

K.I.S.S.

...

Author: Will J. Suh

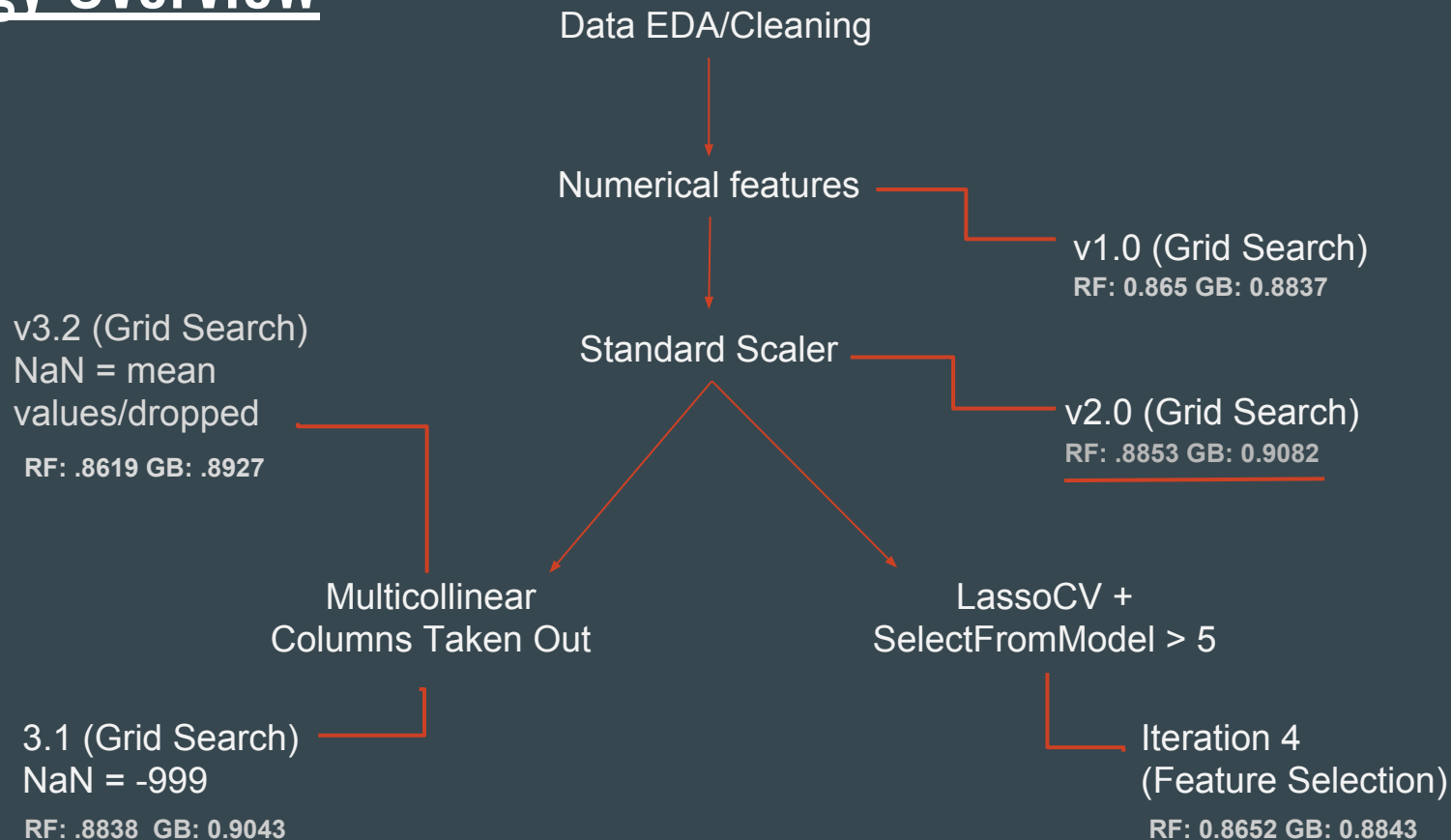


99 Problems but a NaN Ain't One

1. 81 features
2. Many missing values and a lot of noise
3. Numerical and object categories
 - a. Divide categories so I can avoid dealing with “True nulls”
4. Unscaled data



Strategy Overview



Optimizations

- v2.0 performed the best.
- Experiment with removing features or feature engineering then run Gradient Boosting model.
- Try XG Boost
- Experiment incorporating categorical features into my model.



Concluding Thoughts

1. Gradient Boosting performed better than Random Forest for all versions.
2. Using Grid Search to select best family of models helped a lot.
3. Less is more (sometimes*)

