Predicting NBA Salaries

John Makhijani & Juna Iafelice

May 5, 2021

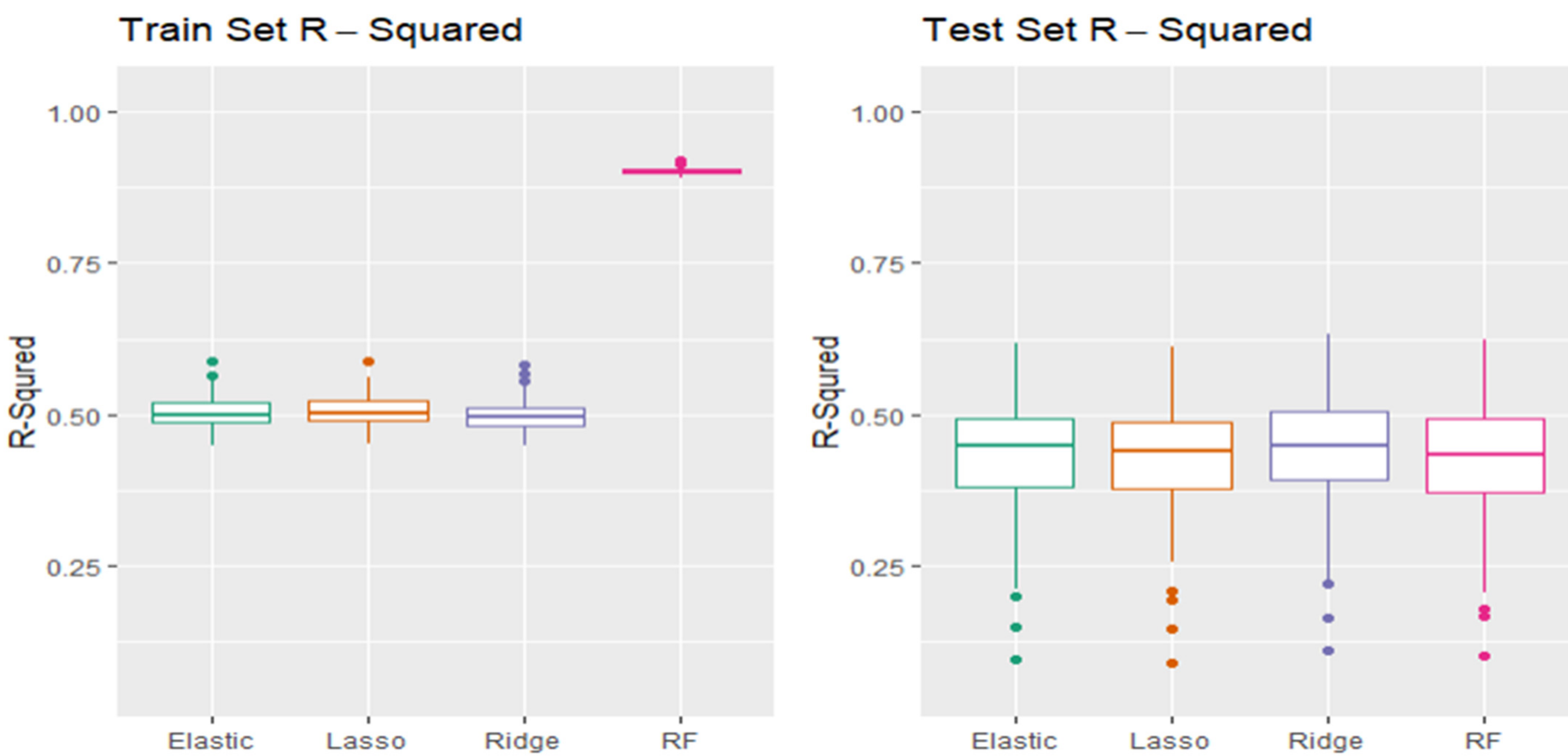
NBA Dataset Description

Goal - Predict NBA player salaries based on player statistics

- Data Breakdown
  - Sample Size (n) – 413 | Predictors (p) – 41 | No missing data
  - Predictors based on 18/19 season, Salary based on 19/20 season
  - Response Variable – Salary
  - Predictors – Field Goals, Rebounds, Three Pointers, Games, Minutes, Points, Age, PER, VORP, WS, etc.
  - Data Source – basketball-reference.com and espn.com

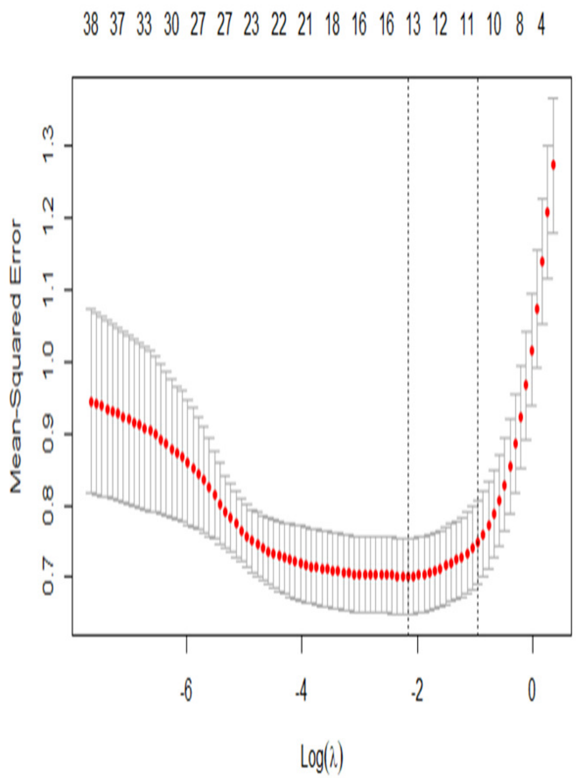
- Adjustments
  - Top 5 salaries removed as outliers from n=418 dataset
  - Natural log taken of Salary data

R^2_{test} and R^2_{train}

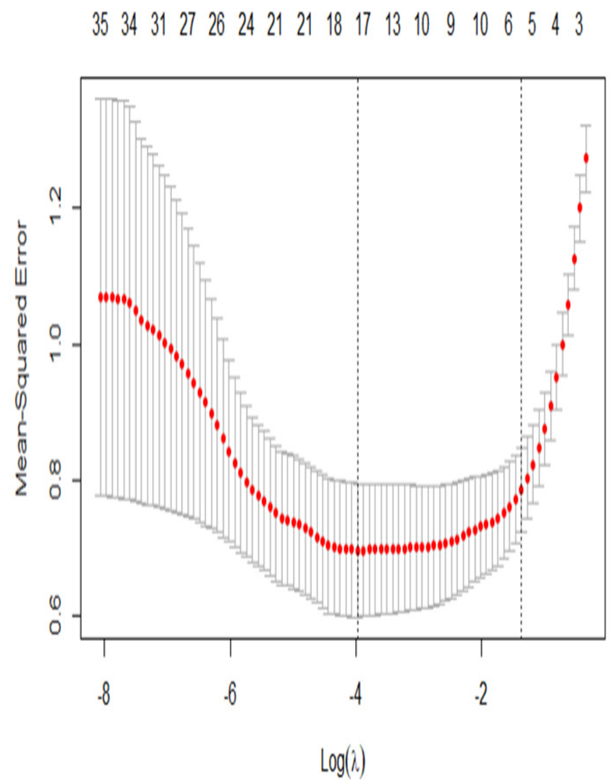


Cross Validation Curves

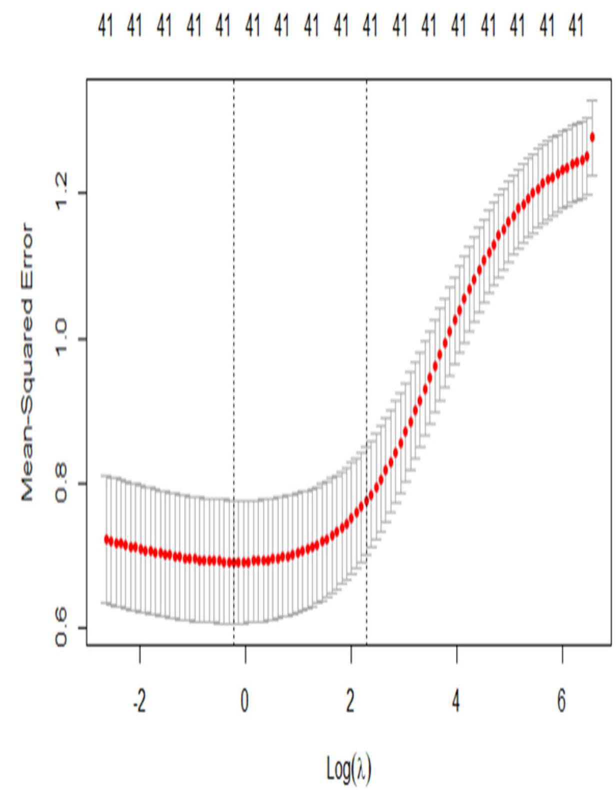
Elastic-net



Lasso

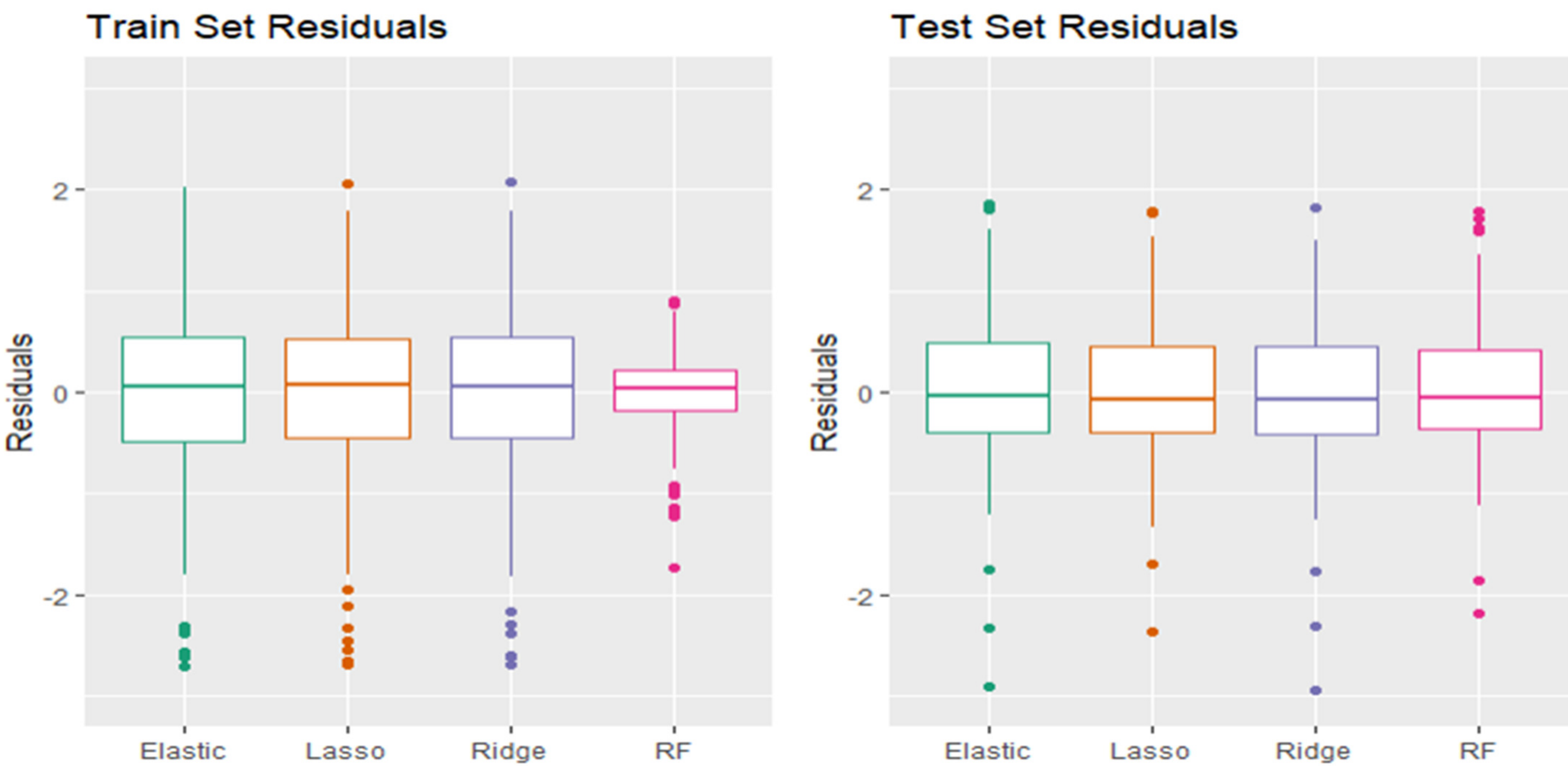


Ridge

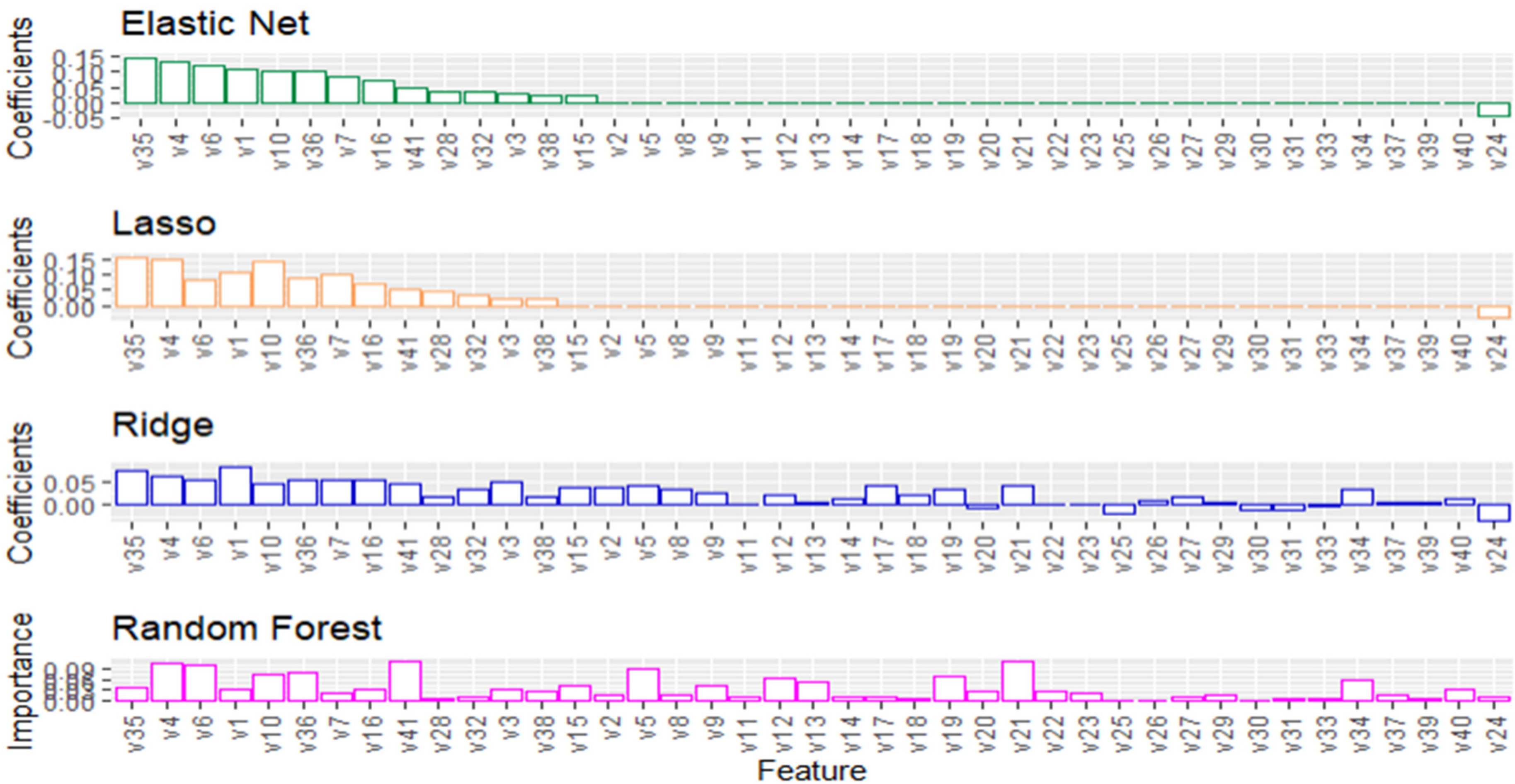


	ELASTIC NET	LASSO	RIDGE	RANDOM FOREST
MEAN CV TIME(SECS)	0.217	0.269	0.099	1.027

Residuals


Train Set Residuals
Test Set Residuals
Residuals
2
0
-2
Elastic
Lasso
Ridge
RF

Estimated Coefficient



Model Performance / Accuracy Tradeoff

	90% Test R^2 Interval	Time
ELASTIC NET	0.272 - 0.555	0.332 secs
LASSO	0.265 - 0.559	0.272 secs
RIDGE	0.276 - 0.563	0.194 secs
RANDOM FOREST	0.253 - 0.567	1.275 secs

Conclusion

- We see an obvious overfitting issue with the training set R^2 values of the Random Forest model that is not seen in the 3 other methods
- Ridge has the best performance in terms of R^2 on the Test set
(We are not considering Random Forest because of the overfitting issue)
- For the trade-off between model accuracy and processing time, Ridge gives us the highest R^2 and the fastest time to run which makes it the best model to fit to predict NBA Salaries