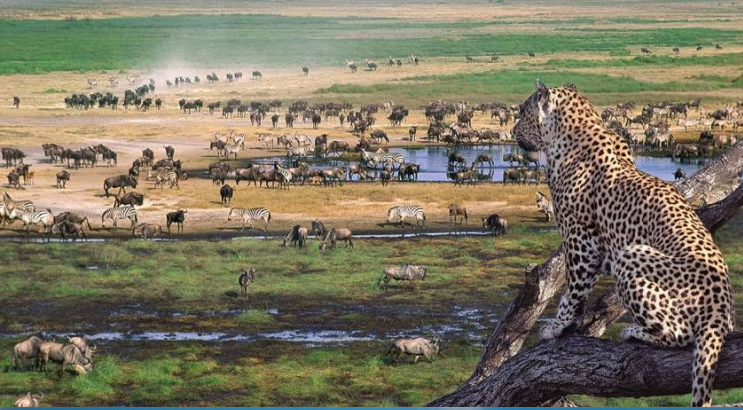




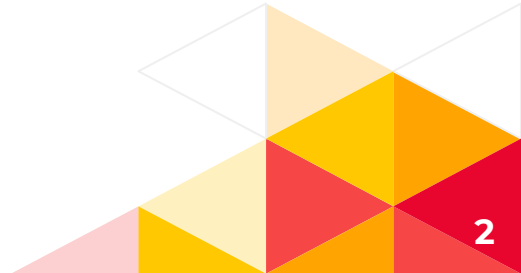
The Circle of Water

Srikrishna Venkat

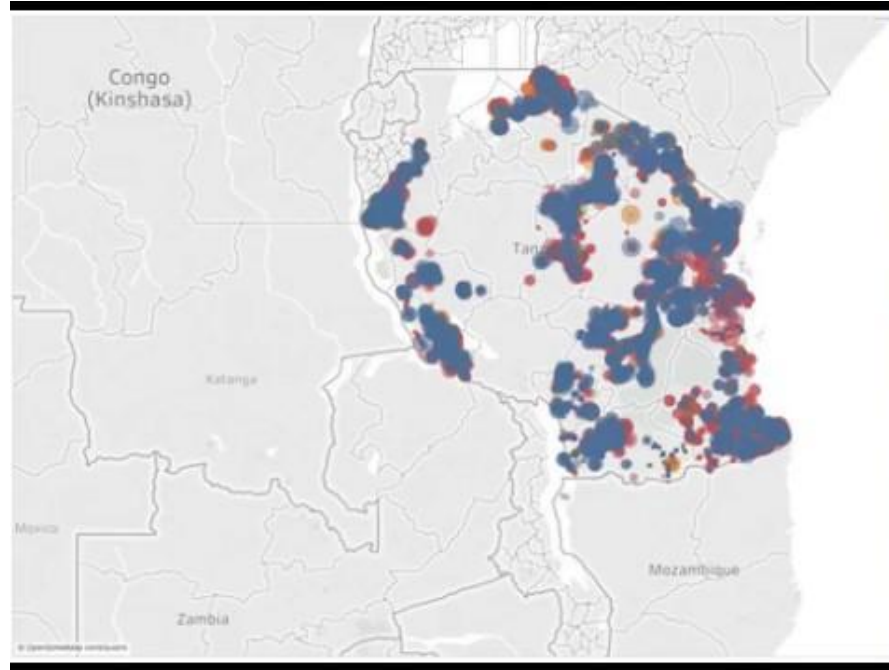


Facts about Tanzania

- ◀ Located in South Eastern Africa.
- ◀ 57 million inhabitants.
- ◀ 70 % is rural with ~20,000 villages.
- ◀ Water scarcity is biggest issue.



Geographical Distribution of Water Pumps



Data Sources, Features and Tools

- ▶ Driven Data
- ▶ Taarifa and Tanzania Ministry of Water
- ▶ 40 features -> 94 features
- ▶ Multiclass classification

Pandas

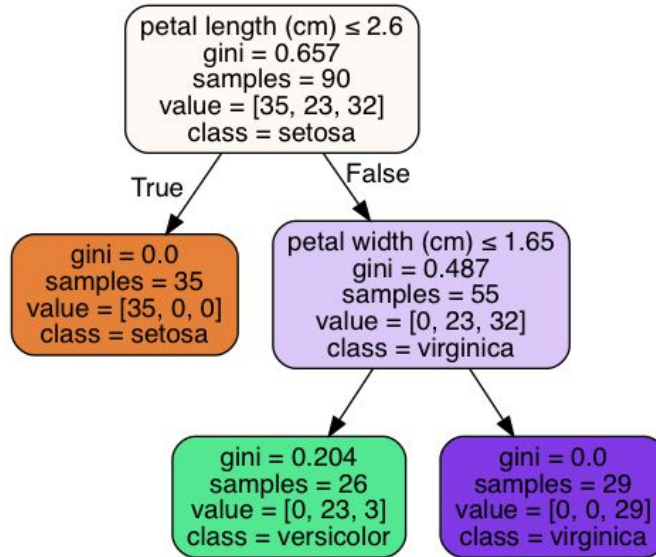


matplotlib

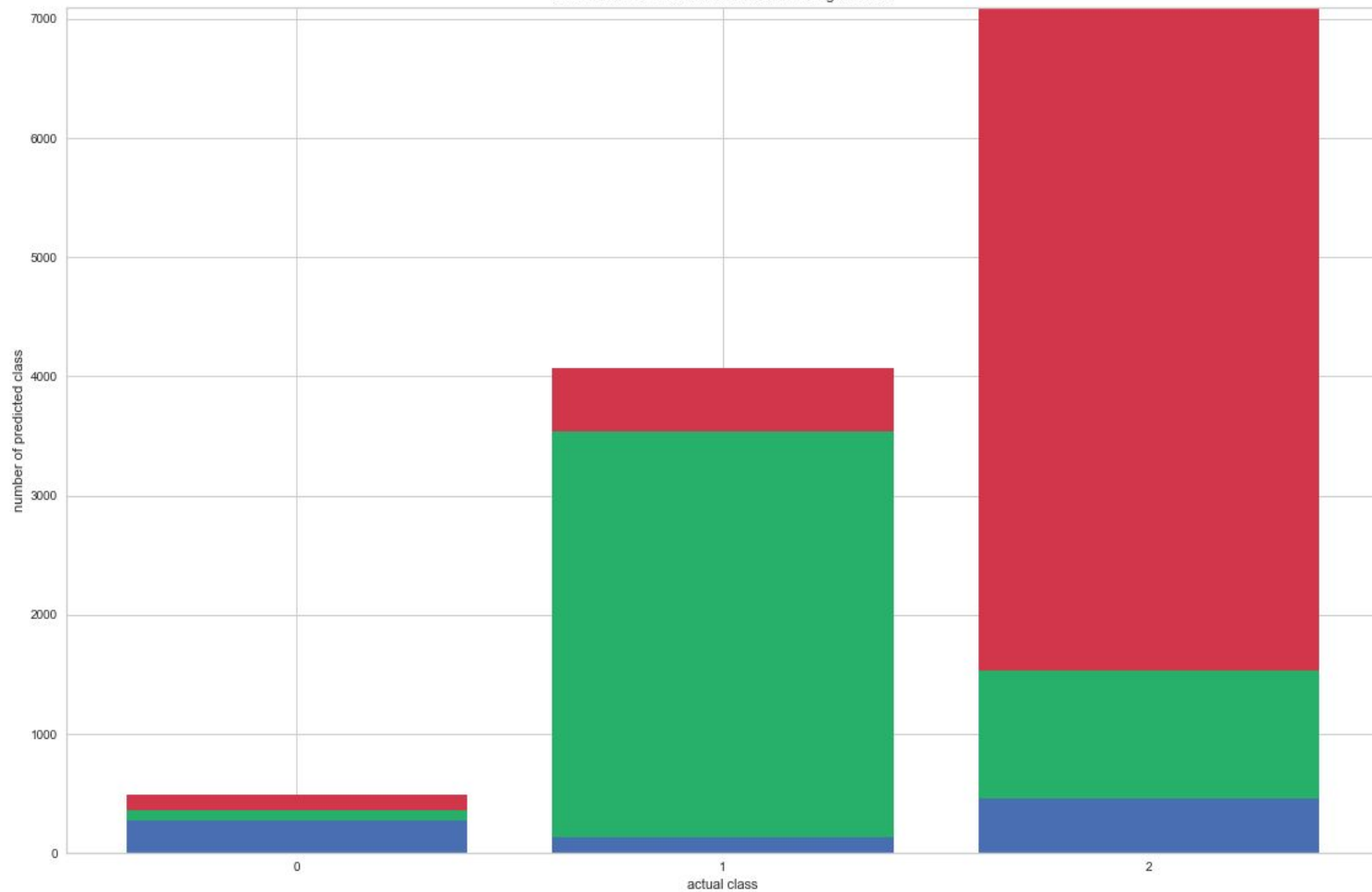


Modelling

- ▶ GradientBoostingClassifier in scikit-learn
- ▶ Ensemble method using Decision trees.
- ▶ 79.93% accuracy!



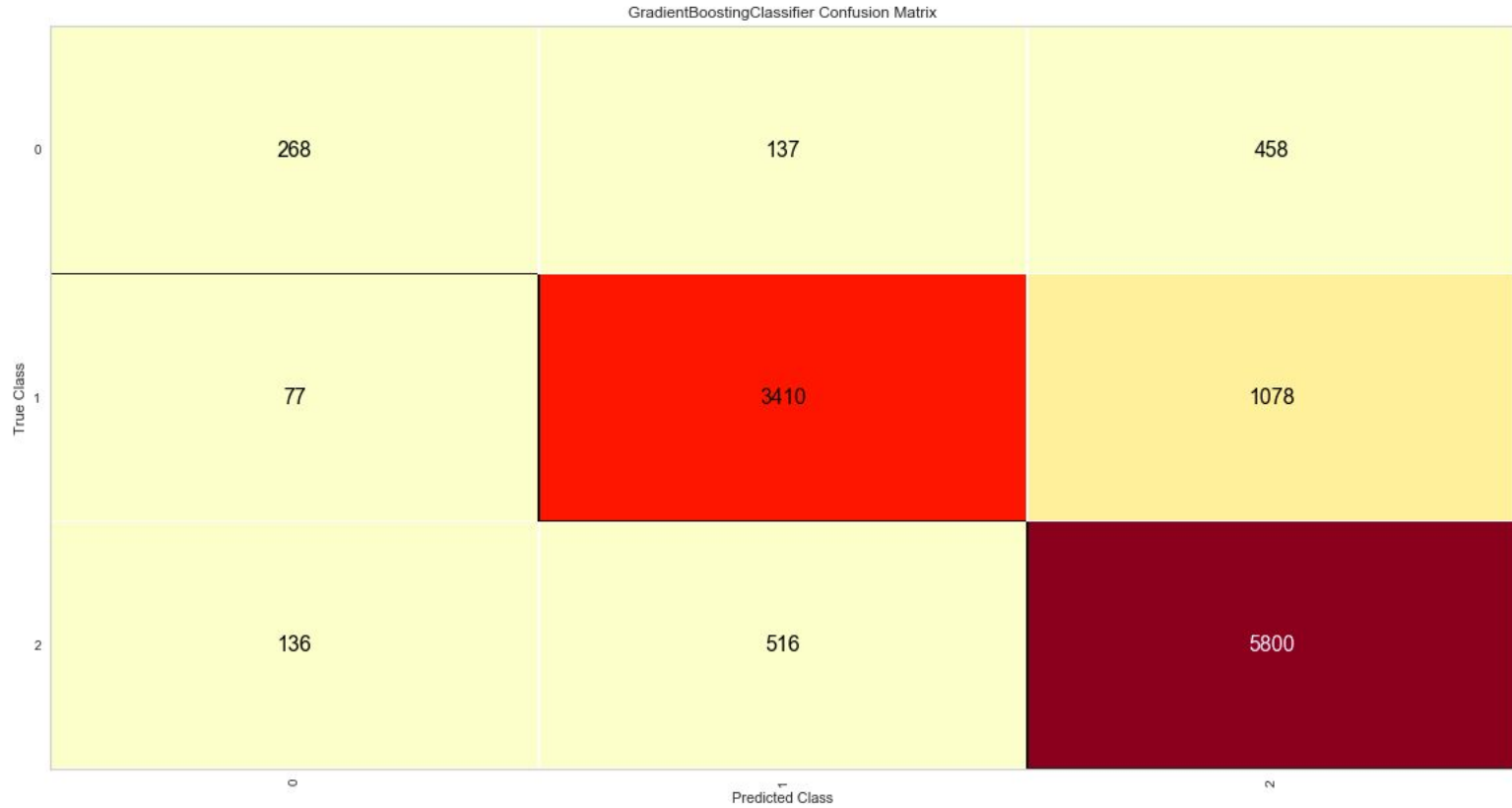
Class Prediction Error for GradientBoostingClassifier

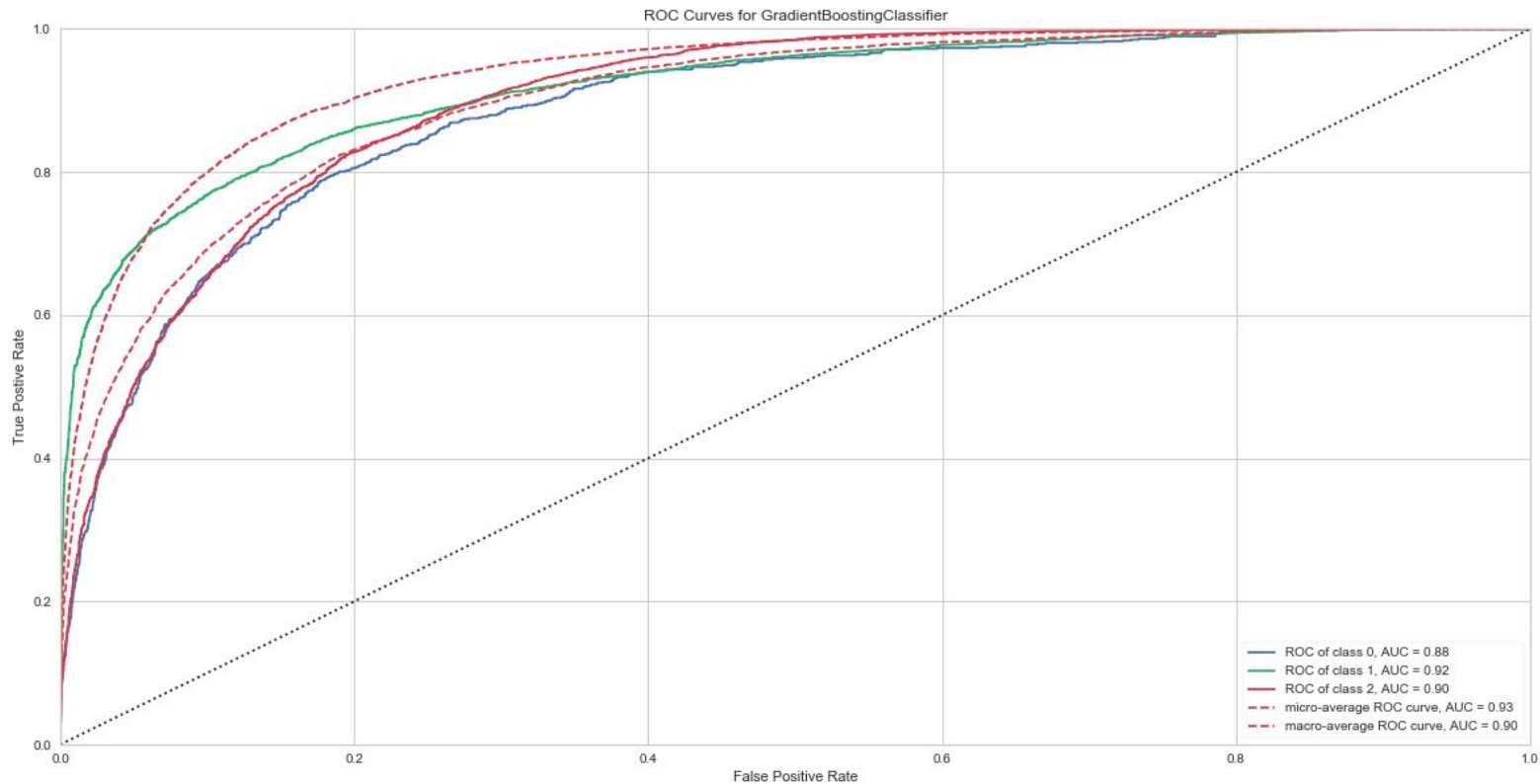


Class Prediction
Error

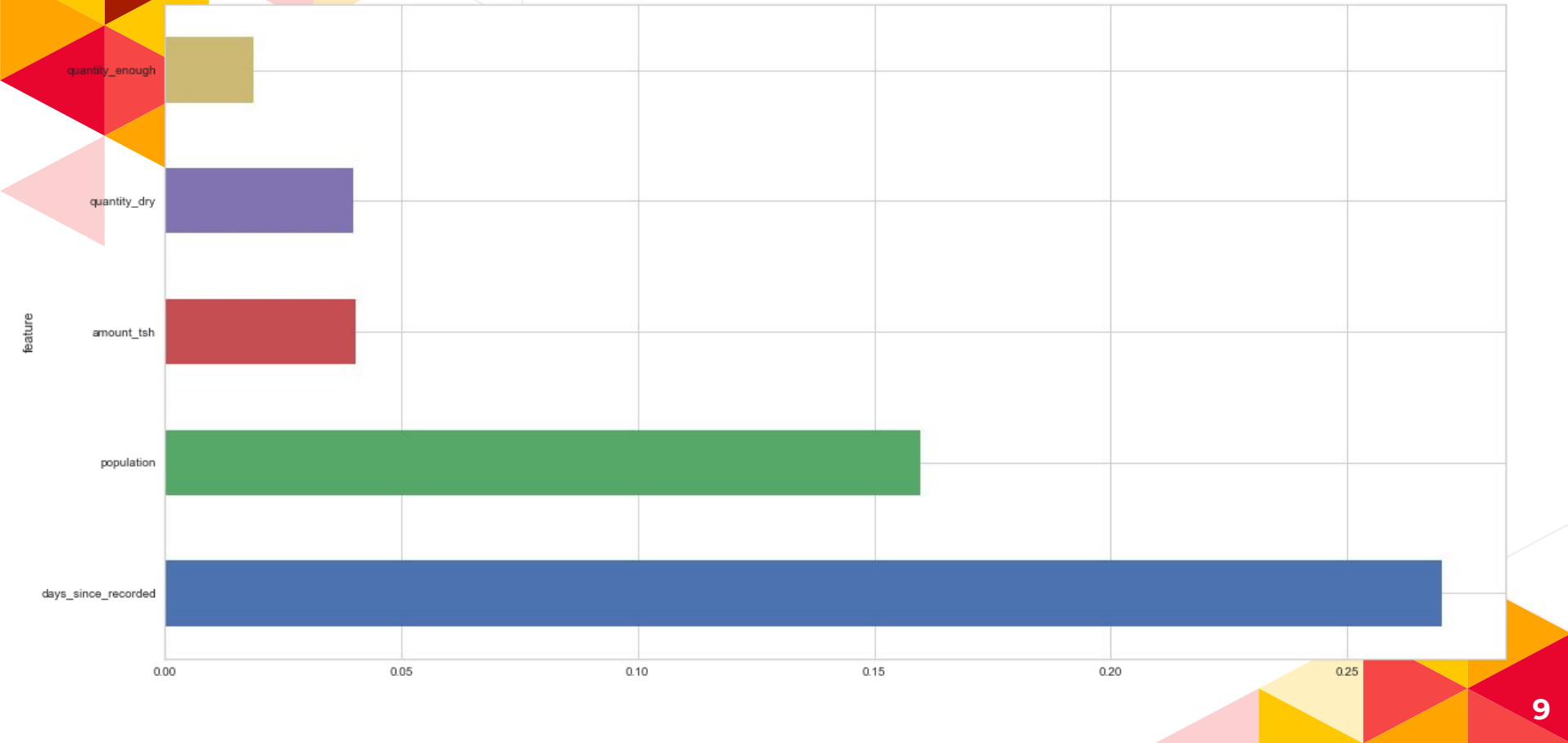
0
1
2

Confusion Matrix





Model Feature Importance





Summary

- ▶ Model does a good job of classification
- ▶ Has trouble recognizing, 'Functional Needs Repair' class
- ▶ Features: Age, population and quantity.

