Mushroom Classification

Harry Carter

Business Problem

Messy classification

Inaccessible without tutelage

Dangerous

“The Guide clearly states that there is no simple rule for determining the edibility of a mushroom”

Dataset

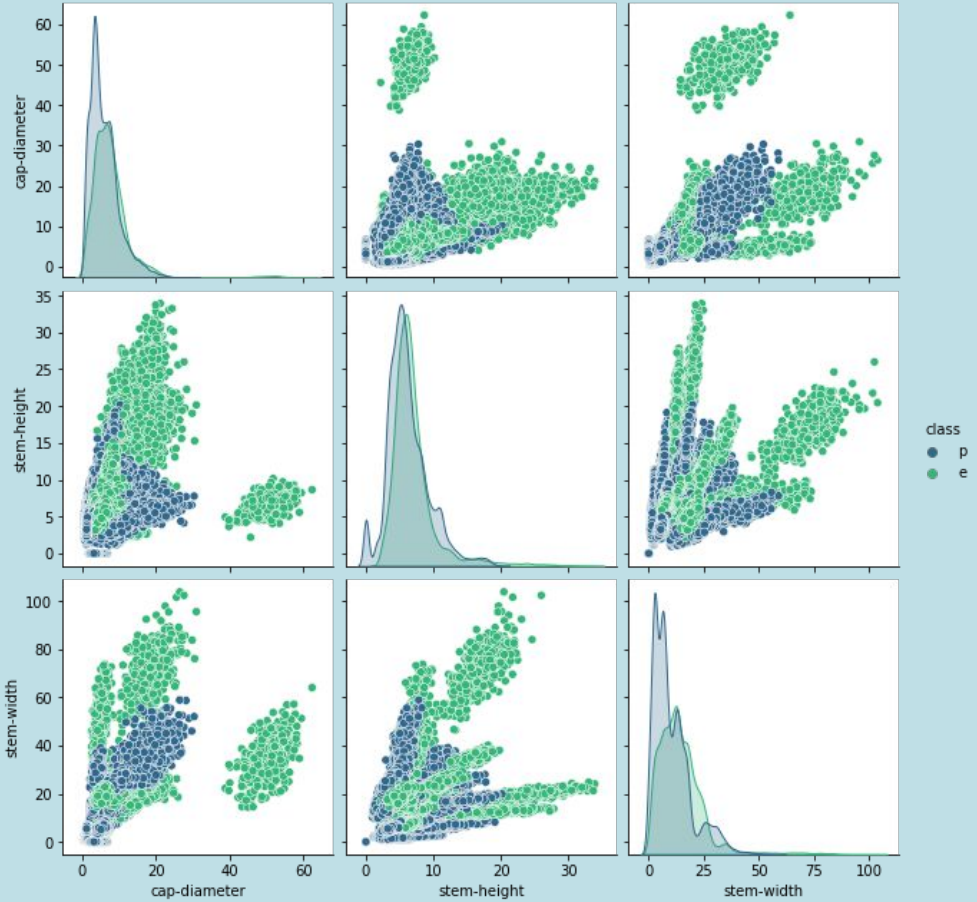
Inspired by transcription of The Audubon Society Field Guide to North American Mushrooms (Knopf)(1981)

Supplemental

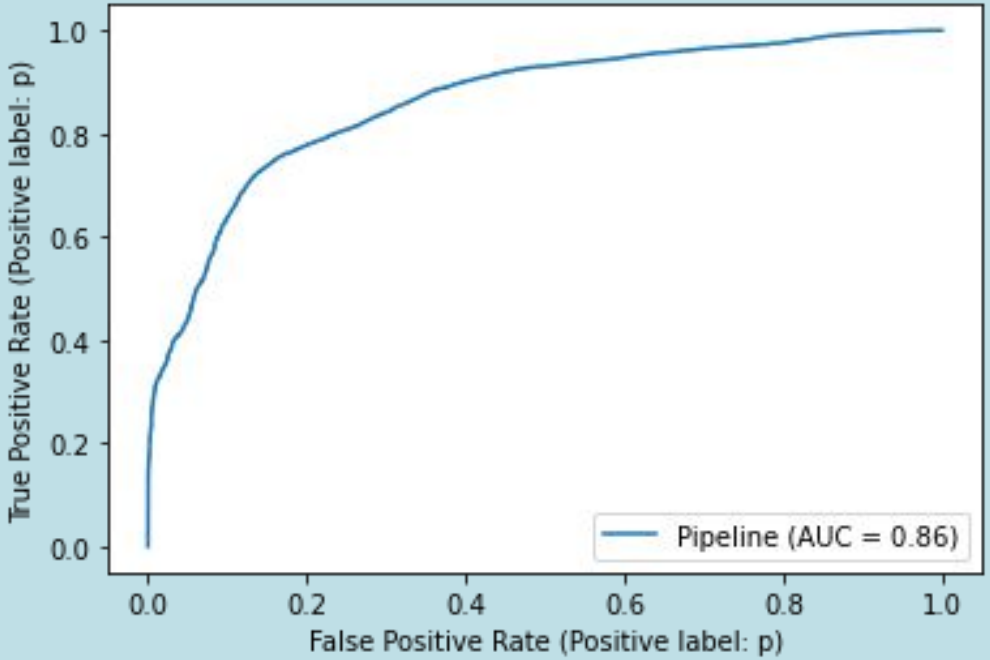
173 species

North American Centric

Synthetic - 353 examples per shroom



Baseline Model

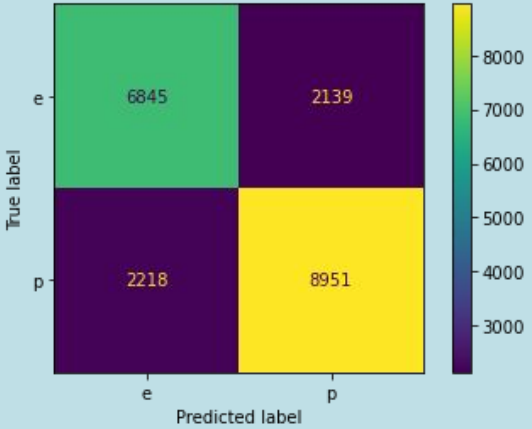


Logistic regression

Accuracy = 87%

Recall = 86%

Not acceptable!

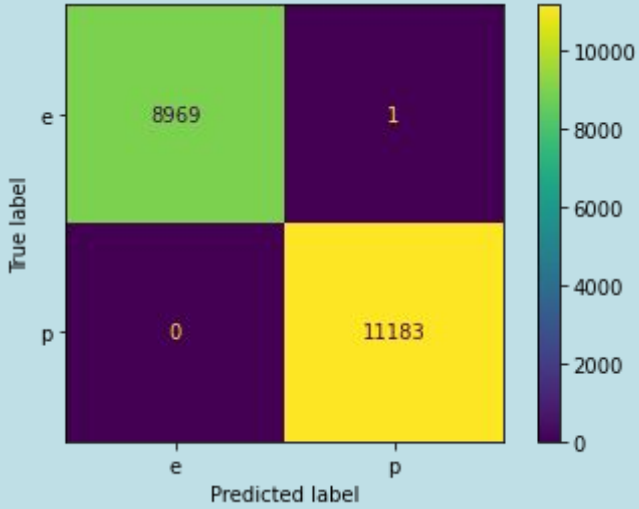


Random Forest Model

Trains at 100%

Tests at 99.995%

1 False negative/positive



Best Predictors:

Ring type at 3.1%

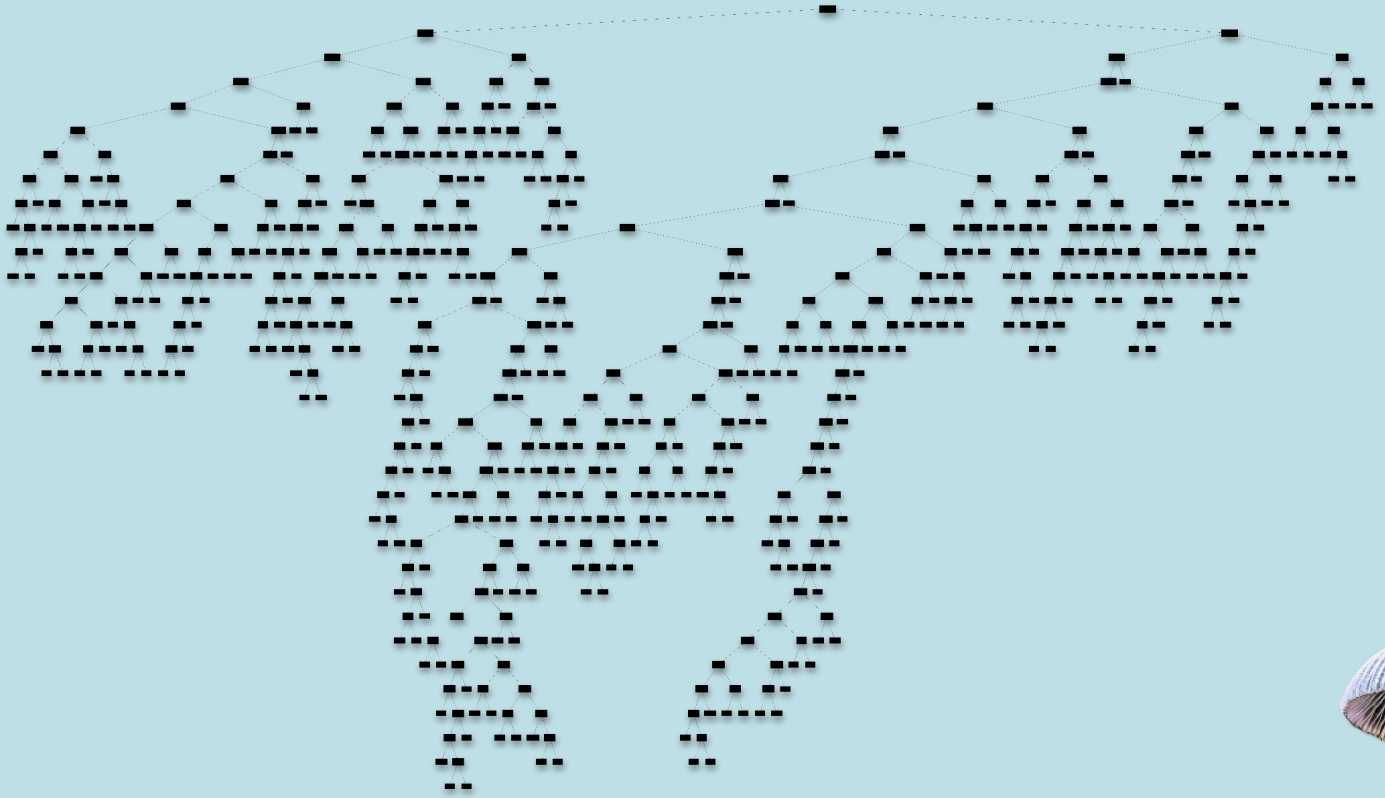
Gill color at 2.9%

Gill attachment at 2.5%

Cap diameter at 2.5%

Stem surface at 2.3%

Tree Visualization



KNN Model

+ No outliers

+/- Large dataset

- Curse of multidimensionality

N neighbors is 2

Lasso regression

32 minutes and 13 seconds

100% accuracy (and recall) on training and testing data.

Business Recommendations

Image credit: Jane Harrison

Pocket resource

App potential

Safety/emergency resource

Predictive power

Further Recommendations

Image Credit: Jane Harrison

Assess predictive power

Update training data

Reduce computational load

Resources

Gary H. Lincoff, The Audubon Society Field Guide to North American mushrooms (Knopf) (1981)

Secondary mushroom dataset data set UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom+Dataset (last visited Jul 15, 2022)