

K.I.S.S.

...

Author: Will J. Suh

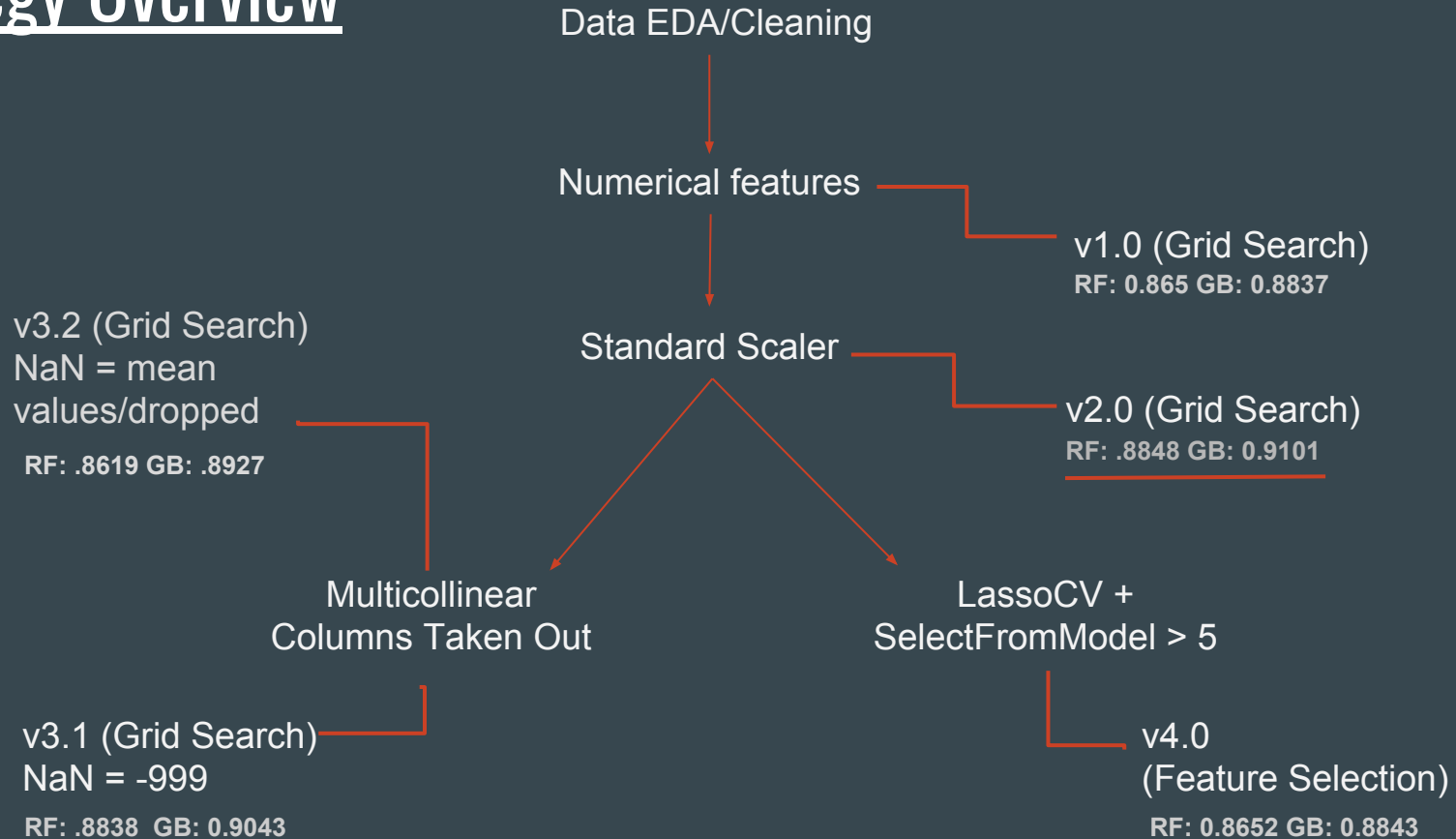


99 Problems but a NaN Ain't One

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



Takeaways

- On the surface v2.0 performed the best.
 - v2.0: **36** features and scored .91
 - v4.0: **5** features and scored .88
- Start quick and dirty → Optimize
 - Easier to defend
 - Easier to explain
 - Increased transparency



Next Steps

1. XG Boost
2. Feature Engineering
3. Less is more (sometimes*)

