

#####

Title: Virtual Experiment using Machine Learning (ML), Principle Component Analysis (PCA), and Bayesian Optimization (BO)

Date of Update: 25/Mar./2021

Version: 1.0

Authors: Kengo Ito<sup>1,2</sup> and Ryo Yamawaki<sup>1</sup>

Affiliation: <sup>1</sup>Graduate School of Medical Life Science, Yokohama City University (YCU), Environmental Factor Analysis Laboratory (EFAL), <sup>2</sup>RIKEN Center for Sustainable Resource Science (CSRS), Environmental Metabolic Analysis Research team (EMAR).

URL: <http://dmar.riken.jp/NMRinformatics/>

License: GPL-3

Reference: Ryo Yamawaki, Akiyo Tei, Kengo Ito, and Jun Kikuchi. "Decomposition Factor Analysis Based on Virtual Experiments Throughout Bayesian Optimization for Compost-Degradable Polymers" *Appl. Sci.* **11**, 2820 (2021).

#####

## Requirements

---

Language

[R](#)

Library

[caret](#)

[rBayesianOptimization](#)

[openxlsx](#)

## How to use

---

Type the code on console (e.g.) :

```
Rscript Virtual_expt.R File_name=test.csv PCA_cols=6-964 Ncomp=2 ML_alg=rf N_Points=10 N_Iter=20  
BO_type=3
```

[Input arguments]

Arg. name	Description
File_name	Input file (*.csv) Column1 = sample name or No. Column2 = objective variable Column3~ = explanatory variable
PCA_cols	Range of column used for PCA
Ncomp	Number of PC used for ML
ML_alg	ML algorithm (method value in <a href="#">caret library</a> )
N_Points	First points used for BO
N_Iter	BO cycle
BO_type	1 = propose only next point 2 = first points are experimental data 3 = first points are predicted data

[Output file]

VE\_results.xlsx : Calculated variables

Sheet name	Description
InData_for_PCA	Input data for PCA
PCA_Score	Scores calculated by PCA
PCA>Loading	Loadings calculated by PCA
InData_for_ML	Input data for ML
ML_Prediction	Result of cross validation
ML_Importance	Importance of variables
BO_Best	Best variables from BO