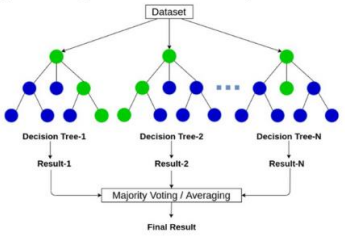
A Statistical Investigation of Rough-Toughed Dolphin Whistle Types

Introduction

- Dolphins have complex vocalizations including clicks, whistles, and burst pulses.
- We are studying variability in dolphin vocalization for the whistles of the rough-toothed dolphin, to help compare communication across different groups.
- We aim to develop a statistical model to classify whistles into categories.

Methods

- ROCCA is a MATLAB-based tool designed for real time species identification.
- ARTwarp is an automated method for categorizing bioacoustic signals
- These tools classify whistles into categories, outputting multiple variables per whistle.
- Random Forest is a machine learning method which builds multiple decision trees, each using different variables.
- A Multinomial GLM extends logistic regression to handle multiple categories instead of just binary classification.



Random Forest Process

Model Comparisons

Multinomial GLM Pros vs Cons

Pros

- It shows the relative importance of each variable in prediction.
- It can help understand variable relationships.

Cons

- There is difficulty when there are too many variables.
- The model may require a large sample size.

Random Forest Pros vs Cons

Pros

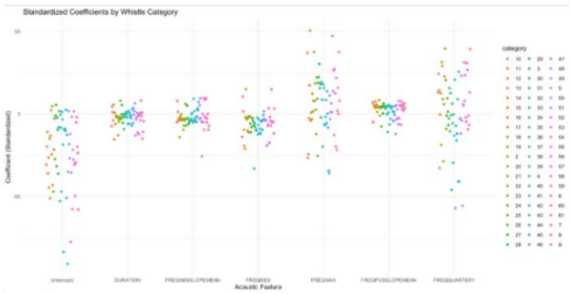
- Captures complex, non-linear patterns.
- It is robust to noise and outliers.

Cons

- It has a limit of 53 categories for classification.
- Doesn't provide p-values or clear coefficients.

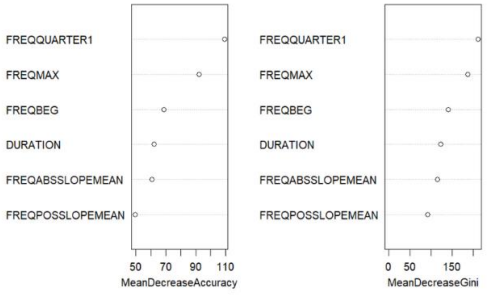
Multinomial GLM Results

- Accuracy = 22.5%
- Residual Deviance = 5304.35



Random Forest Results

- The accuracy of this model was found to be 67.68%



Final Model

- We decided on Multinomial GLM for various reasons.
- It allows for hypothesis testing and confidence intervals
- It is easier to interpret compared to the Random Forest model.
- It is faster to run than Random Forest