

# Package ‘MSML’

November 21, 2023

**Title** Model selection based on Machine Learning (ML)

**Version** 1.0.0.0

**Description** Model selection based on Machine Learning (ML) approach following modified Recursive feature elimination (RFE) process.

**License** ``use_mit_license()``, ``use_gpl3_license()`` or friends to pick a license

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.2

**Depends** R (>= 2.10)

**LazyData** true

## R topics documented:

data_test . . . . .	1
data_train . . . . .	2
data_valid . . . . .	3
model_combination . . . . .	3
model_evaluation . . . . .	4
<b>Index</b>	<b>5</b>

---

data_test	<i>7 sets of PRSs for test dataset and target phenotype</i>
-----------	---

---

## Description

A dataset containing 7 sets of PRSs for test dataset and target phenotype

## Usage

data\_test

**Format**

A data frame for test dataset:

**V1** PRS1, for bin1

**V2** PRS2, for bin1

**V3** PRS3, for bin1

**V4** PRS4, for bin1

**V5** PRS5, for bin1

**V6** PRS6, for bin1

**V7** PRS7, for bin1

**target** Target Phenotype, value

---

data\_train

*7 sets of PRSs for training data set and target phenotype*

---

**Description**

A dataset containing 7 sets of PRSs for training data set and target phenotype

**Usage**

data\_train

**Format**

A data frame for training dataset:

**V1** PRS1, for bin1

**V2** PRS2, for bin1

**V3** PRS3, for bin1

**V4** PRS4, for bin1

**V5** PRS5, for bin1

**V6** PRS6, for bin1

**V7** PRS7, for bin1

**target** Target Phenotype, value

data\_valid

7 sets of PRSs for validation dataset and target phenotype

**Description**

A dataset containing 7 sets of PRSs for validation dataset and target phenotype

**Usage**

```
data_valid
```

**Format**

A data frame for validation dataset:

**V1** PRS1, for bin1

**V2** PRS2, for bin1

**V3** PRS3, for bin1

**V4** PRS4, for bin1

**V5** PRS5, for bin1

**V6** PRS6, for bin1

**V7** PRS7, for bin1

**target** Target Phenotype, value

---

`model_combination` *model\_combination function This function will generate PRS based on all possible combinations of model. The total number of models required to explore the combinations of these 'n' features can be calculated by summing the combinations for each possible number of features, ranging from 1 to 'n' ( $\sum_{i=1}^n C(n,i)$ ). where  $C(n,k)$  represents the binomial coefficient or "n choose k," with n denoting the total number of features and k indicating the number of features to include in each model.*

---

**Description**

`model_combination` function This function will generate PRS based on all possible combinations of model. The total number of models required to explore the combinations of these 'n' features can be calculated by summing the combinations for each possible number of features, ranging from 1 to 'n' ( $\sum_{i=1}^n C(n,i)$ ). where  $C(n,k)$  represents the binomial coefficient or "n choose k," with n denoting the total number of features and k indicating the number of features to include in each model.

**Usage**

```
model_combination(data_train, data_valid, data_test, mv)
```

**Arguments**

<code>data_train</code>	This is the matrix for training dataset
<code>data_valid</code>	This is the matrix for validation dataset
<code>data_test</code>	This is the matrix for test dataset
<code>mv</code>	The total number of columns in <code>data_train/data_valid/data_test</code>

**Value**

This function will generate all possible model outcomes for validation and test dataset

**Examples**

```
data_train <- data_train
data_valid <- data_valid
data_test <- data_test
mv=8
model_combination(data_train,data_valid,data_test,mv)
```

---

`model_evaluation`     *model\_evaluation function*

---

**Description**

This function will identify best model in validation and test dataset.

**Usage**

```
model_evaluation(dat, mv, tn, prev)
```

**Arguments**

<code>dat</code>	This is the matrix for all the combinations of model
<code>mv</code>	The total number of columns in <code>data_train/data_valid/data_test</code>
<code>tn</code>	The total no of best models to be identified
<code>prev</code>	The prevalence of disease in the data

**Value**

This function will generate all possible model outcomes for validation and test dataset

**Examples**

```
dat <- read.table("models_test_all")
mv=8
tn=15
prev=0.047
model_evaluation(dat,mv,tn,prev)
```

# Index

- \*Topic **All**
  - model\_combination, 3
- \*Topic **Identify**
  - model\_evaluation, 4
- \*Topic **best**
  - model\_evaluation, 4
- \*Topic **combinations**
  - model\_combination, 3
- \*Topic **datasets**
  - data\_test, 1
  - data\_train, 2
  - data\_valid, 3
- \*Topic **models**
  - model\_evaluation, 4
- \*Topic **model**
  - model\_combination, 3
- \*Topic **possible**
  - model\_combination, 3

$C(n, i)$ , 3

data\_test, 1  
data\_train, 2  
data\_valid, 3

model\_combination, 3  
model\_evaluation, 4