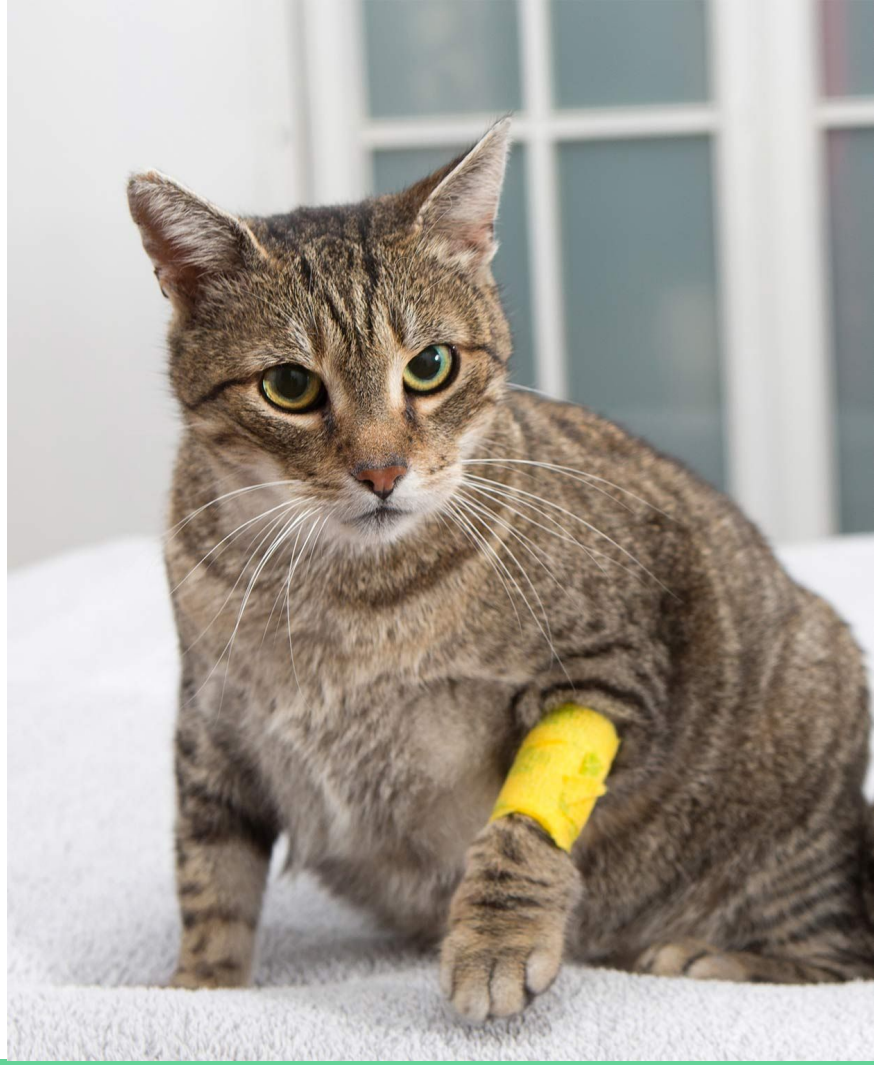
Predicting Shelter Animal Outcomes

Luke Tibbott

Motivation

- Data science for good

- Tool for shelters



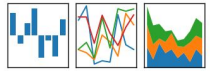
Workflow and tools

Data → Analysis → Model → Engineering

kaggle

AUSTIN ANIMAL CENTER

pandas



NumPy

python™

scikit learn

CatBoost

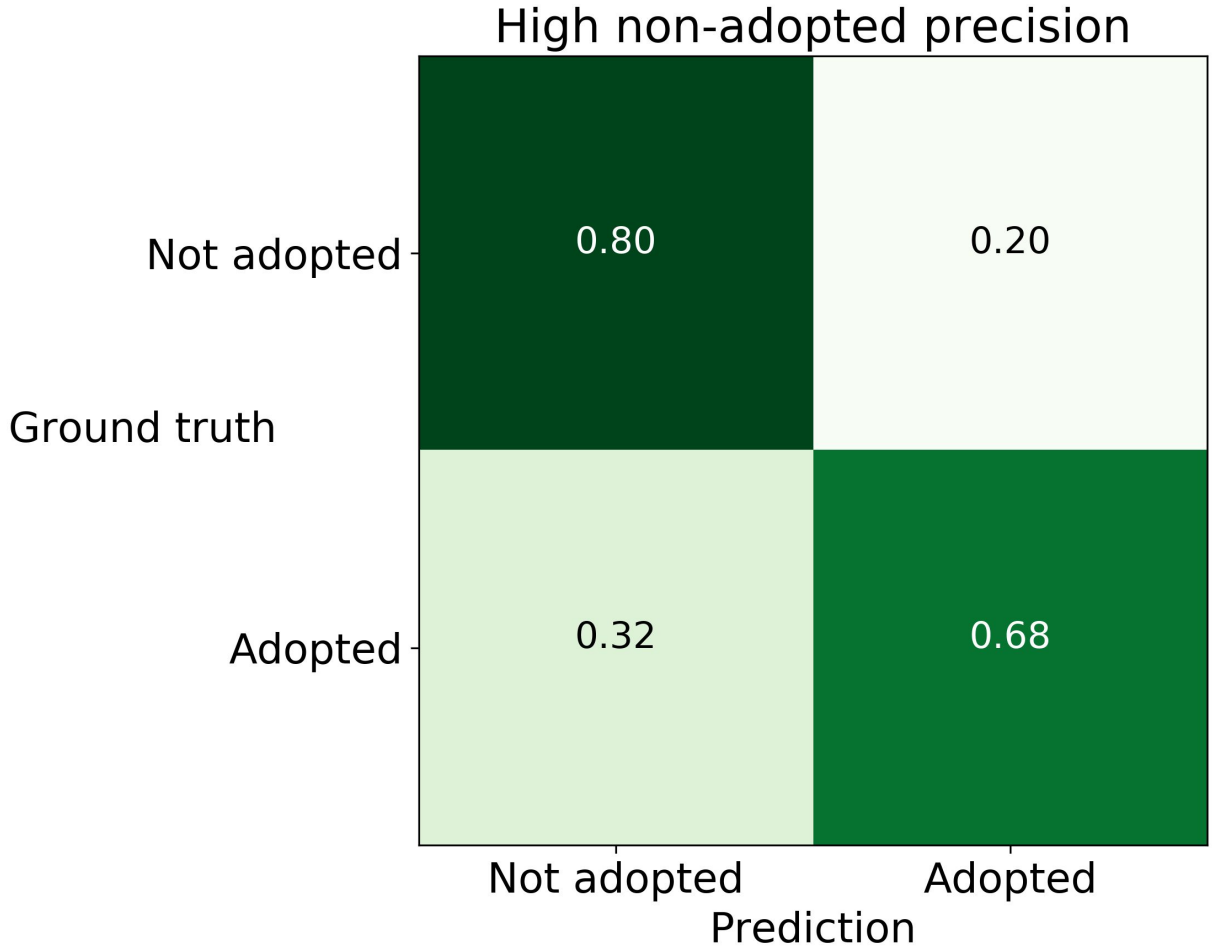
matplotlib

Flask

git

Results

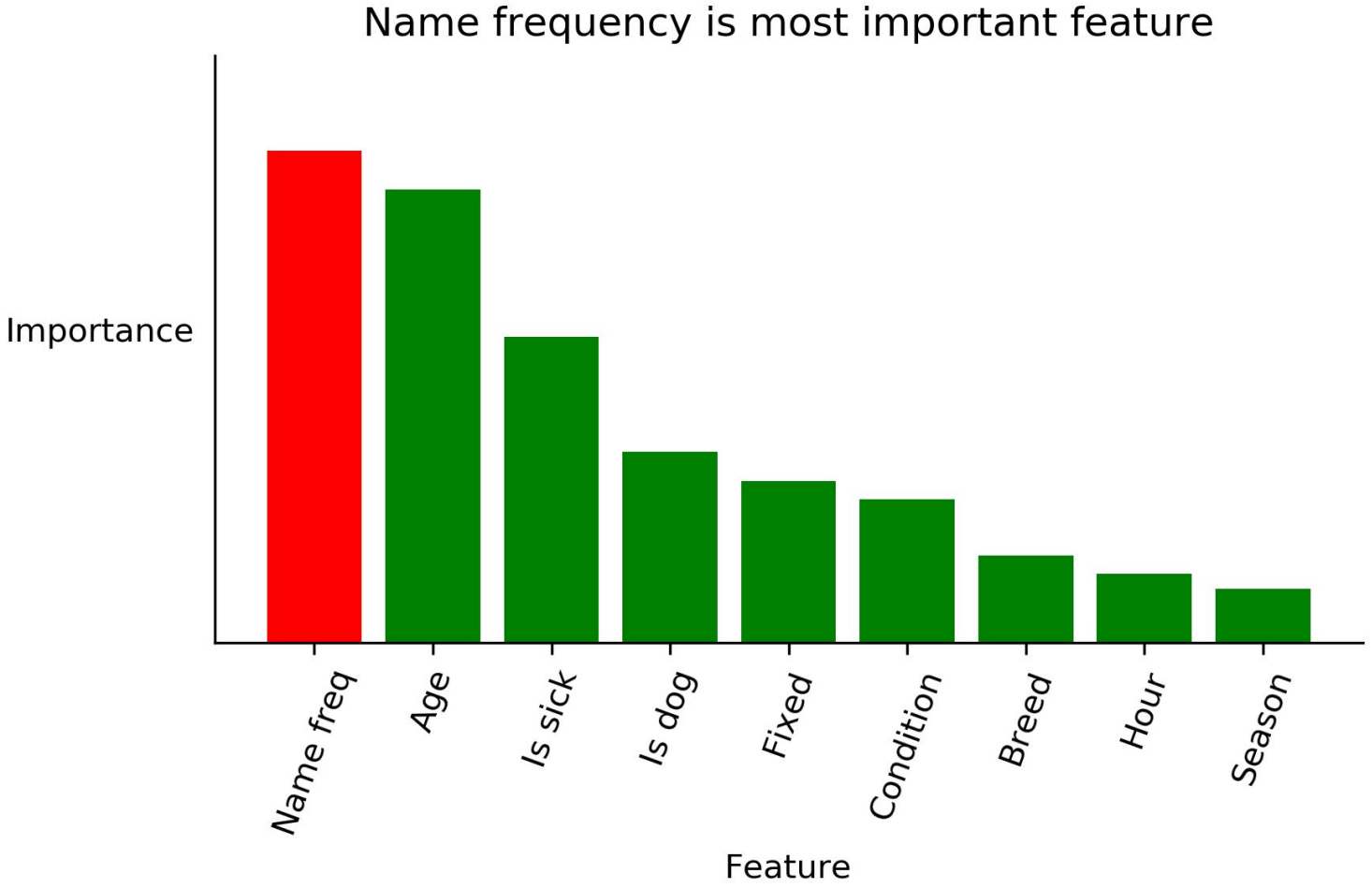
- .80 precision for non-adopted
- Less precise for adopted



Conclusions

- Good at identifying less adoptable animals

- Name is most important


Name frequency is most important feature
Importance
Name freq
Age
Is sick
Is dog
Fixed
Condition
Breed
Hour
Season
Feature

Future Work

- Real world usage
- Weight data
- Days in shelter regression

Thank you!

luke.tibbott@gmail.com

luketibbott

luke-tibbott