Predicting Players' Salaries

Paul Lim
04/28/2017

Objective

Salary Prediction

- Eye test versus advanced metrics.

Data Collection

Web Scraping

Merge into one dataframe

Preliminary Feature Selection
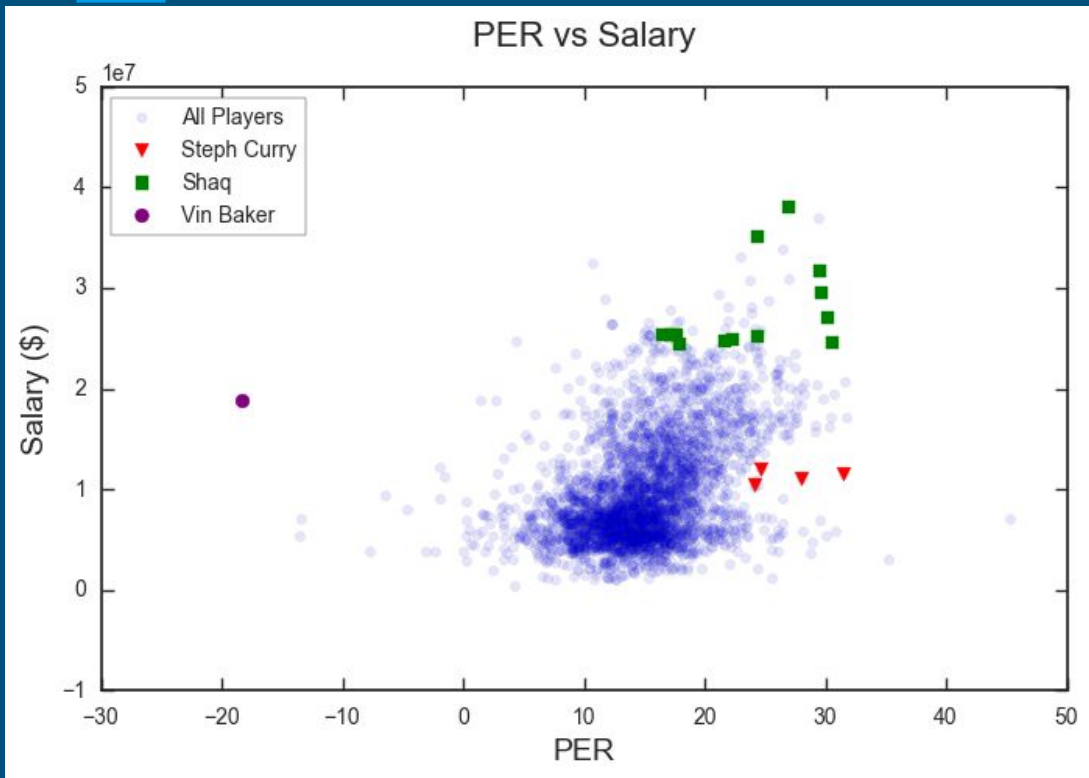
- **Outliers!**
- Vin Baker makes more money than Steph Curry?

Modeling

Approach
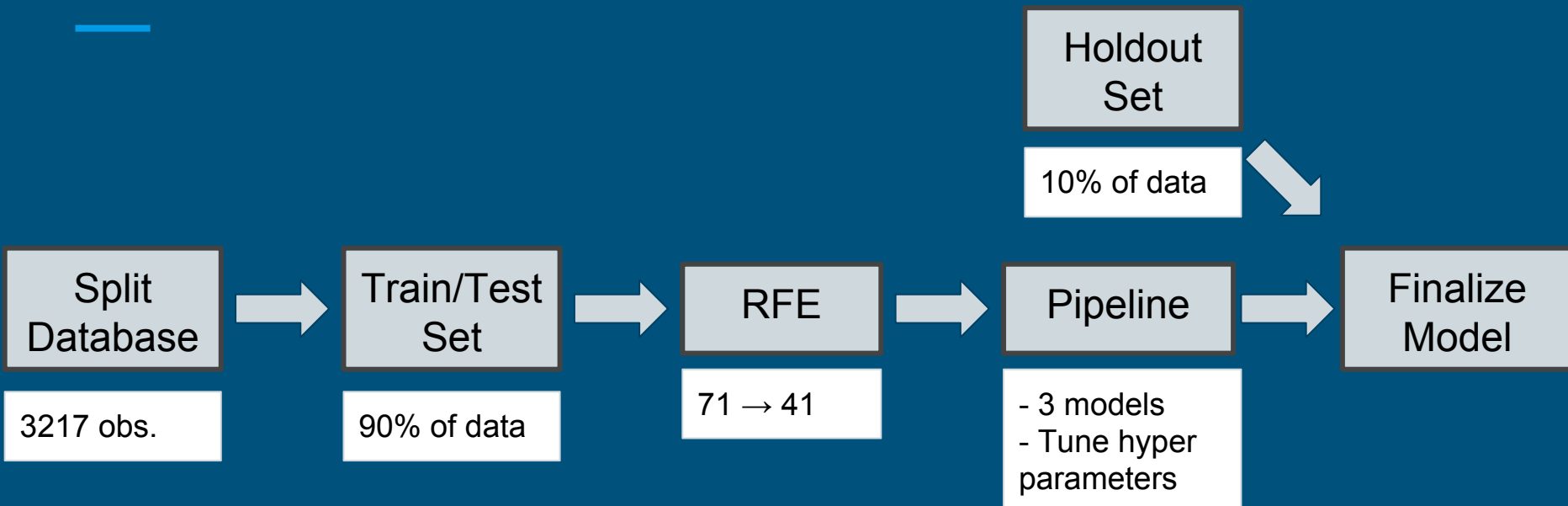
Model Scores

Model:	Score:
Lasso	0.433
Random Forest	0.369
Gradient Boost	0.465

Gradient Boost Regressor performed the best!

Gradient Boost Performance

Predicted vs True Salaries
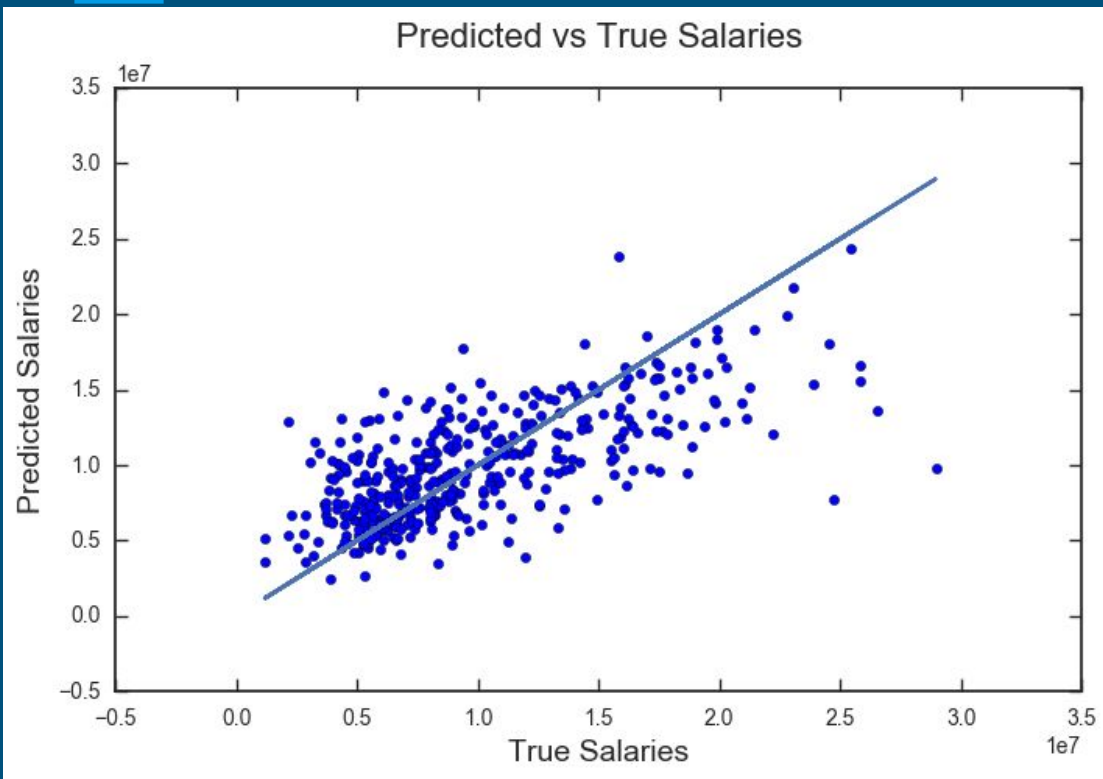
Score = 0.465

Hyper Parameters:

- *min_samples_split* = 20
- *max_depth* = 5
- *max_features* = 'sqrt'
- *n_estimators* = 100

Features:

- Age, PF, USG%, AST%

N_Estimators (GB)
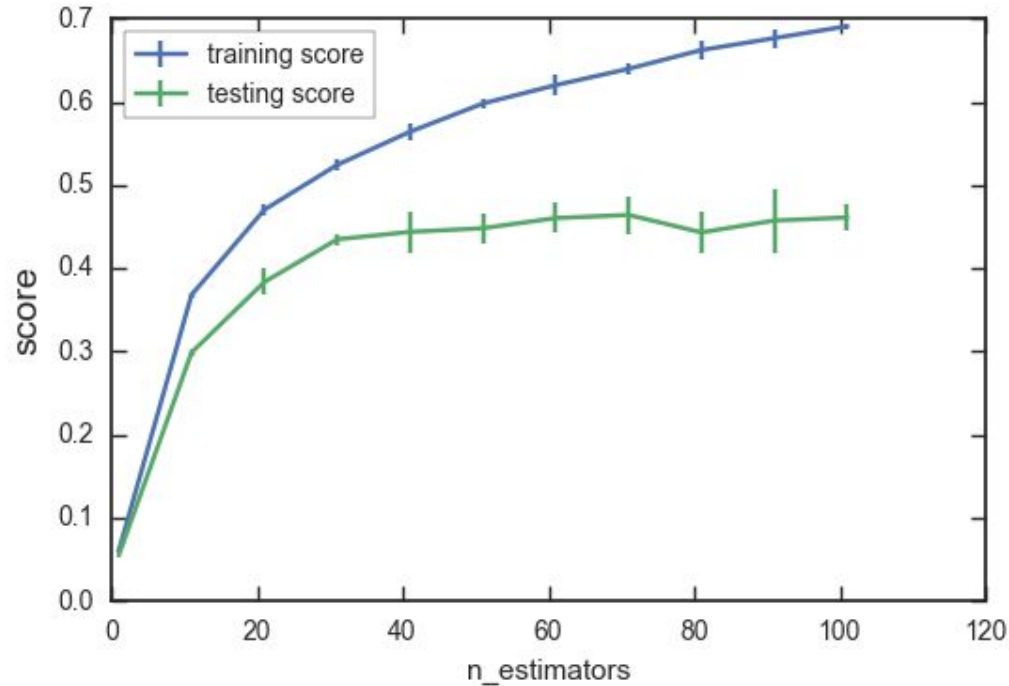
Future Works

Additional Items

1. Continue feature engineering
2. Research other data sources to gain additional observations
3. Try more advanced models

Appendix

Lasso Performance

Lasso: Predicted vs True Results

Score = 0.433

Hyper Parameters:

- alpha = 100

Random Forest Performance

RandomForest: Predicted vs True Results
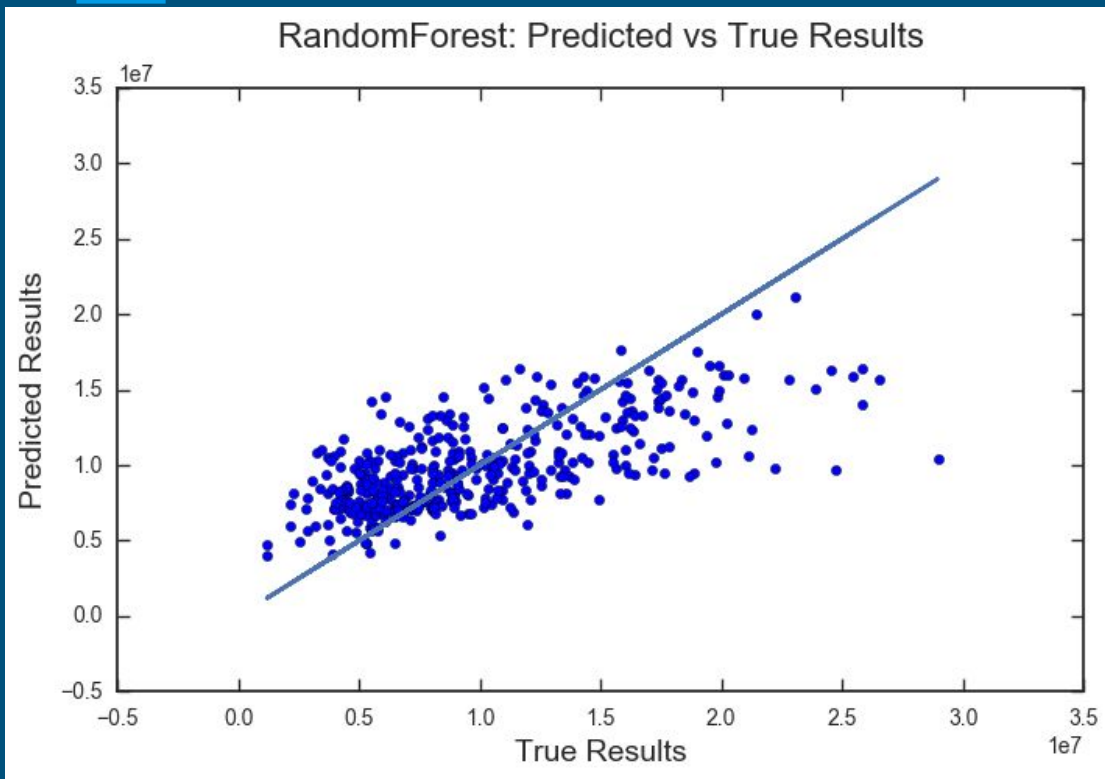
Score = 0.369

Hyper Parameters:

- min_samples_split = 5
- max_depth = 5
- max_features = 'auto'
- n_estimators = 300

Gradient Boosting Features

```
[(0.039979288201189275, 'AST%'),
 (0.034430834077222035, 'PF'),
 (0.03269383653901576, 'Age_cube'),
 (0.030205956306421169, 'FGA'),
 (0.029652310877804493, 'USG%'),
 (0.02957027807958075, 'BLK%'),
 (0.029388523589018594, 'Age'),
 (0.029101000300381846, 'TOV%'),
 (0.028110336549877175, 'DRB%'),
 (0.027743156826659378, 'FG'),
 (0.027681541879088792, 'BPM'),
 (0.027285327422077968, 'MP'),
 (0.026464547758716619, 'FG%'),
 (0.026423264635552043, 'DRB'),
 (0.02630417727828049, 'USG%_sq'),
 (0.02627463138497501, 'AST'),
 (0.025771241214545899, 'WS/48'),
 (0.025737486040819848, 'ORB%'),
 (0.025582074786042165, 'ORB'),
 (0.025436920757196144, '2PA'),
 (0.025210471898029149, 'PER'),
 (0.0240735454101354, '3PAr'),
 (0.023922390005604567, 'FT'),
 (0.02330981074539569, 'FTA'),
 (0.022759560808408495, 'PER_sq'),
 (0.022472003945057332, 'DBPM'),
 (0.022354467928711642, 'OWS'),
 (0.022068454392837111, 'G_sq'),
 (0.021581501886616498, 'TRB%'),
 (0.021188718958678125, '2P'),
 (0.020696583701643682, 'GS'),
 (0.02026503876609427, 'FTr'),
 (0.020185112752095669, '2P%'),
 (0.0193905135158883386, 'G_cube'),
 (0.019249563510110625, 'GS_sq'),
 (0.018436926568250019, '3PA'),
 (0.018094805218174755, 'OBPM'),
 (0.017953431851572749, '3P%'),
 (0.01637450631553275, 'FT%'),
 (0.013738415910571662, 'TS%'),
 (0.012837441406130955, 'eFG%')]
```
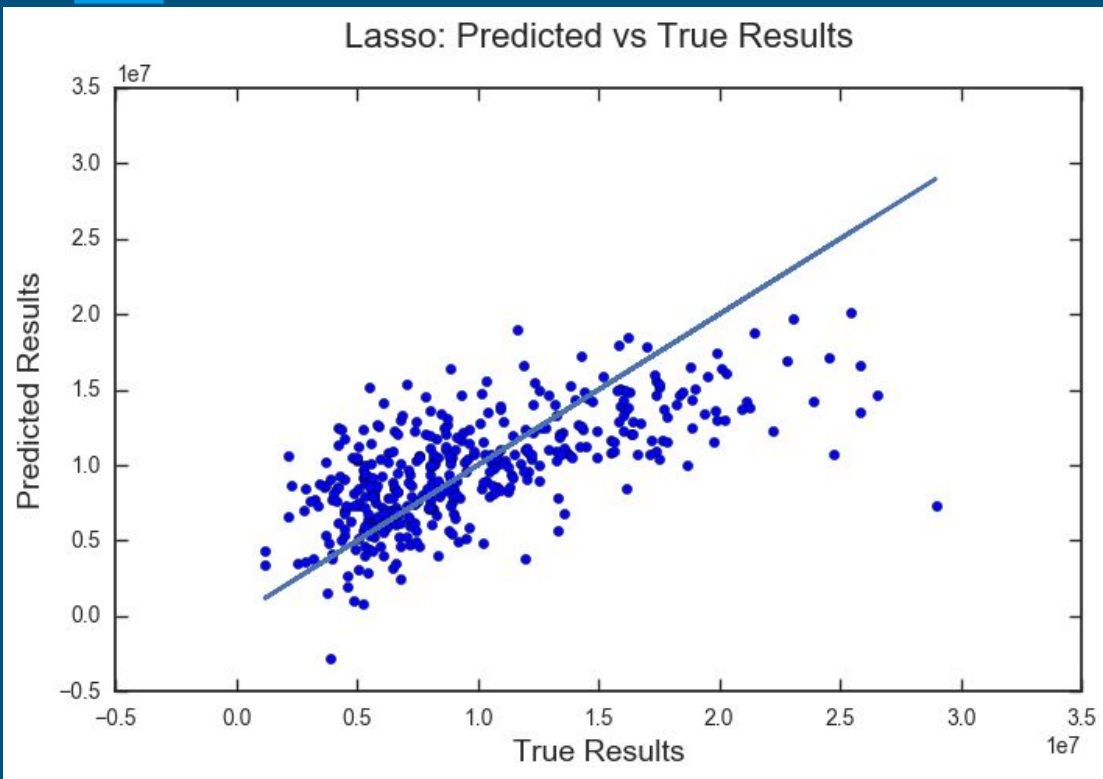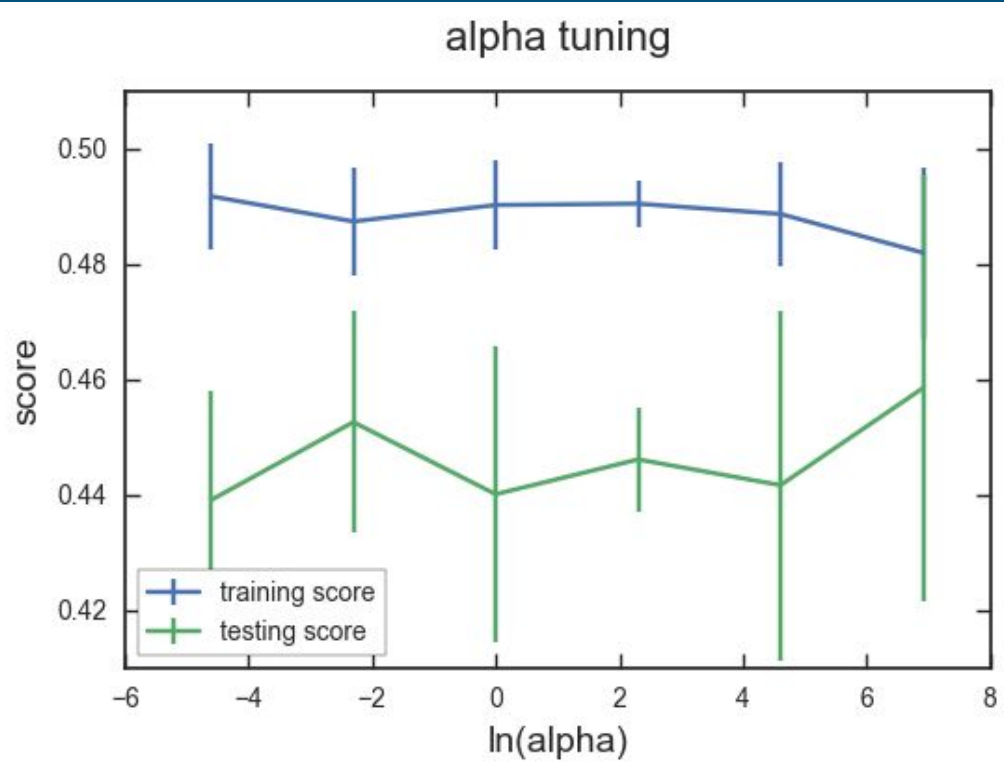
Random Forest Features

```
[(0.21915235461482402, 'PER_sq'),
 (0.077377427094435097, 'Age_cube'),
 (0.074530022580351865, 'PER'),
 (0.0731924457744582807, 'Age'),
 (0.072213724124162035, 'Age_sq'),
 (0.061163557920318919, 'USG%'),
 (0.058723019302410473, 'USG%_cube')
 (0.046612430865444344, 'USG%_sq'),
 (0.041440574263541467, 'GS_sq'),
 (0.041292793464155784, 'GS'),
 (0.015446957957005712, '2P'),
 (0.015266474351825083, 'DRB%'),
 (0.014762796307324084, 'DRB'),
 (0.012752097753354423, 'DBPM'),
 (0.010860198809678147, 'FTA'),
 (0.010249971062386643, 'FG'),
 (0.0083056726266983426, '2P%'),
 (0.0061821467492086859, 'ORB'),
 (0.0059869100775622871, 'BLK'),
 (0.0053197784595053373, 'BLK/TRB'),
 (0.0048570375810426695, 'PTS'),
 (0.0046551494113846676, 'eFG%'),
 (0.0045737118879394679, 'PF'),
 (0.0045428038314191824, 'FG%'),
 (0.0043683383012614761, '2PA'),
 (0.0042526993208324964, 'WS/48'),
 (0.0041993980405974083, 'TRB%'),
 (0.0041785641064511388, 'FT%'),
 (0.0041639404033720828, 'WS/OWS'),
 (0.0040390172057284783, 'STL/AST'),
 (0.0040327966761899313, 'WS'),
 (0.0040279164362791052, 'FTr'),
 (0.0039950560604004062, 'STL/TOV'),
 (0.0038356782518835825, 'BLK%'),
 (0.0037321619726945312, 'AST%'),
 (0.0037145982240704999, 'TRB'),
 (0.0035192936503895232, 'VORP'),
 (0.0035047935900841794, 'BPM'),
 (0.003344213892163511, 'DWS'),
 (0.0030674102005504161, 'AST/TOV'),
 (0.0030642881057164834, 'OWS'),
```
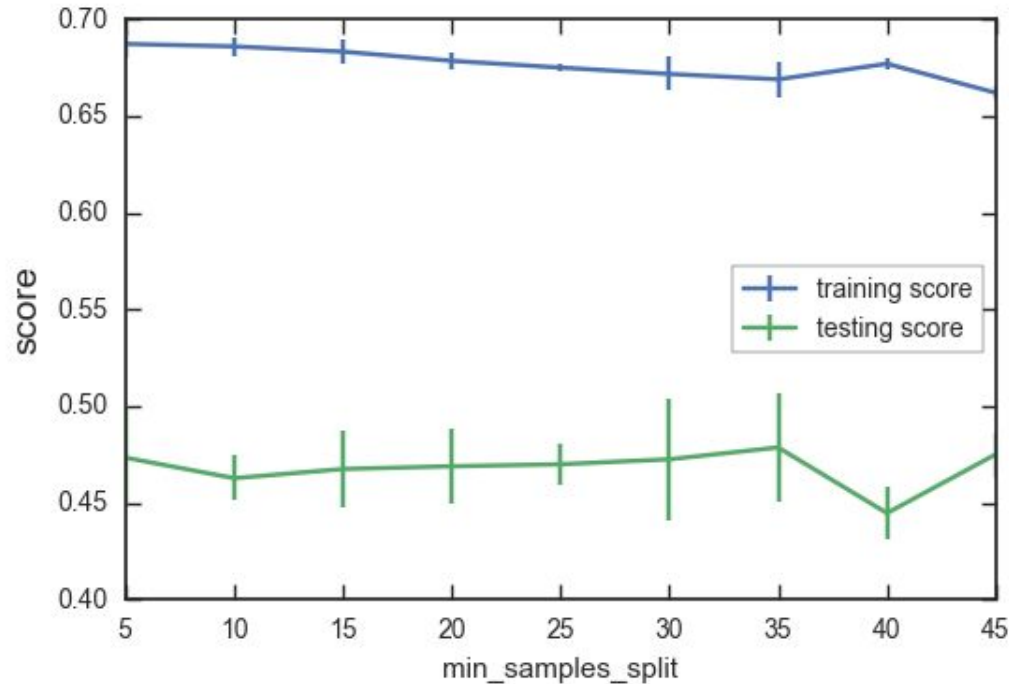
Lasso Features

```
[(20809637.504010659, 'WS/48'),
 (4334278.801074787, 'WS/PER'),
 (3120097.2331944867, '3PAr'),
 (2764710.0896681128, 'FTr'),
 (2507781.2391417096, 'OWS/PER'),
 (2207568.7163218916, 'FG%'),
 (1972058.4062925798, '2P%'),
 (1853437.4721244236, 'TS%'),
 (1607382.3668013043, 'Age'),
 (1276023.4507610183, 'TRB%'),
 (953750.13401092729, 'DWS'),
 (921270.92150262278, 'STL%'),
 (815237.27855895984, 'ORB%'),
 (811037.81779628713, 'VORP'),
 (758613.42186131305, 'BPM'),
 (495054.41241291654, 'DRB%'),
 (470071.99158422532, 'AST/TOV'),
 (464267.56190868397, 'OWS'),
 (437495.92518800939, 'Pos_PF'),
 (210418.3154952332, 'Pos_SG'),
 (192206.51616339243, 'WS'),
 (164279.412235503, 'BLK%'),
 (119857.03346106537, 'PER'),
 (108920.78653531287, 'TOV_rk'),
 (102982.36834964711, 'PTS_rk'),
 (95680.679220600257, 'STL/TOV'),
 (76911.217136375781, 'G'),
 (75089.422737606175, 'USG%'),
 (65938.969619748255, 'OBPM'),
 (65712.522192984514, 'BLK_rk'),
 (58244.034015487865, 'PTS/TOV'),
 (46010.286815258725, 'AST%'),
 (45733.499460716477, 'MP_rk'),
 (33291.383126448854, 'DBPM'),
 (31492.089097692755, 'TRB_rk'),
 (26820.173146698609, 'GS'),
 (24772.558260432157, 'STL_rk'),
 (15520.243275691148, 'USG%_sq'),
 (15247.170229498362, '3P'),
 (13524.328365056637, 'FG'),
 (13522.136851627827, '2P'),
```

Alpha Tuning (Lasso)

Min Samples Split (GB)

Max Features (GB)