Package 'VIMPS'

February 21, 2024

Title Calculate Variable Importance with Knock Off Variables

Version 1.0

Description The variable importance is calculated using knock off variables. Then output can be provided in numerical and graphical form. Meredith L Wallace (2023) <doi:10.1186/s12874-023-01965-x>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports caret, ggplot2, ranger, knockoff, ROCR

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

NeedsCompilation no

Author Meredith Wallace [aut, cre] (<https://orcid.org/0000-0003-3951-890X>)

Maintainer Meredith Wallace <lotzmj@upmc.edu>

Repository CRAN

Date/Publication 2024-02-21 20:40:06 UTC

R topics documented:

---

calc_vimps   *calc_vimps*

Description

Calculate the variable importance of the domains for a given dataset

Usage

```
calc_vimps(
  dat,
  dep_var,
  doms,
  calc_ko = TRUE,
  calc_dom = FALSE,
  num_folds = 10,
  num_kos = 100,
  model_all = normal_model,
  model_subset = one_tree_model,
  mtry = NULL,
  min.node.size = NULL,
  iterations = 500,
  ko_path = NULL,
  results_path = NULL,
  output_file_ko = NULL,
  output_file_dom = NULL
)
```

Arguments

dat	A dataframe of data
dep_var	The dependent variable in the dat
doms	A dataframe of the variables in dat and the domain they belong to
calc_ko	True/False to calculate the knock_off importance
calc_dom	True/False to calculate the domain importance
num_folds	The number of folds to use while calculating the classification threshold for predictions
num_kos	The number of sets of knock off variables to create
model_all	The model to use in full ensemble mode in calculations
model_subset	The model to use sigularly for building ensembles from
mtry	The mtry value to use in the random forests
min.node.size	The min.node.size value to use in the random forests
iterations	Number of trees to build while calculating variable importance

ko_path	Where to store the knock off variable sets
results_path	Where to store the intermediary results for calculating variable importance
output_file_ko	Where to store the results of the knock off variable importance
output_file_dom	
	Where to store the results of the domain variable importance

Value

List with 1) Threshold for binary class labeling 2) Model metrics using all variables 3) Model metrics using knock-off variables 4) Variable importance with knock-offs

Examples

```
calc_vimps(
  data.frame(
    X1=c(2,8,3,9,1,4,3,8,0,9,2,8,3,9,1,4,3,8,0,9),
    X2=c(7,2,5,0,9,1,8,8,3,9,7,2,5,0,9,1,8,8,3,9),
    Y=c(0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,1,1,1,1,1)),
"Y",
data.frame(domain=c('X1','X2'),
variable=c('X1','X2')),
num_folds=2,
num_kos=1,
iterations=50)
```

---

graph_results *graph_results*

Description

Graph the variable importance results from calc_vimps

Usage

```
graph_results(results, object)
```

Arguments

results	The results from calc_vimps
object	Which object from results to use for graphing results

Value

No return value

Index