**Description**

This code uses a NNET model to predict the yield of 6-*endo-trig* radical cyclization reactions. Descriptors were extracted with Python scripts (included in the compressed folder “python scripts for descriptor extraction”). The code for the training and testing of other models such as: SIMPLS, KNN, Random Forest, Treebag are included in “other models.R”.

**Instructions for Using “6-endo-radical NNET model.R”**

1. System requirements

This code is supported for macOS and Windows, and has been tested on the following systems:

* Windows 11 Home (21H2)

R version 4.1.2, Platform: x86\_64-w64-mingw32/x64 (64-bit)

* macOS Monterey (Version 12.1)

R version 4.1.0, Platform: aarch64-apple-darwin20 (64-bit)

1. Installation guide

Download and install the following programs:

R (>= 3.6) (Download at <https://cran.r-project.org/mirrors.html> – choose any mirror link)

R Studio (Download at <https://www.rstudio.com/products/rstudio/download/>)

Typical install time on a typical desktop computer: 15 mins.

1. Instructions for use
2. Open “6-endo-radical NNET model.R” in R Studio and modify the working directory to the location of the “Newhouse\_ML\_in\_synthesis” folder that includes the data table.

setwd("XX://Newhouse\_ML\_in\_synthesis")

[Note: Before running the R code on a Mac, change the file locations such that folders are separated by “/” instead of “\\”.]

1. Install relevant packages. This needs to be done only once per computer.

install.packages("caret", “nnet”, "prospectr", "NeuralNetTools", "ggplot2", 'e1071'…)

1. Run the code by selecting each section and then ctrl (“command” in macOS) + enter, which will perform the following steps:

Loading data table "6-endo-radical library.csv"

Splitting the data into training (70%) and test (30%) sets.

Pre-processing of the data.

Training a NNET model using the training set.

Calculating R^2 and RMSE values for the model using the test set.

Plotting a calibration plot of the model using the test set.

Generating a variable importance plot and structure for the model.

Y-randomization test and random data test.

literature validation

Prediction for Clovane substrate refinement

Prediction for experimental validation