

Final Report Presentation

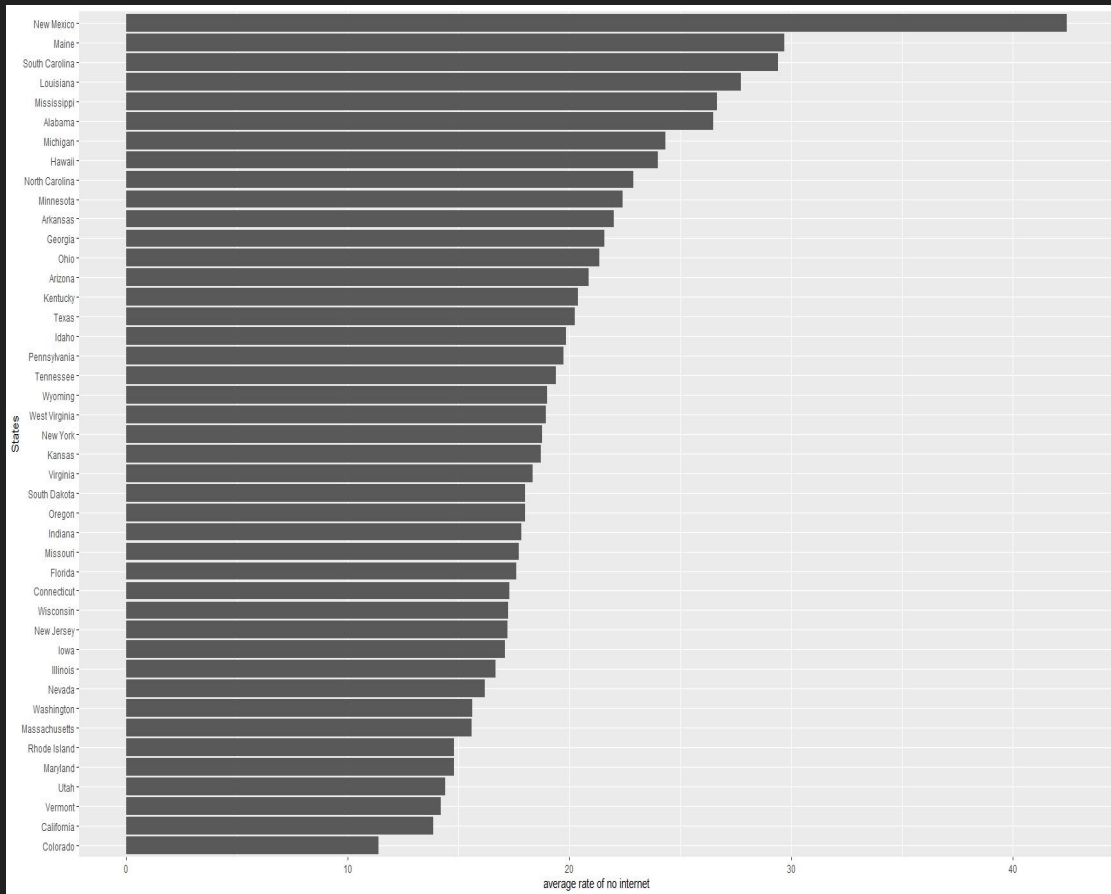
Jack Wang

Introduction

- Muddy broadband data by FCC
 - Self-reporting from FCC
- Pandemic highlights internet connectivity importance
- Could potentially help predict internet connectivity

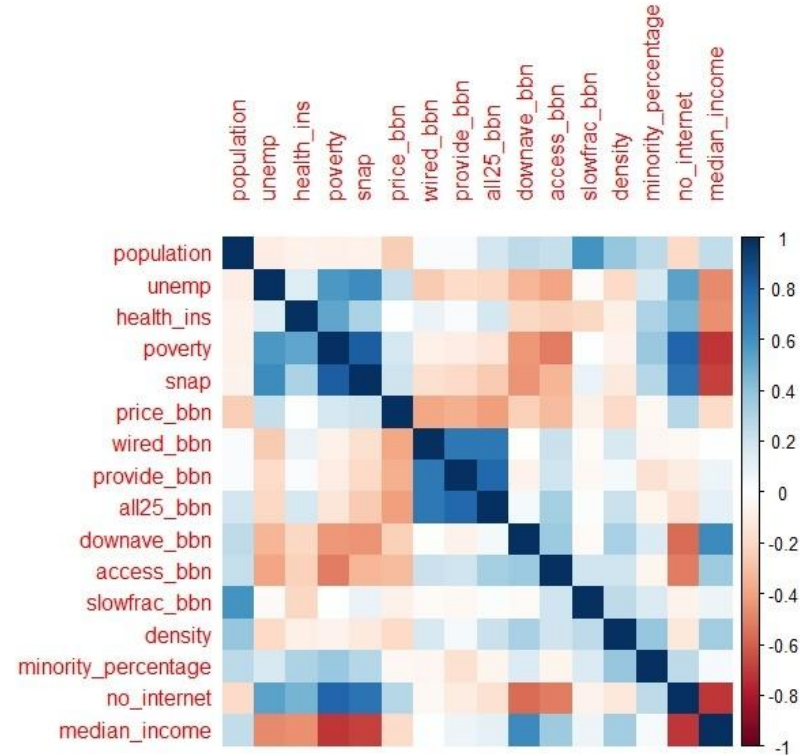
Exploratory Data Analysis

- Substantial differences among states
 - State as a dummy variable



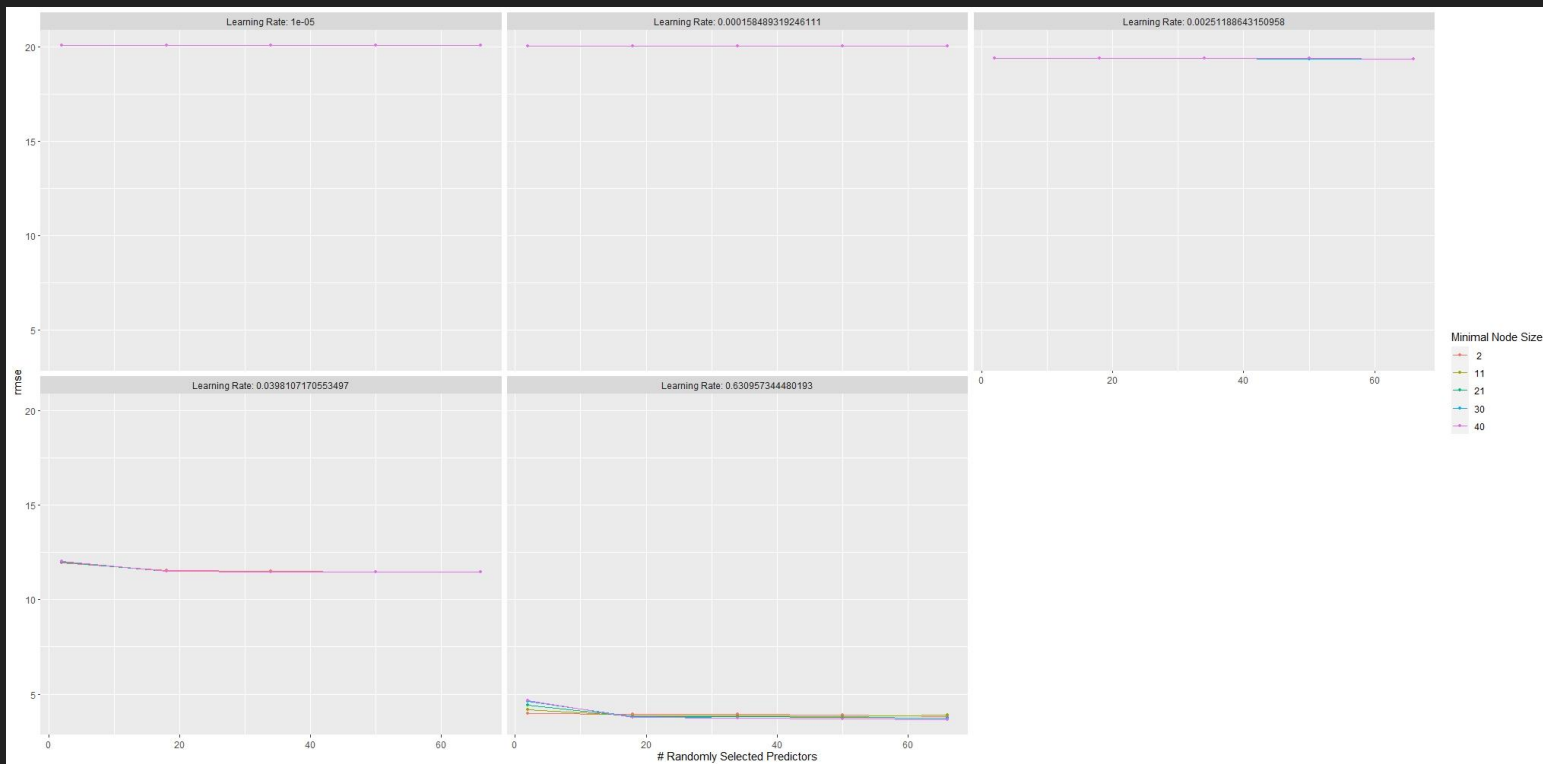
Exploratory Data Analysis

- Which variable correlate with no_internet?
- Which variable could impute for missing minority_percentage values?



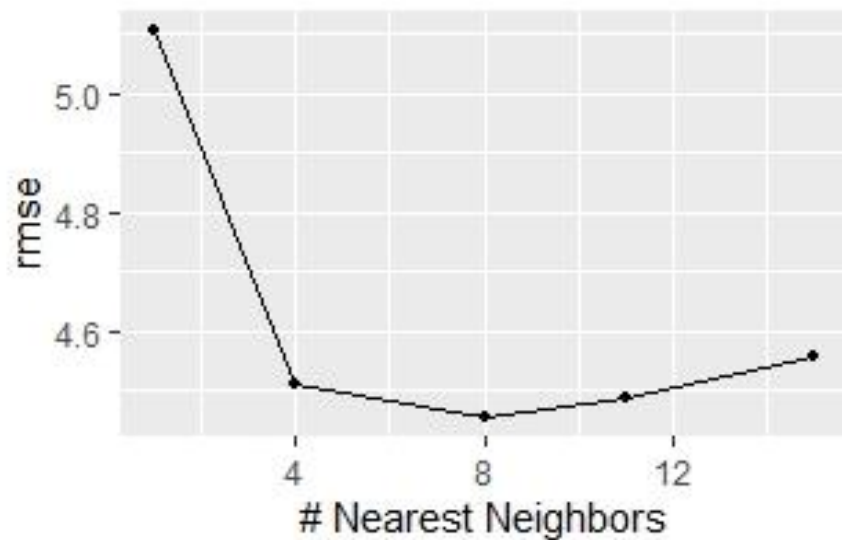
Initial Models Result

Boosted Tree

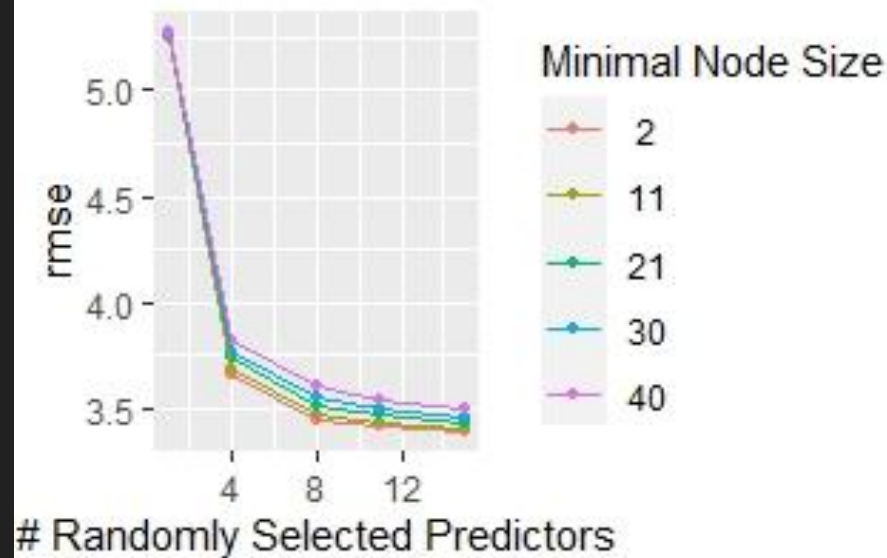


Initial Data Results

K-Nearest Neighbor



Random Forest



Final Model Result

Results of Null Model

	<code>.metric</code>	<code>.estimator</code>	<code>.estimate</code>
	<code><chr></code>	<code><chr></code>	<code><dbl></code>
1	mape	standard	30.9
2	rmse	standard	6.73

Result of RF on Test

```
# A tibble: 2 x 3
```

	<code>.metric</code>	<code>.estimator</code>	<code>.estimate</code>
	<code><chr></code>	<code><chr></code>	<code><dbl></code>
1	rmse	standard	3.05
2	mape	standard	12.3

Conclusion

- Analysis
 - No overfitting
 - $>10\%$ MAPE
- Rooms for improvement
 - More predictors
 - Selective predictors
 - `Step_interact()`