**Instructions to run GEM-ITH-Ensemble with Fish\_code.py**

**Tunable parameters**

*GENERAL PARAMETERS*

Line 50: Resampling

Line 56: Kfold splits

Line 63: Training size proportion

*GEM-ITH - BAYESIAN SEARCH*

Line 77: Number of iterations for Bayesian search

Line 88: LASSO parameter search space

Line 107: Random Forest parameter search space

Line 129: XGBoost parameter search space

Line 157: SVM parameter search pace

Line 207: Kfold splits

*GEM and STACKING - GRID SEARCH*

Line 319: LASSO grid search space

Line 332: Random Forest grid search space

Line 344: XGBoost grid search space

Line 357: SVM grid search space

*BASE LEARNERS - GRID SEARCH*

Line 473: LASSO grid search space

Line 482: Random Forest grid search space

Line 491: XGBoost grid search space

Line 501: SVM grid search space

The code is comprised of 5 main sections as follows:

1. Importing libraries, loading the data, and preprocessing
2. GEM-ITH model with hyperparameters tuned by Bayseian search
3. GEM model with hyperparameters tuned by Grid search
4. Stacked generalization models with hyperparameters tuned by Grid search
5. Individual base learners with hyperparameters tuned by Grid search

This code is meant to be run as one entire piece after the user selects their parameters. However, each section is written in a way that after loading the first section, any of the other sections could be run independently if the user desired. The models’ search spaces are defined separately in all sections for two reasons: timing the computation time of all algorithms, and being able to run any of them separately.

After importing necessary modules, lines 36 - 54 load the dataset and create empty dictionaries for storage. Lines 58 - 520 contain the GEM-ITH-ENSEMBLE Algorithm. Lines 67 - 168 tune the models’ hyperparameters with Bayesian search. Lines 170-205 define the optimization model. Lines 209 - 268 initiate the GEM-ITH algorithm to find the optimal weights and hyperparameters. Lines 270 - 312 make the final base model predictions on the test set using the found hyperparameters. They then also aggregate the base models predictions with the found optimal weights. Lines 317 - 427 perform a grid search to tune the base models’ hyperparameters and then run the GEM algorithm. Lines 429 - 467 create the stacked ensemble models. Lines 469 - 507 perform a grid search for the base learners. Lastly, lines 509 - 552 aggregate the results and save the predictions into a csv file.

**Instructions on how to run the code**

1. Load necessary modules
2. Place the data set in the same directory as the main Python file (fish\_code.py)
3. Set the tunable parameters as mentioned in the beginning of this document.
4. Run the entire code to generate the results and output csv files.