

# Rashomon Code Flow

Simon Dovan Nguyen

April 14, 2025

## Contents

### 1 One Iteration Function

1

All code can be found on [Github](#). The links below lead to the function that performs each action.

## 1 One Iteration Function

### 1. Set Up:

- (a) Set seed
- (b) [Load Dataset](#)
- (c) [Train Test Candidate Split Data](#)
  - `df_Test`: 20%
  - `df_Candidate`: 80%
  - Initial `df_Train`: Remaining

### 2. [Learning Procedure](#)

For  $i = 0$  to  $\text{len}(\text{df\_Train})$ :

#### (a) Prediction Method:

- i. [Train TreeFarms](#) model on `df_Train`
  - Results in  $K$  number of trees
- ii. [Predict label](#) for `df_Test` for all  $K$  tree in TreeFarms
- iii. [Calculate F1 score \(micro average\)](#) from each tree in TreeFarms model for the `df_Test`
  - Note that the F1 score is over all  $K$  trees, including the duplicate trees, even if we are only using the set of unique trees for the selection process ([Line 51](#)).
  - Note that my `TestErrorFunction` still stores both F1 scores: (1) using the duplicate trees and (2) using only the unique trees. ([Line 31-55](#)).

#### (b) [Selection](#)

- i. (UNREAL) Restrict ourselves to only the unique trees ([Lines 44 - 47](#)).
- ii. Predict the labels of `df_Candidate` for each (unique) tree in TreeFarms ([Lines 31 - 42](#)).
- iii. Calculate recommendation metric (Vote Entropy) for each observation in the candidate set based off of the (unique/duplicate) trees ([Lines 57 - 75](#)):

$$\text{VoteEntropy}(y, x) = - \sum_{y \in \{0,1\}} \frac{\text{vote}_C(y, x)}{|C|} \log \frac{\text{vote}_C(y, x)}{|C|} \quad (1)$$

where

$$\text{vote}_C = \sum_{c \in C} \mathbb{I}\{c(x) = y\} \quad (2)$$

is the number of "votes" that label  $y$  receives for  $x$  amongst the models in the Rashomon set of trees  $C$ .

- iv. Recommend the candidate observation  $x$  with the highest vote entropy ([Lines 77 - 81](#)):

$$\arg \max_x \text{VoteEntropy}(y, x)$$

#### (c) [Update \(Lines 69-71\)](#)

- i. Add that observation `df_Training`
- ii. Remove that observation from `df_Candidate`

#### (d) Repeat Steps (a) - (c)