hdm-framework

A classification framework to enhance your habitat distribution models

1. Dataset preparation

Pre-processing:
- Spatial filtering: selection of countries
- Temporal filtering: selection of years
- Hierarchical filtering: selection of EUNIS level
- Taxon names harmonizing: use of GBIF

Covariate data:
- External criteria addition: pruning plot locations

2. Parameters evaluation

Model fitting:
- Algorithm choice: selection between MLP, RFC, XGB, TNC, FTT
- Hyperparameters combination: finding optimal values
- Model evaluation: use of k-fold CV

Data representation:
- Encoding choice: CA, PA, RR
- Noise addition: dropout, normal

3. Model training

Data aggregation:
- More data: use of whole set

Metric choice:
- Different metrics: accuracy, F-1 Score
- Different averages: macro, micro
- Different top_k: 1, 3

Features ablation:
- Variables choice: species, location, altitude, country, ecoregion, dune, coast

4. Habitats prediction

Pre-processing:
- Taxon names harmonizing: use of GBIF
- One-hot-encoding and label encoding: use of previously fitted encoders

Dataset:
- Vegetation plots: use of your own (unlabelled) data

Predictions:
- Habitat prediction: use of EUNIS typology