# Classification of Endangered Species

Module 3: Machine Learning

Adam Dick // FLATIRON SCHOOL

## Features

Species Group

State Distribution

Vertebrate
Invertebrate
Plant

U.S. Fish & Wildlife Service

9,612 Species Records x 15 Categorical Features

## Classifiers

Dummy
Logistic Regression
K Nearest Neighbors
Decision Tree
Random Forest

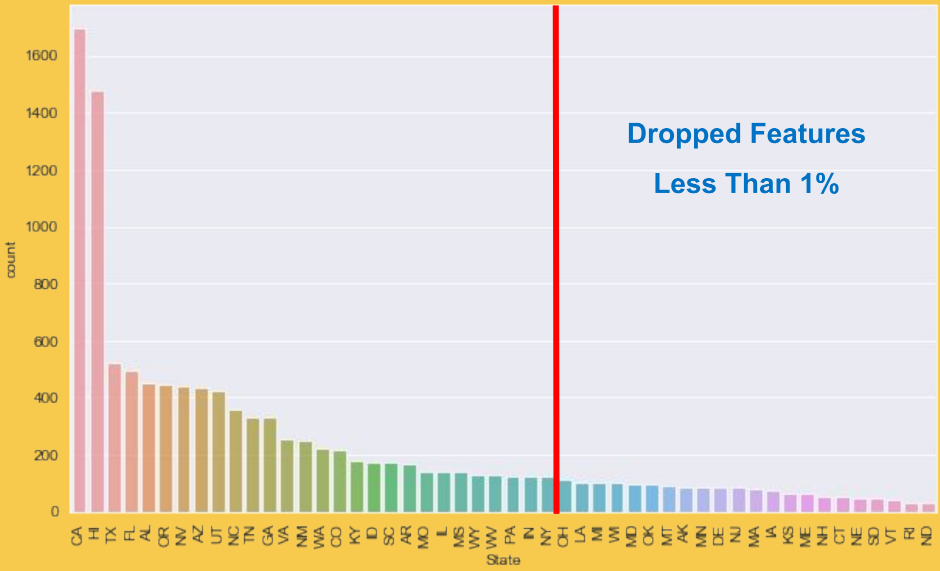
## Target

### Federal Listing Status

Endangered
Threatened
Not Listed

# Exploratory Data Analysis of Features

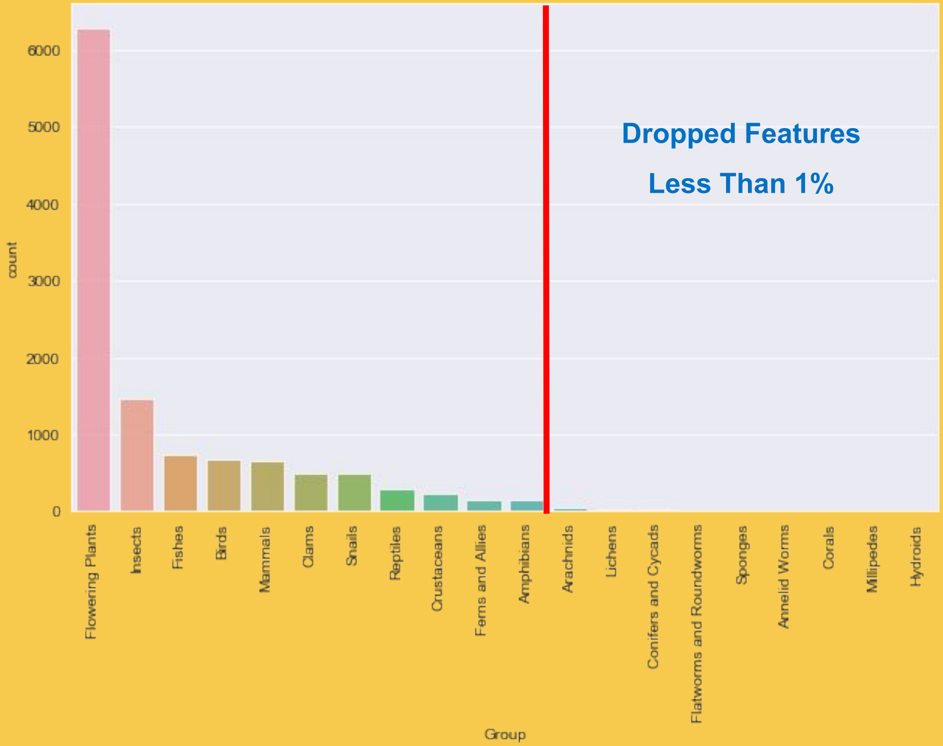
### State Distribution



Dropped Features

Less Than 1%

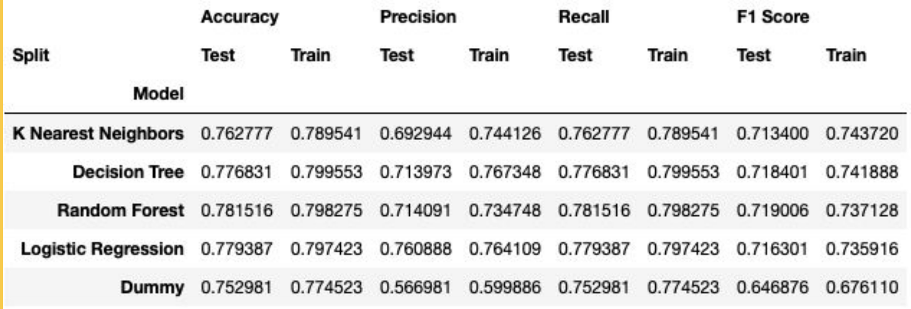
### Species Group



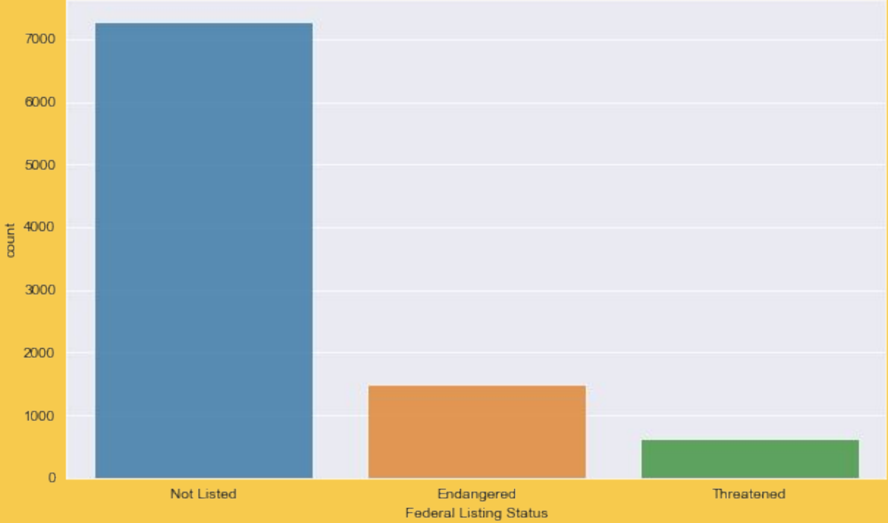
Dropped Features

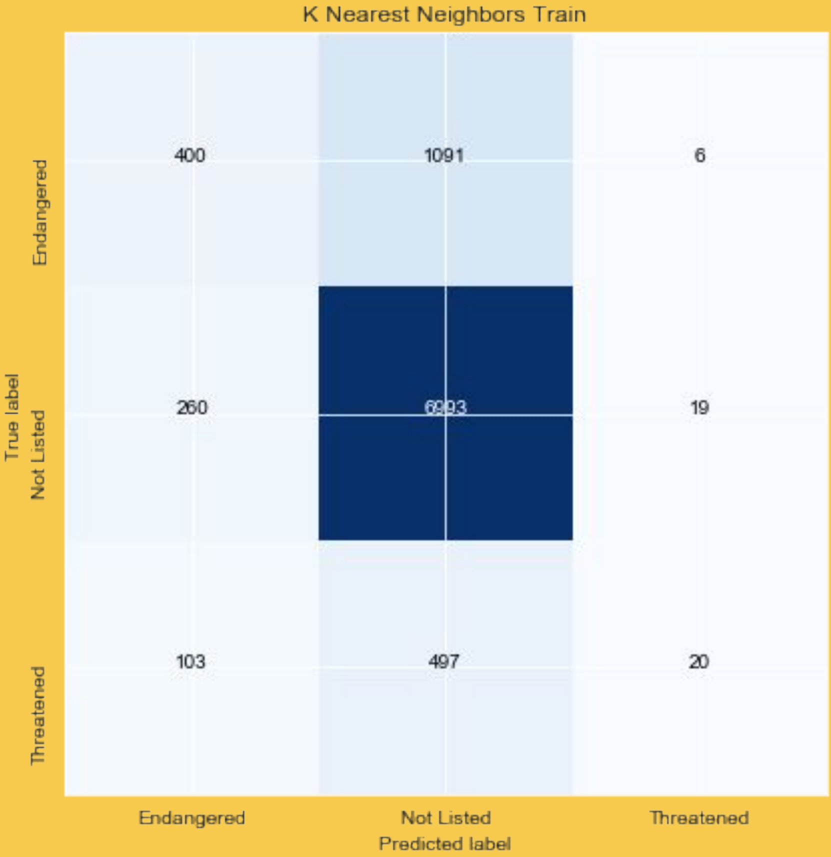
Less Than 1%

# Baseline Classifiers



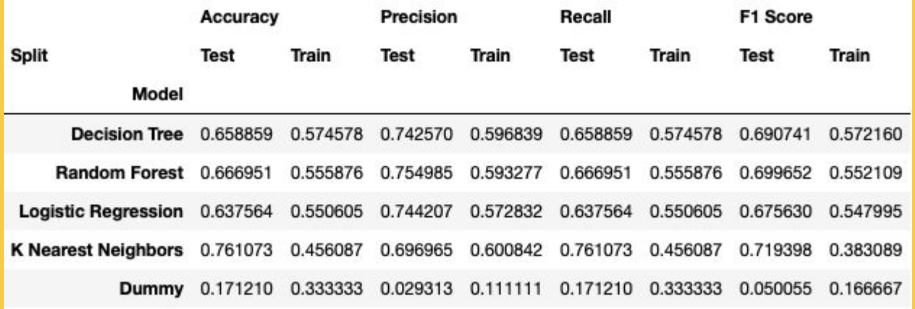
|  | Accuracy |  | Precision |  | Recall |  | F1 Score |  |
|---|---|---|---|---|---|---|---|---|
| Split | Test | Train | Test | Train | Test | Train | Test | Train |
| Model |  |  |  |  |  |  |  |  |
| K Nearest Neighbors | 0.762777 | 0.789541 | 0.692944 | 0.744126 | 0.762777 | 0.789541 | 0.713400 | 0.743720 |
| Decision Tree | 0.776831 | 0.799553 | 0.713973 | 0.767348 | 0.776831 | 0.799553 | 0.718401 | 0.741888 |
| Random Forest | 0.781516 | 0.798275 | 0.714091 | 0.734748 | 0.781516 | 0.798275 | 0.719006 | 0.737128 |
| Logistic Regression | 0.779387 | 0.797423 | 0.760888 | 0.764109 | 0.779387 | 0.797423 | 0.716301 | 0.735916 |
| Dummy | 0.752981 | 0.774523 | 0.566981 | 0.599886 | 0.752981 | 0.774523 | 0.646876 | 0.676110 |



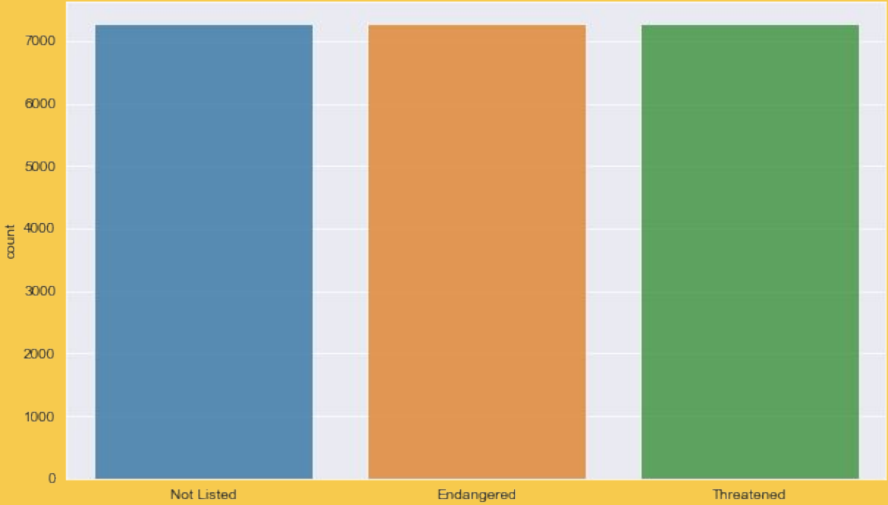


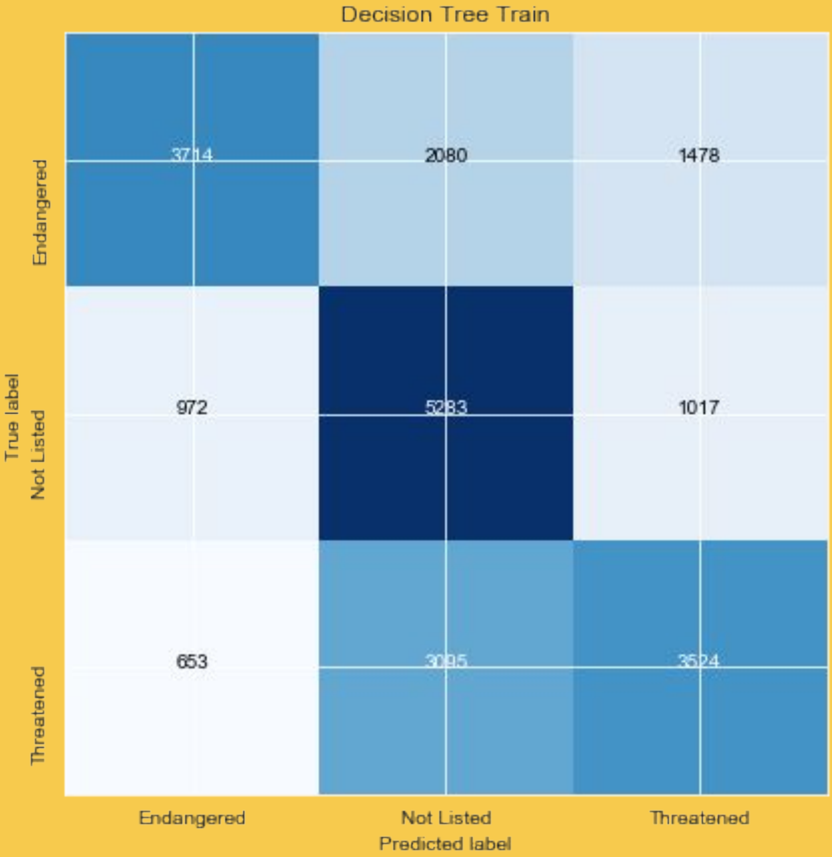
**11,737 Species Records x 42 Dummy Features**

# Class Imbalance with SMOTE Oversampling



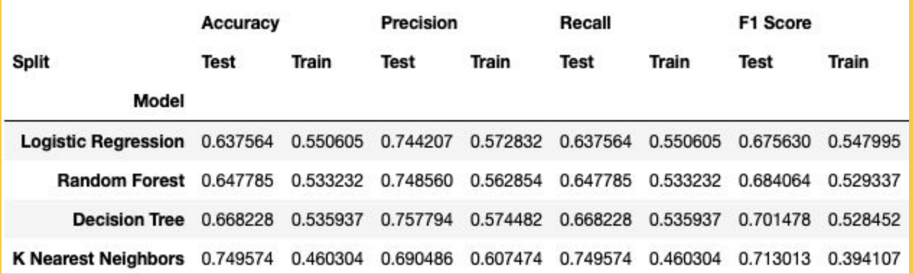
| | Accuracy | | Precision | | Recall | | F1 Score | |
|---|---|---|---|---|---|---|---|---|
| Split | Test | Train | Test | Train | Test | Train | Test | Train |
| Model | | | | | | | | |
| Decision Tree | 0.658859 | 0.574578 | 0.742570 | 0.596839 | 0.658859 | 0.574578 | 0.690741 | 0.572160 |
| Random Forest | 0.666951 | 0.555876 | 0.754985 | 0.593277 | 0.666951 | 0.555876 | 0.699652 | 0.552109 |
| Logistic Regression | 0.637564 | 0.550605 | 0.744207 | 0.572832 | 0.637564 | 0.550605 | 0.675630 | 0.547995 |
| K Nearest Neighbors | 0.761073 | 0.456087 | 0.696965 | 0.600842 | 0.761073 | 0.456087 | 0.719398 | 0.383089 |
| Dummy | 0.171210 | 0.333333 | 0.029313 | 0.111111 | 0.171210 | 0.333333 | 0.050055 | 0.166667 |



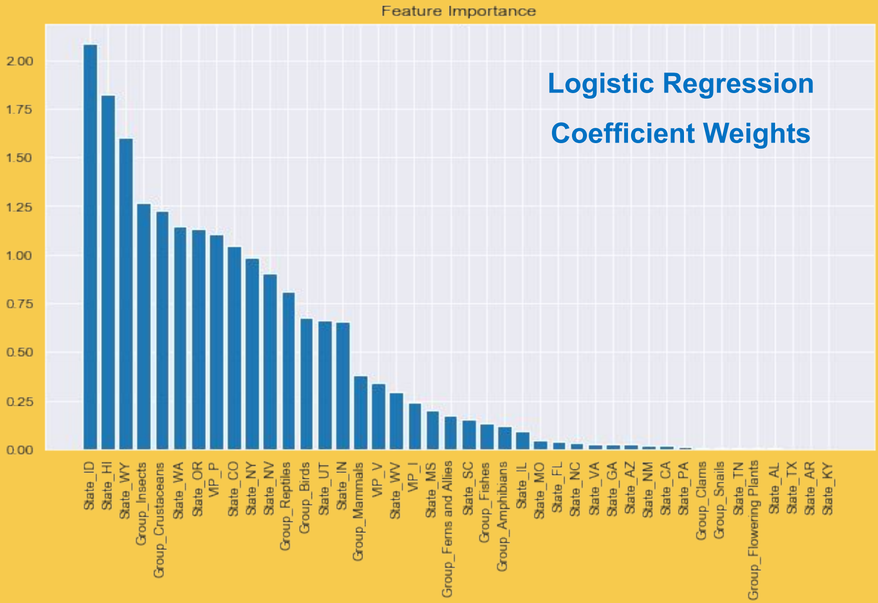


**21,816 Species Records x 42 Dummy Features**

# Grid Search Cross Validation with Balanced Classes

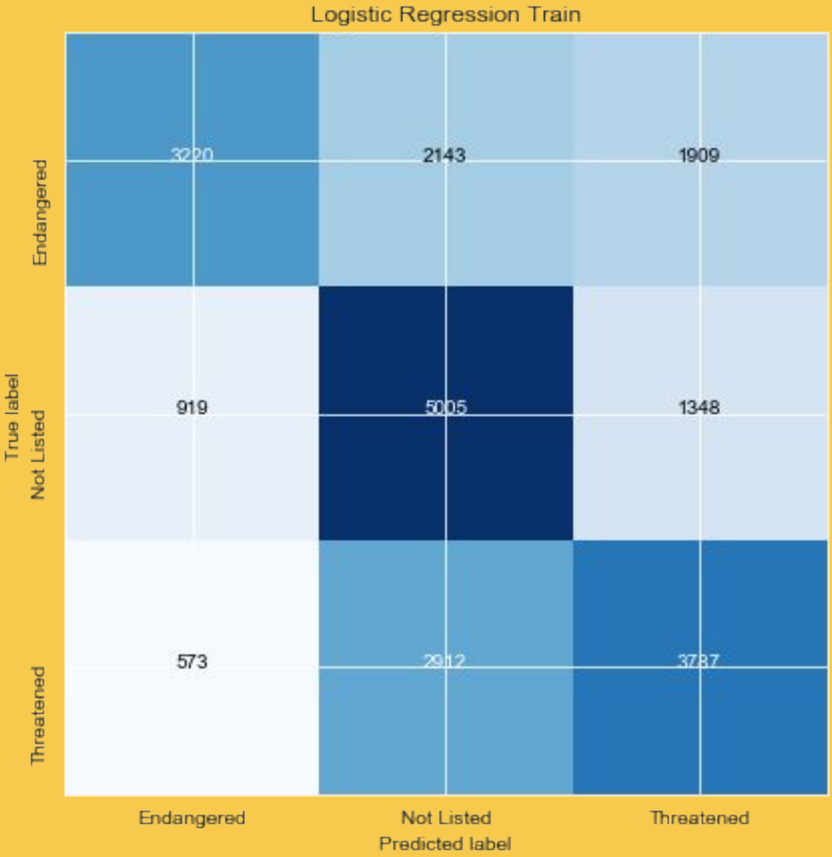


| | Accuracy | | Precision | | Recall | | F1 Score | |
|---|---|---|---|---|---|---|---|---|
| Split | Test | Train | Test | Train | Test | Train | Test | Train |
| Model | | | | | | | | |
| Logistic Regression | 0.637564 | 0.550605 | 0.744207 | 0.572832 | 0.637564 | 0.550605 | 0.675630 | 0.547995 |
| Random Forest | 0.647785 | 0.533232 | 0.748560 | 0.562854 | 0.647785 | 0.533232 | 0.684064 | 0.529337 |
| Decision Tree | 0.668228 | 0.535937 | 0.757794 | 0.574482 | 0.668228 | 0.535937 | 0.701478 | 0.528452 |
| K Nearest Neighbors | 0.749574 | 0.460304 | 0.690486 | 0.607474 | 0.749574 | 0.460304 | 0.713013 | 0.394107 |



**Logistic Regression**

**Coefficient Weights**



# Classification of Endangered Species

Module 3: Machine Learning

Adam Dick // FLATIRON SCHOOL

//FLATIRON SCHOOL