

# SPORTS ARTICLES REGRESSION ANALYSIS

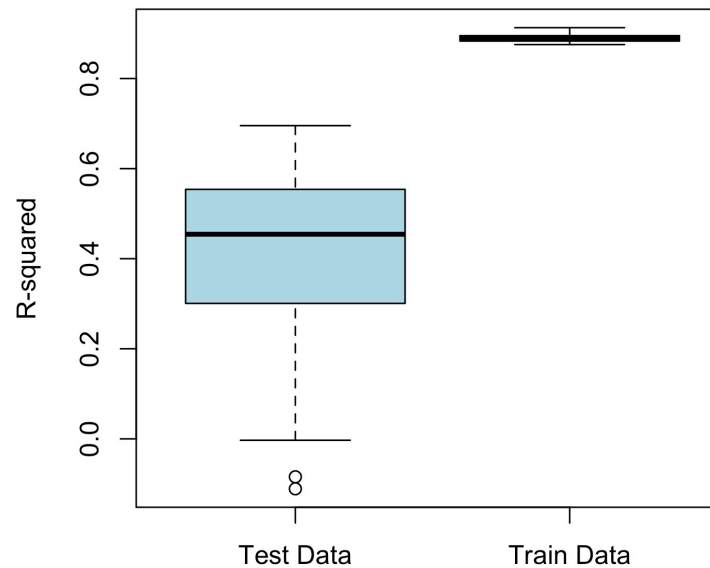
Alexandra Eberts

# SPORTS ARTICLES FOR OBJECTIVITY ANALYSIS DATASET

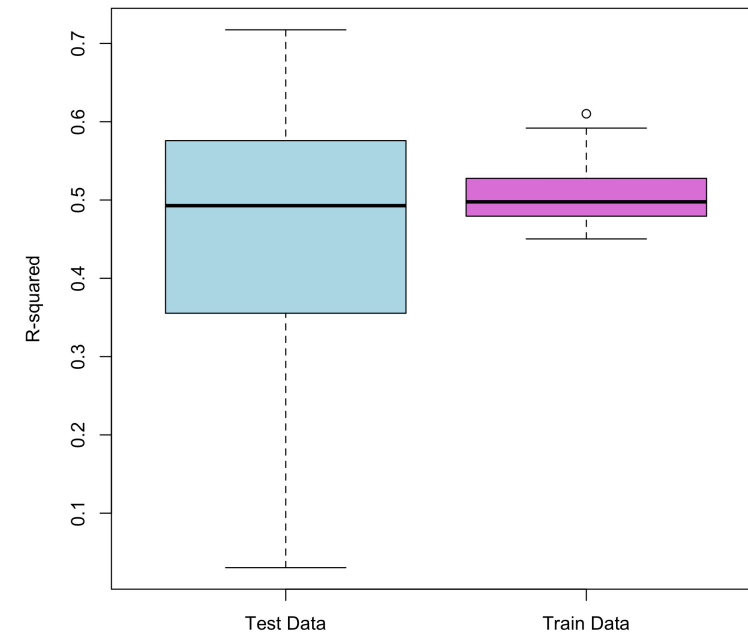
- Data from 1000 sports articles
- Original goal: classify article as objective or subjective
- Goal for this project: analyze data to predict text complexity
- $p = 54, n = 1000$
- Response variable: Text Complexity Score
- <https://archive.ics.uci.edu/ml/datasets/Sports+articles+for+objectivity+analysis>

# R-SQUARED TESTS

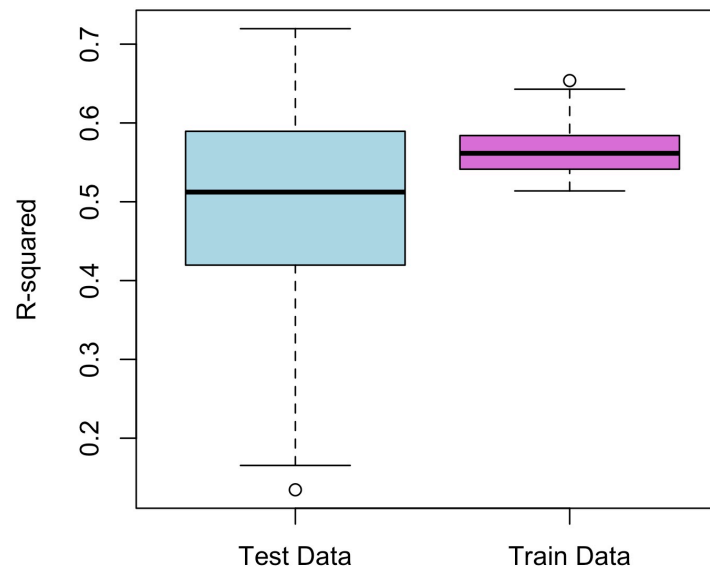
## Random Forest R-squared Test



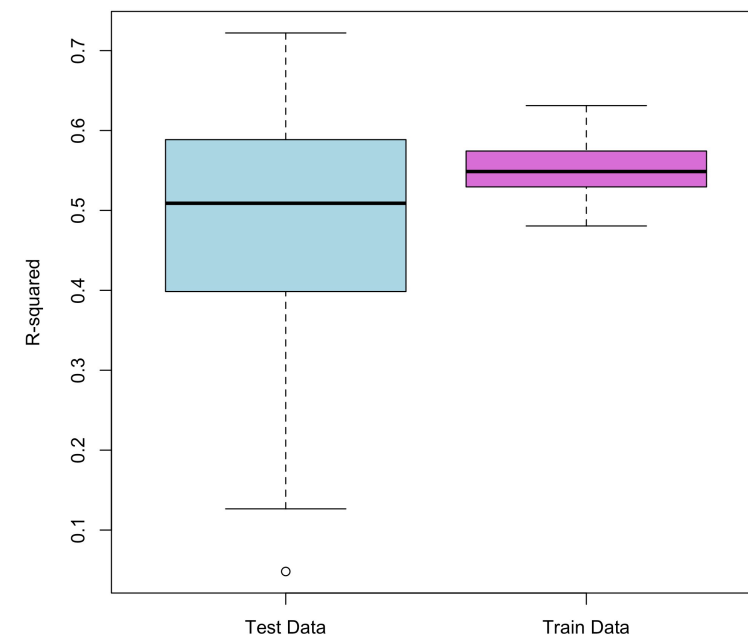
## Ridge Regression R-squared Test

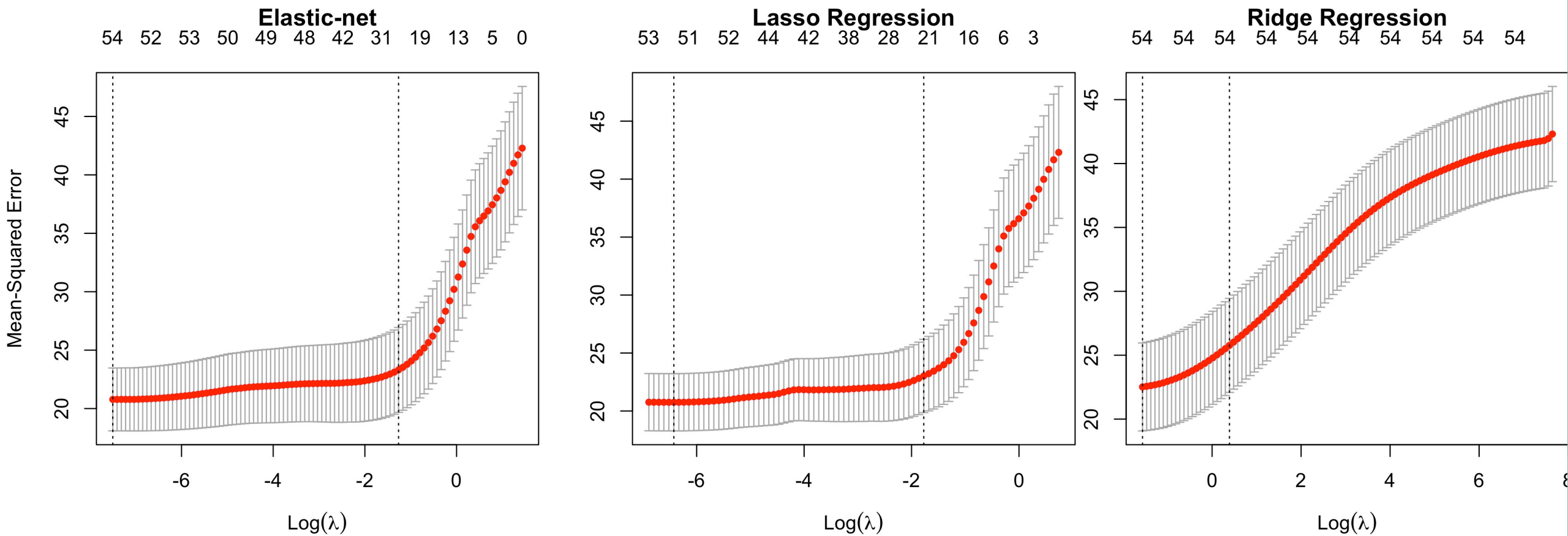


## Elastic-Net R-squared Test



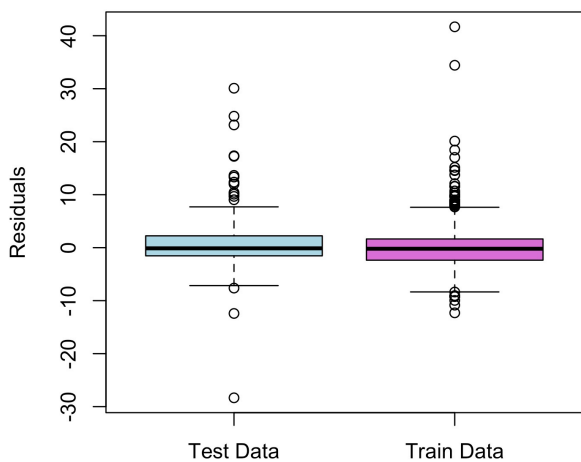
## Lasso Regression R-Squared Test



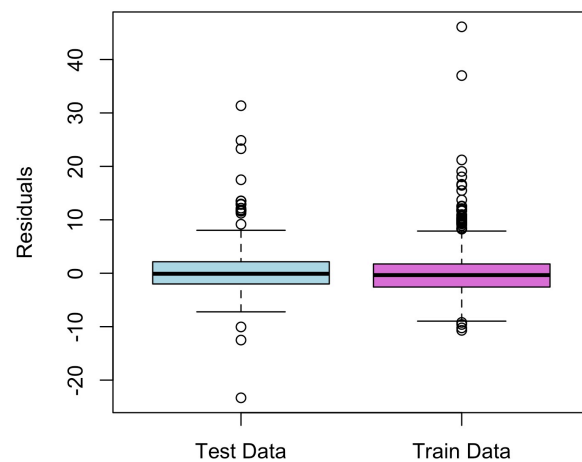


10 FOLD CV CURVES

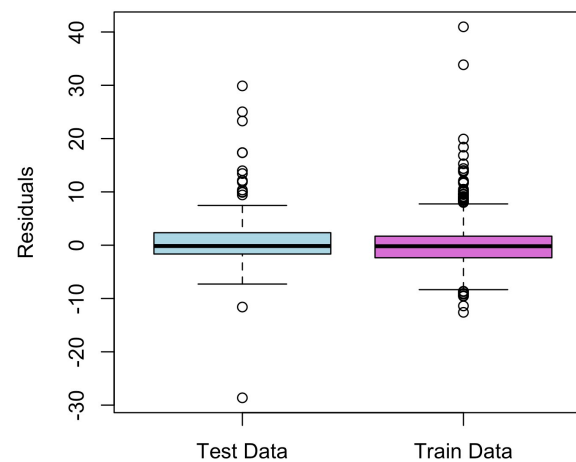
**Lasso Regression**



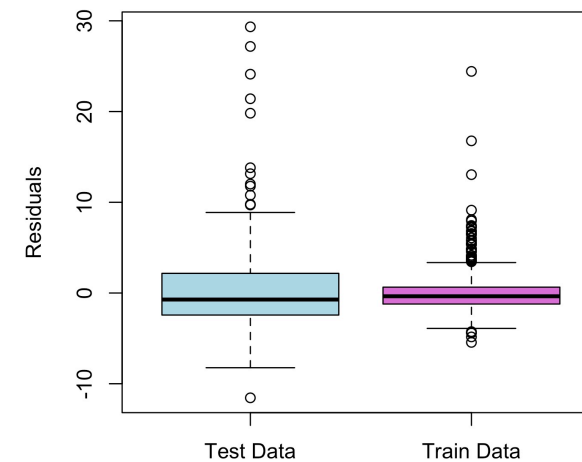
**Ridge Regression**



**Elastic Net**



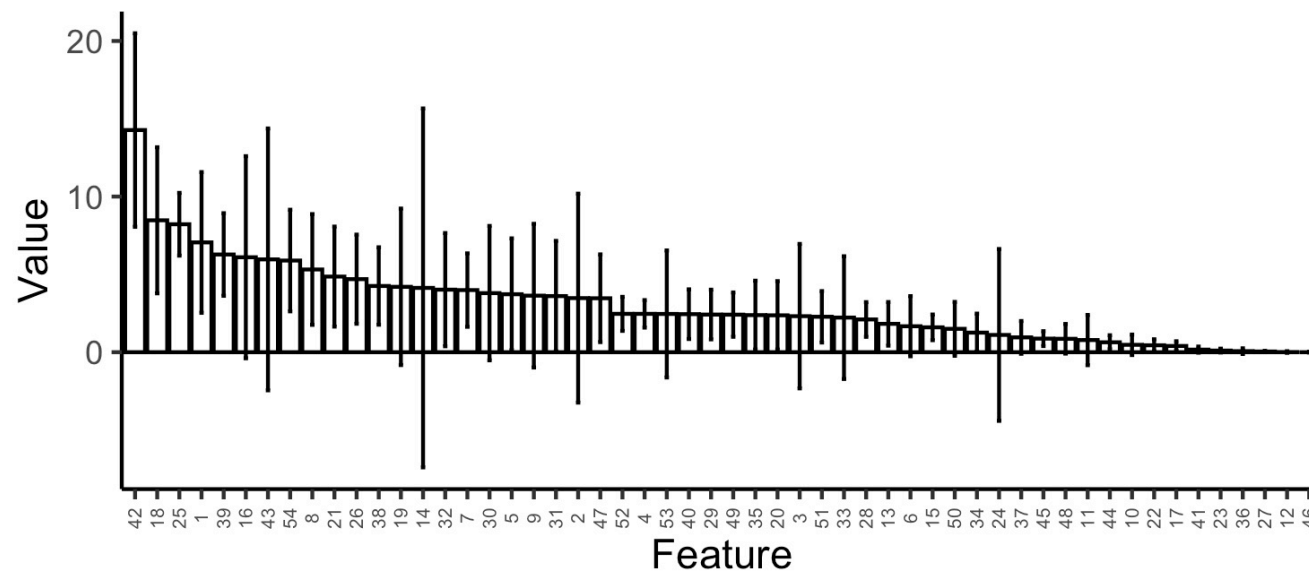
**Random Forest**



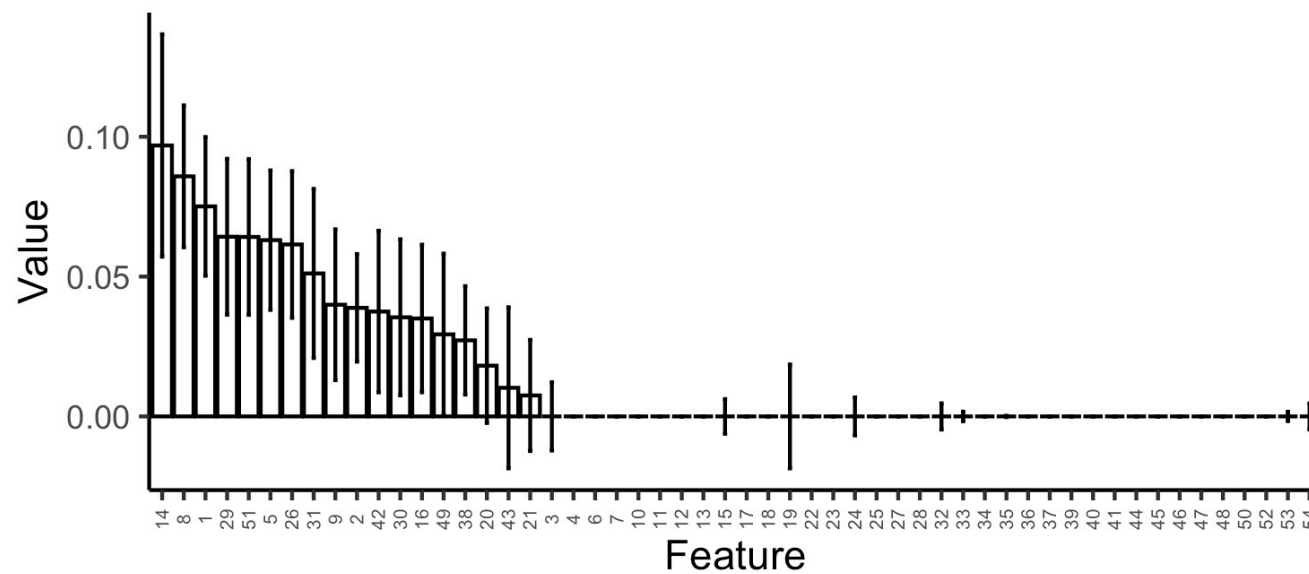
RESIDUALS

# ESTIMATED COEFFICIENTS

Random Forest



Elastic-Net



## SUMMARY

Model	Performance- $R^2$ Test
Random Forest	Train: 89%, Test: 41%
Elastic-Net	Train: 56%, Test: 50%
Ridge	Train: 51%, Test: 46%
Lasso	Train: 56%, Test: 49%