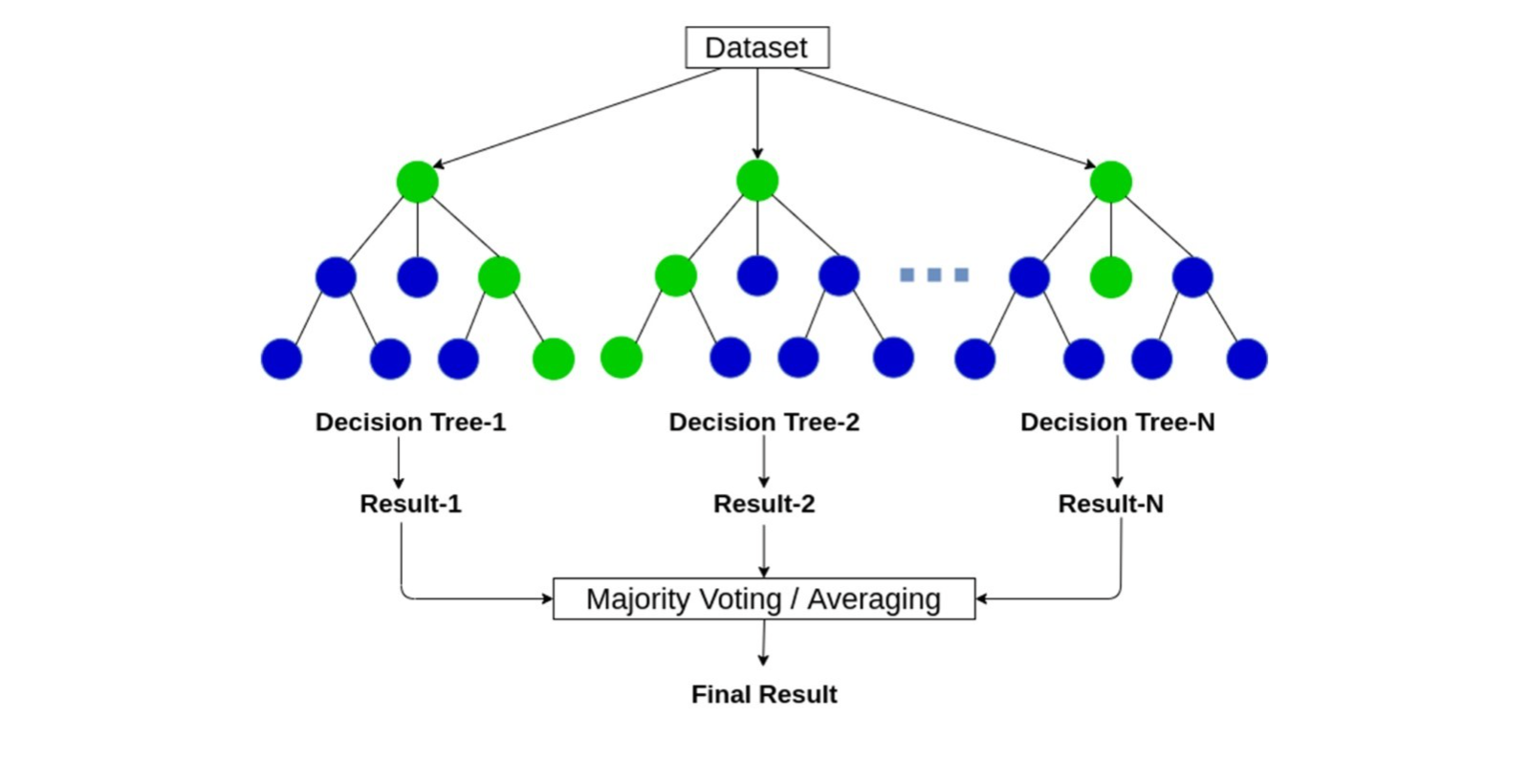
**Introduction**

* 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 be able to use dolphin whistle characteristics, to classify the category of Steno Dolphin whistle.

**Random Forest Methods Explanation**

* Random Forest is a machine learning method for classification and prediction.
* It works by building multiple decision trees, each using different variables.
* The final prediction is based on the majority vote across all the trees.

Images:



**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 like Multinomial GLM.

**Multinomial GLM Pros vs Cons**

Pros:

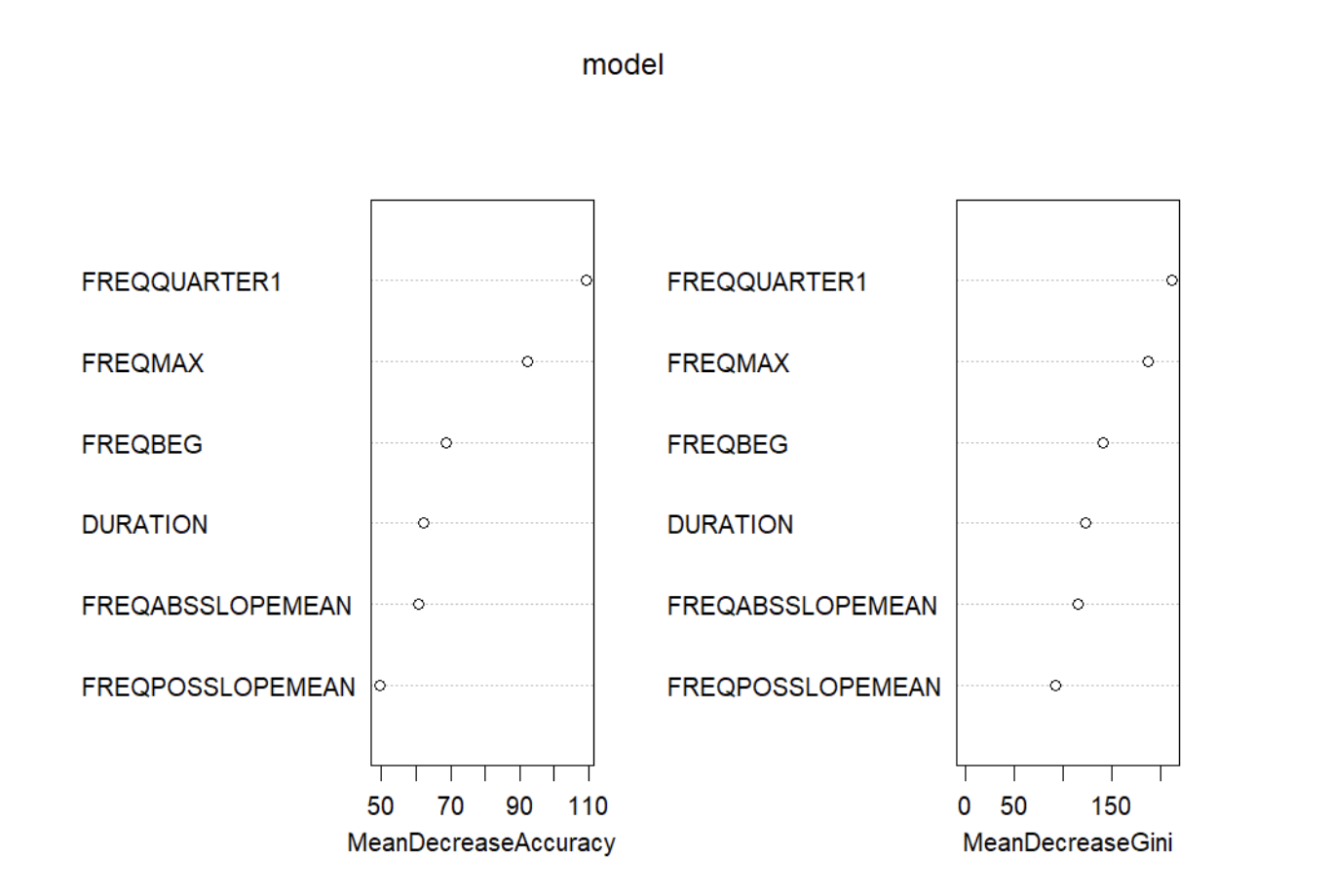
* It shows the relative importance of each variable in prediction.
* It can help understand the relationship between variables in the data.

Cons:

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

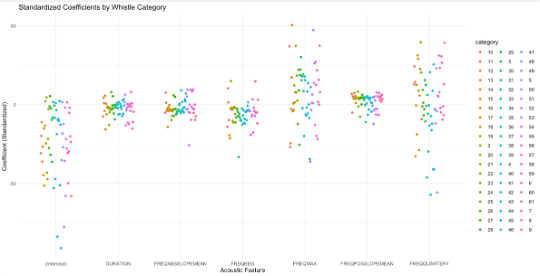
**Random Forest Results**

* Using set seeds ‘123’ and ‘456’ the accuracy of the Random Forest model fitted with the data was 74.32% and 63.39% respectively.
* After averaging the results of 100 trials with random seeds, the accuracy of this model was found to be 67.68%



**Multinomial GLM results**

* The most informative variables to predict categories are Duration, FREQBEG, FREQMAX, FREQQUARTER1, FREQABSSLOPEMEAN, FREQPOSSPLOPEMEAN.
* We have Residual Deviance = 5304.35 and AIC = 6144.35 indicating that acoustic features are informative for classifying whistles.



**Comparison**

**Final Model**

* We decided on Multinomial GLM for various reasons.
* It allows for hypothesis testing and confidence intervals, unlike Random Forest, which helps for more scientific conclusions.
* It is easier to interpret compared to the Random Forest model.
* It is faster to run than Random Forest

References

Introduction

<https://seaworld.org/animals/all-about/bottlenose-dolphin/communication/#:~:text=They%20use%20these%20unique%20whistles,broadcast%20affiliation%20with%20other%20individuals>.

<https://www.fisheries.noaa.gov/species/rough-toothed-dolphin#:~:text=Rough%2Dtoothed%20dolphins%20have%20been%20part%20of%20a%20declared%20unusual,scientists%20study%20unusual%20mortality%20events>.

<https://whalewatchtenerife.org/rough-toothed-dolphins/#:~:text=Rough%20Toothed%20Dolphin%20Info,Killer%20whales%20and%20Humpback%20whales>.

Random Forest Methods Explanation

<https://www.youtube.com/watch?v=v6VJ2RO66Ag>

<https://www.ibm.com/think/topics/random-forest>

Images:

<https://anasbrital98.github.io/blog/2021/Random-Forest/>

<https://dida.do/what-is-random-forest>

Random Forest Pros vs Cons

<https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/>

<https://www.mygreatlearning.com/blog/random-forest-algorithm/>

Multinomial GLM Pros vs Cons

<https://www.wallstreetmojo.com/multinomial-logistic-regression/>

<https://www.mygreatlearning.com/blog/multinomial-logistic-regression/>