Jumpshots, removed them when predicting Jumpshot Percentage

Predictors in the different models

- Added % of FGA that are Jumpshots to my Jumpshot Percentage model
- Correlation remover removed different predictors in each model (PTS vs FGA)

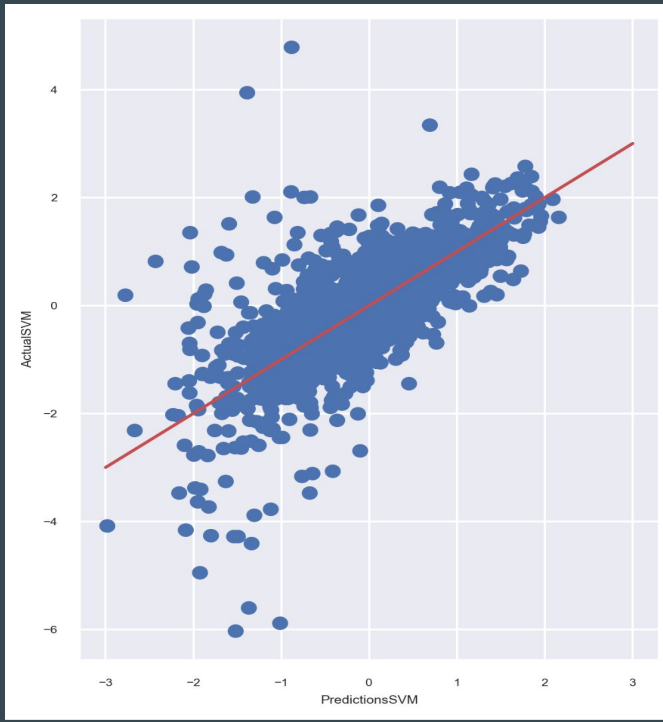
Data Collected From Basketball Reference

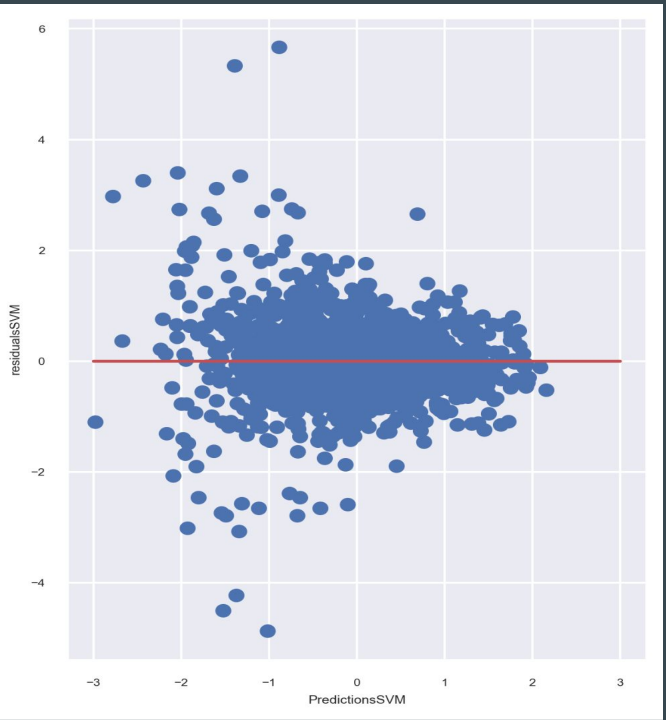
Model Building, Evaluation, and Performance

Training and testing RMSE and R2 for each model (RMSE, R2)	Linear Regression	PLS Regression	Ridge Regression	Support Vector Machine Regression	Neural Network Regression	Random Forest Regression
% of FGA that are Jumpshots	Training: (0.463, 0.785) Testing: (0.457, 0.792)	Training: (0.463, 0.785) Testing: (0.457, 0.792)	Training: (0.463, 0.785) Testing: (0.457, 0.792)	Training: (0.391, 0.847) Testing: (0.427, 0.819)	Training: (0.400, 0.840) Testing: (0.426, 0.819)	Training: (0.375, 0.859) Testing: (0.455, 0.793)
Jumpshot Percentage	Training: (0.694, 0.523) Testing: (0.702, 0.485)	Training: (0.694, 0.524) Testing: (0.702, 0.485)	Training: (0.694, 0.524) Testing: (0.702, 0.485)	Training: (0.607, 0.634) Testing: (0.675, 0.524)	Training: (0.604, 0.632) Testing: (0.666, 0.537)	Training: (0.544, 0.707) Testing: (0.681, 0.516)

Results

- Jumpshot Percentage much more difficult to predict
- Choose SVM as the best model





Predicting Pre-1980 NBA Shooting Statistics

- Use MLPRegressor for Predictions
- Metric: Z(FT%) + Z(PercJump) + Z(JumpPerc)

Player	Year	Shoot Tal
J. McGlock	1972	5.95
J. McGlock	1971	5.77
J. McGlock	1974	5.74
J. McGlock	1973	5.65
Bill Sharm	1959	5.55

Player	Career Shooting Talent Average
Bill Sharman	4.94
Jon McGlocklin	4.58
Vince Boryla	4.22
RIck Barry	4.03
Calvin Murphy	4.00

Data Collected From Basketball Reference

Conclusion

Contributions

- Through this project, we gain a deeper understanding of the skill of past NBA players
- Allow for better future comparison between past and present players

Future Research

- Many more old stats can be predicted...
- Update the model for the pre-data era (1980 - 1998)
- Statistical Significance of my metric

Data Collected From Basketball Reference

Sources

- All Data via Basketball Reference
- Class Notes