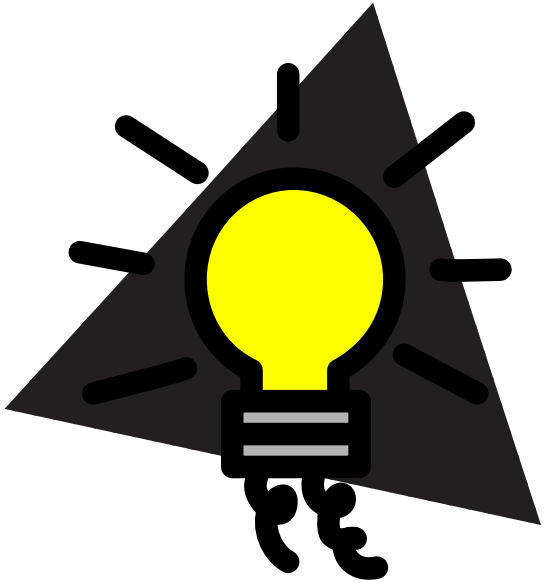
Water pump distribution in Tanzania

Rahul Krishnan

Presentation

Business Objective

To build a model for the Govt of Tanzania that will help predict the status of a water pump based on certain input information.



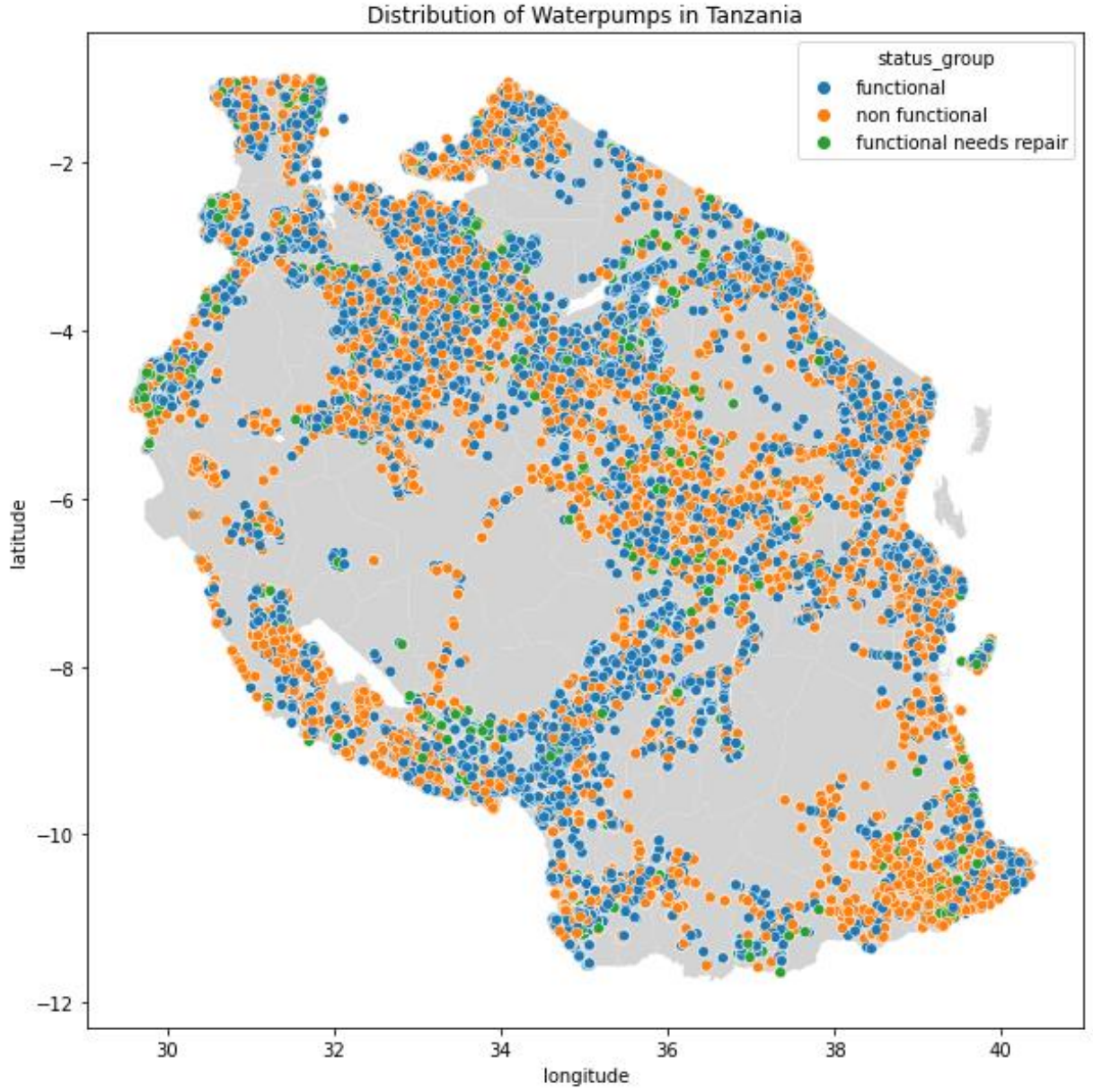
The water pump will be classified as follows:

1. Functional
2. Non-Functional
3. Functional needs repair

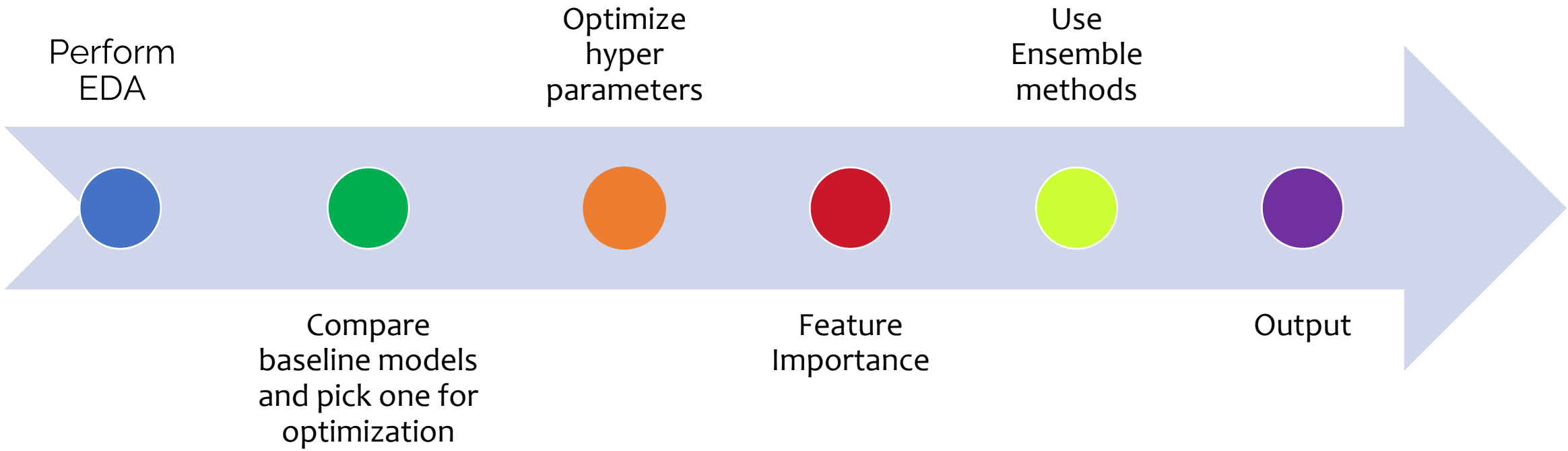
Dataset

Dataset sourced from : https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/page/23/

Water pump distribution by class



Process Steps

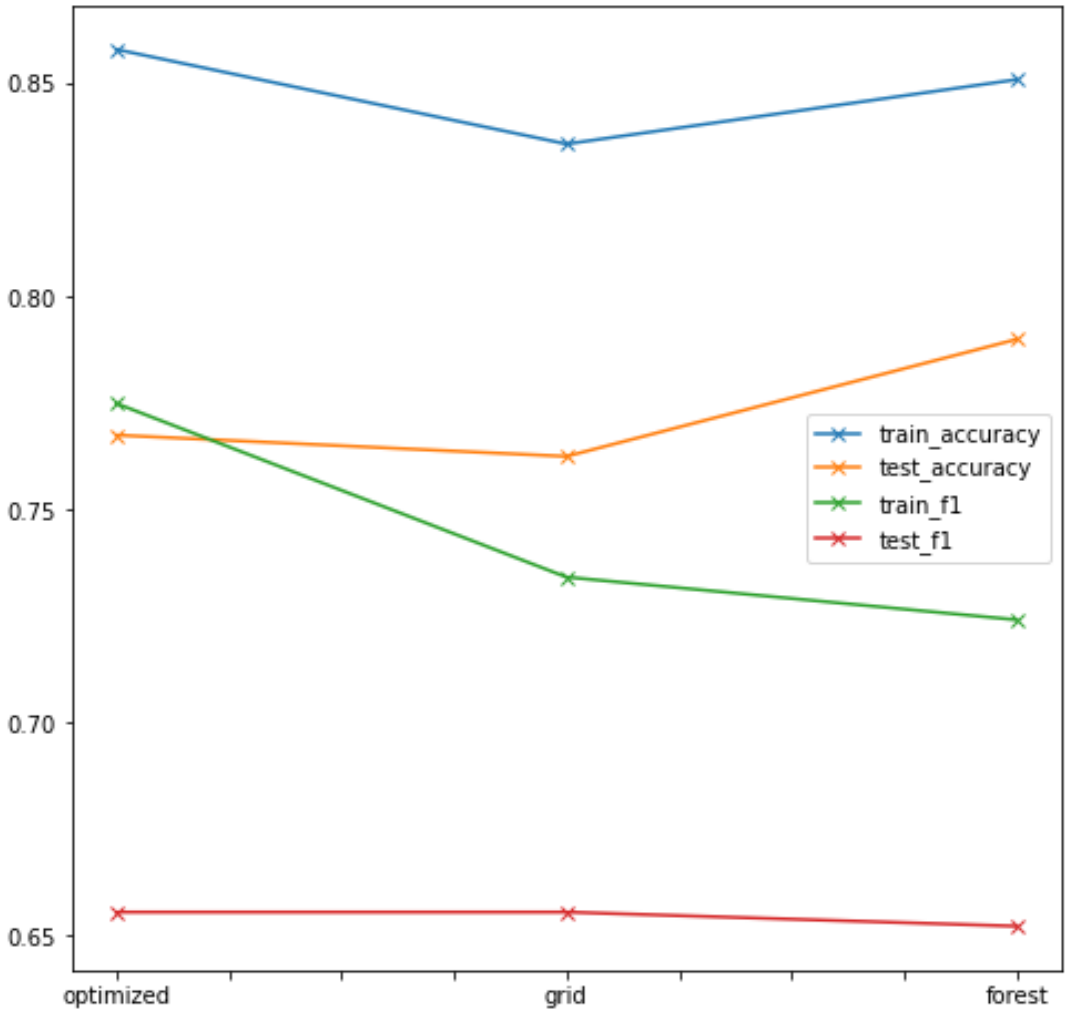


Baseline models stats

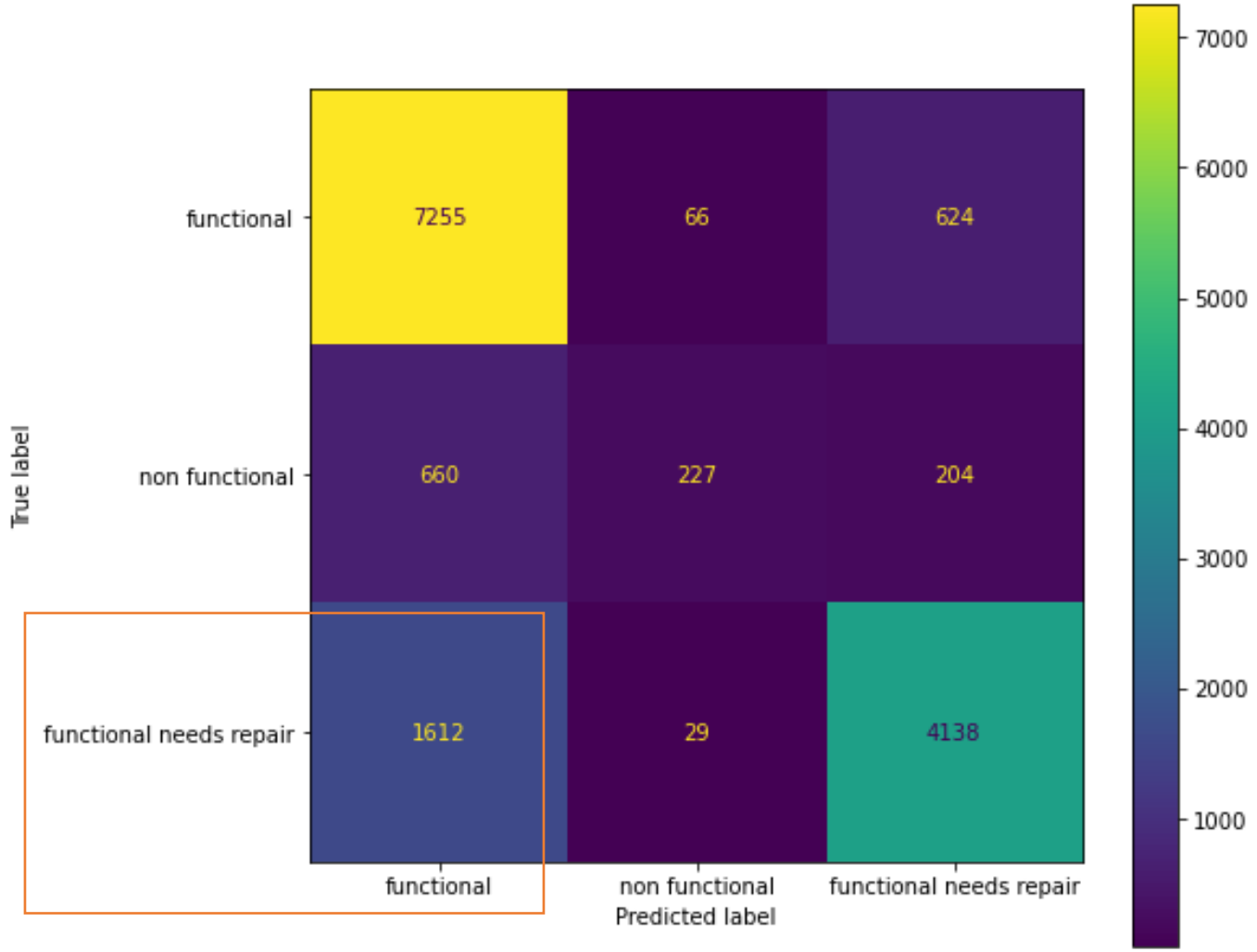
Since this is a multi-classification problem, we will look at the F1-score , which combines both accuracy and recall.

	Logistic Regression	Decision Tree	KNN
F1 - Score	0.33	0.64	0.49

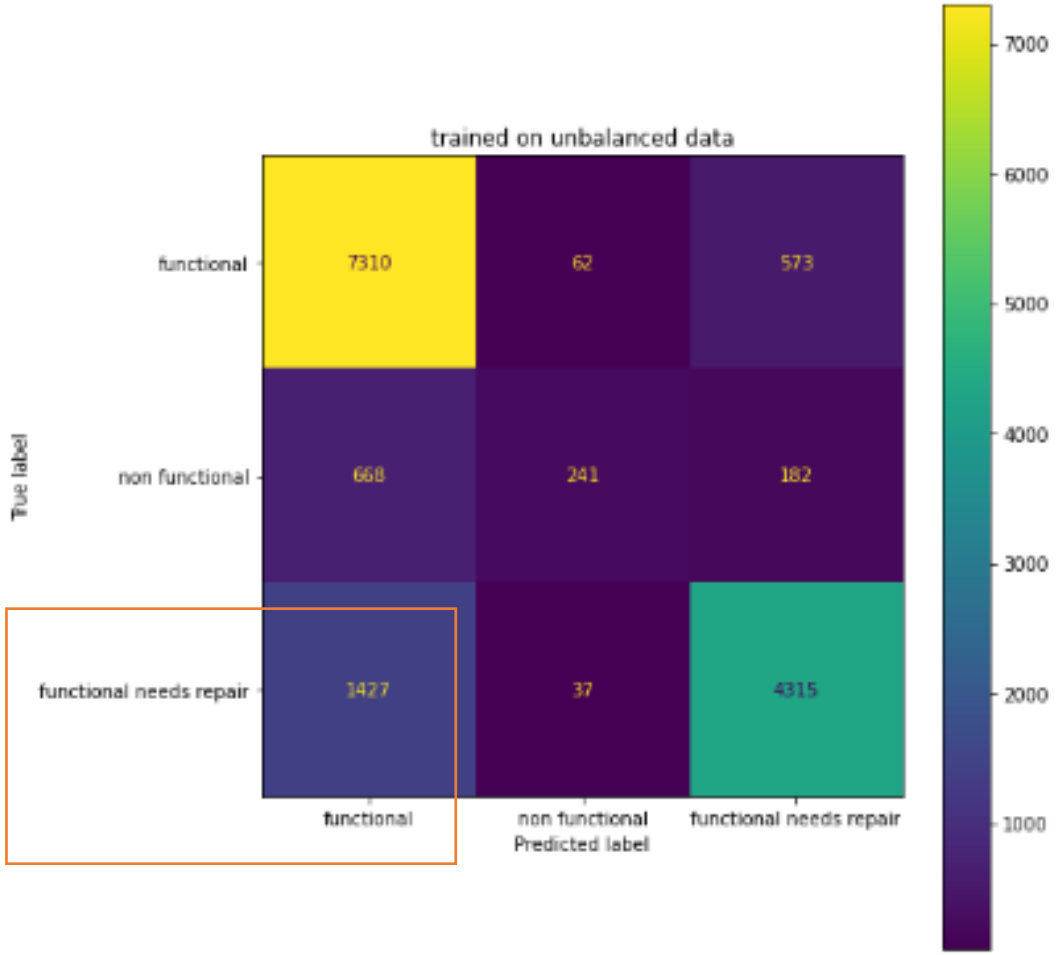
Compare Optimized, GridSearch and RandomForest models

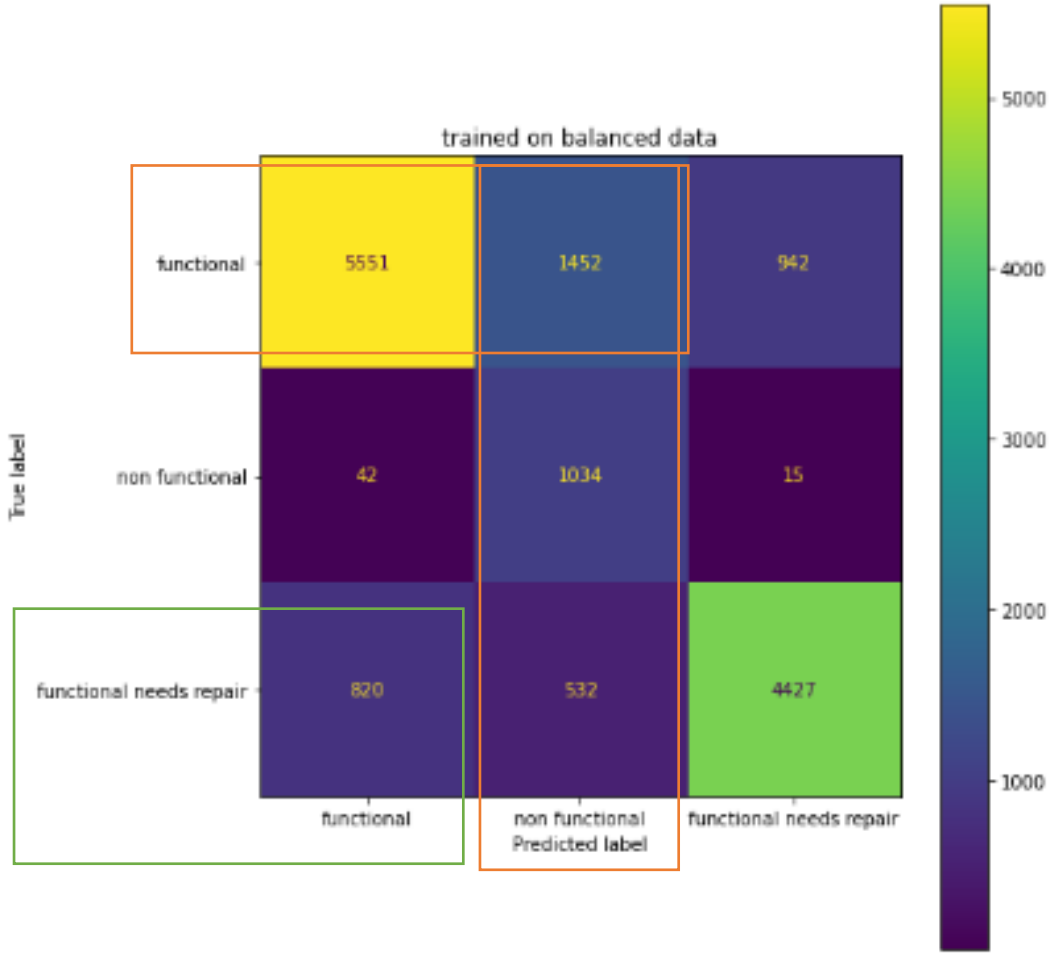


Confusion Matrix – Random Forest

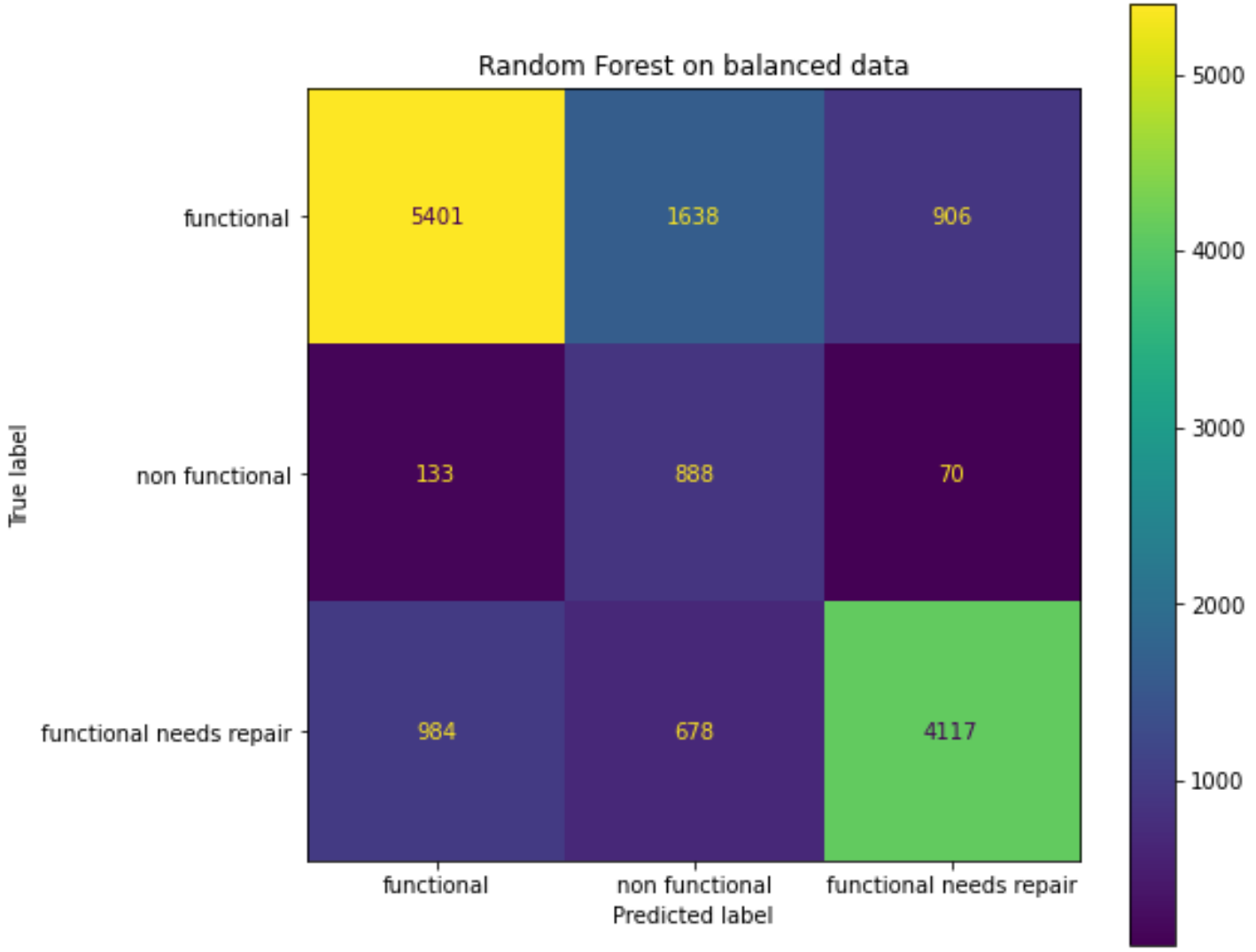


Balanced vs Unbalanced





Final Results



Next Steps

- Re-frame this as a binary classification problem i.e., functional vs non-functional and see if we can build a better model.
- Optimize balanced dataset models.

Thank You!

This Presentation is Prepared by

Rahul Krishnan