Package: rQSAR (via r-universe)

September 30, 2024

Title QSAR Modeling with Multiple Algorithms: MLR, PLS, and Random Forest

Version 1.0.0

Description Quantitative Structure-Activity Relationship (QSAR) modeling is a valuable tool in computational chemistry and drug design, where it aims to predict the activity or property of chemical compounds based on their molecular structure. In this vignette, we present the 'rQSAR' package, which provides functions for variable selection and QSAR modeling using Multiple Linear Regression (MLR), Partial Least Squares (PLS), and Random Forest algorithms.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Depends R (>= 3.6.0), dplyr, corrplot, tibble, gridExtra

Imports utils, rcdk (>= 3.8.1), ggplot2, caret, pls,randomForest, leaps, stats

VignetteBuilder knitr

NeedsCompilation no

Author Oche Ambrose George [aut, cre] (<https://orcid.org/0000-0002-3979-6232>)

Maintainer Oche Ambrose George <ocheab1@gmail.com>

Repository CRAN

Date/Publication 2024-04-02 13:22:04 UTC

Suggests rmarkdown,knitr

Contents

`build_qsar_models` *Build QSAR models with k-fold cross-validation*

Description

This function builds QSAR (Quantitative Structure-Activity Relationship) models using multiple algorithms such as Multiple Linear Regression (MLR), Partial Least Squares (PLS), and Random Forest with k-fold cross-validation.

Usage

```
build_qsar_models(data_file, k = 5)
```

Arguments

`data_file`	The file path of the dataset.
`k`	The number of folds for cross-validation (default is 5).

Value

A list containing MLR, PLS, and Random Forest models with their predictions, actuals, and formulas.

`correlation_plots` *Create correlation plots for QSAR models*

Description

This function creates correlation plots for QSAR models, showing the relationship between predicted and actual values with a correlation coefficient.

Usage

```
correlation_plots(model_results)
```

Arguments

`model_results`	A list containing QSAR model results.

Value

A list of correlation plots for each QSAR model.

`generate_descriptors_from_sdf`
Generate Molecular Descriptors from SDF File

Description

This function reads an SDF (Structure Data File) containing molecular structures and calculates molecular descriptors for each molecule.

Usage

```
generate_descriptors_from_sdf(sdf_file)
```

Arguments

`sdf_file`	Path to the SDF file.

Value

A matrix containing molecular descriptors for each molecule in the SDF file.

`perform_variable_selection`
Perform variable selection using regression subsets

Description

This function performs variable selection using regression subsets method.

Usage

```
perform_variable_selection(file_path, outcome_col, des_sel_meth = "exhaustive")
```

Arguments

`file_path`	The file path of the dataset.
`outcome_col`	The name of the outcome column.
`des_sel_meth`	The method for variable selection (default is "exhaustive").

Value

A data frame containing the selected variables and the outcome.

`residual_plots` *Function to create residual plots with model type labels*

Description

Function to create residual plots with model type labels

Usage

```
residual_plots(model_results)
```

Arguments

`model_results`	A list containing model results

Value

A list of ggplot objects representing residual plots

Index