

# Package ‘daltoolbox’

November 3, 2024

**Title** Leveraging Experiment Lines to Data Analytics

**Version** 1.0.787

## Description

The natural increase in the complexity of current research experiments and data demands better tools to enhance productivity in Data Analytics. The package is a framework designed to address the modern challenges in data analytics workflows. The package is inspired by Experiment Line concepts. It aims to provide seamless support for users in developing their data mining workflows by offering a uniform data model and method API. It enables the integration of various data mining activities, including data preprocessing, classification, regression, clustering, and time series prediction. It also offers options for hyper-parameter tuning and supports integration with existing libraries and languages. Overall, the package provides researchers with a comprehensive set of functionalities for data science, promoting ease of use, extensibility, and integration with various tools and libraries. Information on Experiment Line is based on Ogasawara et al. (2009) <[doi:10.1007/978-3-642-02279-1\\_20](https://doi.org/10.1007/978-3-642-02279-1_20)>.

**License** MIT + file LICENSE

**URL** <https://github.com/cefet-rj-dal/daltoolbox>

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**Imports** FNN, MLmetrics, caret, class, cluster, dbscan, dplyr, e1071, elmNNRcpp, forecast, ggplot2, nnet, randomForest, reshape, tree, reticulate

**Config/reticulate** list( packages = list( list(package = ``scipy"), list(package = ``torch"), list(package = ``pandas"), list(package = ``numpy"), list(package = ``matplotlib"), list(package = ``scikit-learn"), list(package = ``functools"), list(package = ``operator"), list(package = ``sys") ) )

**NeedsCompilation** no

**Author** Eduardo Ogasawara [aut, ths, cre]  
(<<https://orcid.org/0000-0002-0466-0626>>),  
Antonio Castro [aut, ctb],  
Heraldo Borges [aut, ths],  
Diego Carvalho [aut, ths],  
Joel Santos [aut, ths],

Eduardo Bezerra [aut, ths],  
 Rafaelli Coutinho [aut, ths],  
 Federal Center for Technological Education of Rio de Janeiro (CEFET/RJ)  
 [cph]

**Maintainer** Eduardo Ogasawara <eogasawara@ieee.org>

**Repository** CRAN

**Date/Publication** 2024-11-03 05:30:15 UTC

## Contents

action	4
action.dal_transform	5
adjust_class_label	5
adjust_data.frame	6
adjust_factor	6
adjust_matrix	7
adjust_ts_data	7
autoenc_encode	8
autoenc_encode_decode	9
Boston	9
categ_mapping	11
classification	11
cla_dtree	12
cla_knn	13
cla_majority	14
cla_mlp	15
cla_nb	16
cla_rf	16
cla_svm	17
cla_tune	18
cluster	19
clusterer	20
cluster_dbscan	20
cluster_kmeans	21
cluster_pam	22
clu_tune	22
dal_base	23
dal_learner	24
dal_transform	24
dal_tune	25
data_sample	25
do_fit	26
do_predict	27
dt_pca	27
evaluate	28
fit	29
fit.cla_tune	29

fit.cluster_dbscan . . . . .	30
fit_curvature_max . . . . .	30
fit_curvature_min . . . . .	31
inverse_transform . . . . .	32
k_fold . . . . .	32
minmax . . . . .	33
MSE.ts . . . . .	34
outliers . . . . .	34
plot_bar . . . . .	35
plot_boxplot . . . . .	36
plot_boxplot_class . . . . .	36
plot_density . . . . .	37
plot_density_class . . . . .	38
plot_groupedbar . . . . .	39
plot_hist . . . . .	39
plot_lollipop . . . . .	40
plot_pieplot . . . . .	41
plot_points . . . . .	42
plot_radar . . . . .	43
plot_scatter . . . . .	43
plot_series . . . . .	44
plot_stackedbar . . . . .	45
plot_ts . . . . .	45
plot_ts_pred . . . . .	46
predictor . . . . .	47
R2.ts . . . . .	48
regression . . . . .	48
reg_dtree . . . . .	49
reg_knn . . . . .	50
reg_mlp . . . . .	50
reg_rf . . . . .	51
reg_svm . . . . .	52
reg_tune . . . . .	53
sample_random . . . . .	54
sample_stratified . . . . .	55
select_hyper . . . . .	56
select_hyper.cla_tune . . . . .	56
select_hyper.ts_tune . . . . .	57
set_params . . . . .	57
set_params.default . . . . .	58
sin_data . . . . .	58
sMAPE.ts . . . . .	59
smoothing . . . . .	59
smoothing_cluster . . . . .	60
smoothing_freq . . . . .	61
smoothing_inter . . . . .	61
train_test . . . . .	62
train_test_from_folds . . . . .	63

transform . . . . .	64
ts_arima . . . . .	64
ts_conv1d . . . . .	65
ts_data . . . . .	66
ts_elm . . . . .	66
ts_head . . . . .	67
ts_knn . . . . .	68
ts_lstm . . . . .	69
ts_mlp . . . . .	69
ts_norm_an . . . . .	70
ts_norm_diff . . . . .	71
ts_norm_ean . . . . .	72
ts_norm_gminmax . . . . .	73
ts_norm_swminmax . . . . .	73
ts_projection . . . . .	74
ts_reg . . . . .	75
ts_regs_w . . . . .	75
ts_rf . . . . .	76
ts_sample . . . . .	77
ts_svm . . . . .	78
ts_tune . . . . .	79
zscore . . . . .	80
[.ts_data . . . . .	80

<b>Index</b>	<b>82</b>
--------------	-----------

---

action	<i>Action</i>
--------	---------------

---

## Description

Executes the action of model applied in provided data

## Usage

```
action(obj, ...)
```

## Arguments

obj	object: a dal_base object to apply the transformation on the input dataset.
...	optional arguments.

## Value

returns the result of an action of the model applied in provided data

**Examples**

```
data(iris)
# an example is minmax normalization
trans <- minmax()
trans <- fit(trans, iris)
tiris <- action(trans, iris)
```

---

action.dal\_transform    *Action implementation for transform*

---

**Description**

A default function that defines the action to proxy transform method

**Usage**

```
## S3 method for class 'dal_transform'
action(obj, ...)
```

**Arguments**

obj	object
...	optional arguments

**Value**

returns a transformed data

**Examples**

```
#See ?minmax for an example of transformation
```

---

adjust\_class\_label    *Adjust categorical mapping*

---

**Description**

Converts a vector into a categorical mapping, where each category is represented by a specific value. By default, the values represent binary categories (true/false)

**Usage**

```
adjust_class_label(x, valTrue = 1, valFalse = 0)
```

**Arguments**

x	vector to be categorized
valTrue	value to represent true
valFalse	value to represent false

**Value**

returns an adjusted categorical mapping

---

adjust_data.frame	<i>Adjust to data frame</i>
-------------------	-----------------------------

---

**Description**

Converts a dataset to a data.frame if it is not already in that format

**Usage**

```
adjust_data.frame(data)
```

**Arguments**

data	dataset
------	---------

**Value**

returns a data.frame

**Examples**

```
data(iris)
df <- adjust_data.frame(iris)
```

---

adjust_factor	<i>Adjust factors</i>
---------------	-----------------------

---

**Description**

Converts a vector into a factor with specified levels and labels

**Usage**

```
adjust_factor(value, ilevels, slevels)
```

**Arguments**

value            vector to be converted into factor  
ilevels         order for categorical values  
slevels         labels for categorical values

**Value**

returns an adjusted factor

---

*adjust\_matrix*            *Adjust to matrix*

---

**Description**

Converts a dataset to a matrix format if it is not already in that format

**Usage**

```
adjust_matrix(data)
```

**Arguments**

data            dataset

**Value**

returns an adjusted matrix

**Examples**

```
data(iris)  
mat <- adjust_matrix(iris)
```

---

*adjust\_ts\_data*            *Adjust ts\_data*

---

**Description**

Converts a dataset to a ts\_data object

**Usage**

```
adjust_ts_data(data)
```

**Arguments**

data            dataset

**Value**

returns an adjusted ts\_data

---

autoenc\_encode            *Autoencoder - Encode*

---

**Description**

Creates an deep learning autoencoder to encode a sequence of observations. It wraps the pytorch and reticulate libraries.

**Usage**

```
autoenc_encode(
  input_size,
  encoding_size,
  batch_size = 32,
  num_epochs = 1000,
  learning_rate = 0.001
)
```

**Arguments**

input\_size        input size  
 encoding\_size    encoding size  
 batch\_size       size for batch learning  
 num\_epochs       number of epochs for training  
 learning\_rate    learning rate

**Value**

returns a autoenc\_encode object.

**Examples**

```
#See an example of using `autoenc_encode` at this
#[link](https://github.com/cefet-rj-dal/daltoolbox/blob/main/transf/van_encode.ipynb)
```



---

autoenc\_encode\_decode *Autoencoder - Encode-decode*

---

### Description

Creates an deep learning autoencoder to encode-decode a sequence of observations. It wraps the pytorch and reticulate libraries.

### Usage

```
autoenc_encode_decode(  
  input_size,  
  encoding_size,  
  batch_size = 32,  
  num_epochs = 1000,  
  learning_rate = 0.001  
)
```

### Arguments

input_size	input size
encoding_size	encoding size
batch_size	size for batch learning
num_epochs	number of epochs for training
learning_rate	learning rate

### Value

returns a autoenc\_encode\_decode object.

### Examples

```
#See an example of using `autoenc_encode_decode` at this  
#[link](https://github.com/cefet-rj-dal/daltoolbox/blob/main/transf/van_enc_decode.ipynb)
```

---

Boston

*Boston Housing Data (Regression)*

---

**Description**

housing values in suburbs of Boston.

- crim: per capita crime rate by town.
- zn: proportion of residential land zoned for lots over 25,000 sq.ft.
- indus: proportion of non-retail business acres per town
- chas: Charles River dummy variable (= 1 if tract bounds)
- nox: nitric oxides concentration (parts per 10 million)
- rm: average number of rooms per dwelling
- age: proportion of owner-occupied units built prior to 1940
- dis: weighted distances to five Boston employment centres
- rad: index of accessibility to radial highways
- tax: full-value property-tax rate per \$10,000
- ptratio: pupil-teacher ratio by town
- black:  $1000(Bk - 0.63)^2$  where Bk is the proportion of blacks by town
- lstat: percentage of lower status of the population
- medv: Median value of owner-occupied homes in \$1000's

**Usage**

```
data(Boston)
```

**Format**

Regression Dataset.

**Source**

This dataset was obtained from the MASS library.

**References**

Creator: Harrison, D. and Rubinfeld, D.L. Hedonic prices and the demand for clean air, J. Environ. Economics & Management, vol.5, 81-102, 1978.

**Examples**

```
data(Boston)  
head(Boston)
```

---

categ_mapping	<i>Categorical mapping</i>
---------------	----------------------------

---

**Description**

Categorical mapping provides a way to map the levels of a categorical variable to new values. Each possible value is converted to a binary attribute.

**Usage**

```
categ_mapping(attribute)
```

**Arguments**

attribute      attribute to be categorized.

**Value**

returns a data frame with binary attributes, one for each possible category.

**Examples**

```
cm <- categ_mapping("Species")
iris_cm <- transform(cm, iris)

# can be made in a single column
species <- iris[, "Species", drop=FALSE]
iris_cm <- transform(cm, species)
```

---

classification	<i>classification</i>
----------------	-----------------------

---

**Description**

Ancestor class for classification problems using MLmetrics nnet

**Usage**

```
classification(attribute, slevels)
```

**Arguments**

attribute      attribute target to model building  
slevels        possible values for the target classification

**Value**

returns a classification object

**Examples**

```
#See ?cla_dtree for a classification example using a decision tree
```

---

cla_dtree	<i>Decision Tree for classification</i>
-----------	---

---

**Description**

Creates a classification object that uses the Decision Tree algorithm for classification. It wraps the tree library.

**Usage**

```
cla_dtree(attribute, slevels)
```

**Arguments**

attribute	attribute target to model building
slevels	the possible values for the target classification

**Value**

returns a classification object

**Examples**

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_dtree("Species", slevels)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla_knn	<i>K Nearest Neighbor Classification</i>
---------	--

---

**Description**

Classifies using the K-Nearest Neighbor algorithm. It wraps the class library.

**Usage**

```
cla_knn(attribute, slevels, k = 1)
```

**Arguments**

attribute	attribute target to model building.
slevels	possible values for the target classification.
k	a vector of integers indicating the number of neighbors to be considered.

**Value**

returns a knn object.

**Examples**

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_knn("Species", slevels, k=3)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

`cla_majority`*Majority Classification*

---

### Description

This function creates a classification object that uses the majority vote strategy to predict the target attribute. Given a target attribute, the function counts the number of occurrences of each value in the dataset and selects the one that appears most often.

### Usage

```
cla_majority(attribute, slevels)
```

### Arguments

<code>attribute</code>	attribute target to model building.
<code>slevels</code>	possible values for the target classification.

### Value

returns a classification object.

### Examples

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_majority("Species", slevels)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla_mlp	<i>MLP for classification</i>
---------	-------------------------------

---

### Description

Creates a classification object that uses the Multi-Layer Perceptron (MLP) method. It wraps the nnet library.

### Usage

```
cla_mlp(attribute, slevels, size = NULL, decay = 0.1, maxit = 1000)
```

### Arguments

attribute	attribute target to model building
slevels	possible values for the target classification
size	number of nodes that will be used in the hidden layer
decay	how quickly it decreases in gradient descent
maxit	maximum iterations

### Value

returns a classification object

### Examples

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_mlp("Species", slevels, size=3, decay=0.03)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla_nb	<i>Naive Bayes Classifier</i>
--------	-------------------------------

---

**Description**

Classification using the Naive Bayes algorithm It wraps the e1071 library.

**Usage**

```
cla_nb(attribute, slevels)
```

**Arguments**

attribute	attribute target to model building.
slevels	possible values for the target classification.

**Value**

returns a classification object.

**Examples**

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_nb("Species", slevels)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla_rf	<i>Random Forest for classification</i>
--------	---

---

**Description**

Creates a classification object that uses the Random Forest method It wraps the randomForest library.



**Usage**

```
cla_rf(attribute, slevels, nodesize = 5, ntree = 10, mtry = NULL)
```

**Arguments**

attribute	attribute target to model building
slevels	possible values for the target classification
nodesize	node size
ntree	number of trees
mtry	number of attributes to build tree

**Value**

returns a classification object

**Examples**

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_rf("Species", slevels, ntree=5)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla\_svm

*SVM for classification*

---

**Description**

Creates a classification object that uses the Support Vector Machine (SVM) method for classification. It wraps the e1071 and svm library.

**Usage**

```
cla_svm(attribute, slevels, epsilon = 0.1, cost = 10, kernel = "radial")
```

**Arguments**

attribute	attribute target to model building
slevels	possible values for the target classification
epsilon	parameter that controls the width of the margin around the separating hyperplane
cost	parameter that controls the trade-off between having a wide margin and correctly classifying training data points
kernel	the type of kernel function to be used in the SVM algorithm (linear, radial, polynomial, sigmoid)

**Value**

returns a SVM classification object

**Examples**

```
data(iris)
slevels <- levels(iris$Species)
model <- cla_svm("Species", slevels, epsilon=0.0, cost=20.000)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

model <- fit(model, train)

prediction <- predict(model, test)
predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics
```

---

cla\_tune

*Classification Tune*

---

**Description**

This function performs a grid search or random search over specified hyperparameter values to optimize a base classification model

**Usage**

```
cla_tune(base_model, folds = 10, metric = "accuracy")
```

**Arguments**

base_model	base model for tuning
folds	number of folds for cross-validation
metric	metric used to optimize

**Value**

returns a `cla_tune` object

**Examples**

```
# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, iris)
train <- sr$train
test <- sr$test

# hyper parameter setup
tune <- cla_tune(cla_mlp("Species", levels(iris$Species)))
ranges <- list(size=c(3:5), decay=c(0.1))

# hyper parameter optimization
model <- fit(tune, train, ranges)

# testing optimization
test_prediction <- predict(model, test)
test_predictand <- adjust_class_label(test[, "Species"])
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

cluster

*Cluster*

---

**Description**

Defines a cluster method

**Usage**

```
cluster(obj, ...)
```

**Arguments**

obj	a clusterer object
...	optional arguments

**Value**

clustered data

**Examples**

```
#See ?cluster_kmeans for an example of transformation
```

---

clusterer	<i>Clusterer</i>
-----------	------------------

---

**Description**

Ancestor class for clustering problems

**Usage**

```
clusterer()
```

**Value**

returns a clusterer object

**Examples**

```
#See ?cluster_kmeans for an example of transformation
```

---

cluster_dbscan	<i>DBSCAN</i>
----------------	---------------

---

**Description**

Creates a clusterer object that uses the DBSCAN method It wraps the dbscan library.

**Usage**

```
cluster_dbscan(minPts = 3, eps = NULL)
```

**Arguments**

minPts	minimum number of points
eps	distance value

**Value**

returns a dbscan object

**Examples**

```
# setup clustering
model <- cluster_dbscan(minPts = 3)

#load dataset
data(iris)

# build model
model <- fit(model, iris[,1:4])
clu <- cluster(model, iris[,1:4])
table(clu)

# evaluate model using external metric
eval <- evaluate(model, clu, iris$Species)
eval
```

---

cluster_kmeans	<i>k-means</i>
----------------	----------------

---

**Description**

Creates a clusterer object that uses the k-means method It wraps the stats library.

**Usage**

```
cluster_kmeans(k = 1)
```

**Arguments**

k                    the number of clusters to form.

**Value**

returns a k-means object.

**Examples**

```
# setup clustering
model <- cluster_kmeans(k=3)

#load dataset
data(iris)

# build model
model <- fit(model, iris[,1:4])
clu <- cluster(model, iris[,1:4])
table(clu)

# evaluate model using external metric
eval <- evaluate(model, clu, iris$Species)
eval
```

---

cluster_pam	<i>PAM</i>
-------------	------------

---

**Description**

Creates a clusterer object that uses the Partition Around Medoids (PAM) method It wraps the cluster library.

**Usage**

```
cluster_pam(k = 1)
```

**Arguments**

k                    the number of clusters to generate.

**Value**

returns PAM object.

**Examples**

```
# setup clustering
model <- cluster_pam(k = 3)

#load dataset
data(iris)

# build model
model <- fit(model, iris[,1:4])
clu <- cluster(model, iris[,1:4])
table(clu)

# evaluate model using external metric
eval <- evaluate(model, clu, iris$Species)
eval
```

---

clu_tune	<i>Clustering Tune</i>
----------	------------------------

---

**Description**

Creates an object for tuning clustering models. This object can be used to fit and optimize clustering algorithms by specifying hyperparameter ranges

**Usage**

```
clu_tune(base_model)
```

**Arguments**

base\_model      base model for tuning

**Value**

returns a clu\_tune object.

**Examples**

```
data(iris)

# fit model
model <- clu_tune(cluster_kmeans(k = 0))
ranges <- list(k = 1:10)
model <- fit(model, iris[,1:4], ranges)
model$k
```

---

dal\_base

*Class dal\_base*

---

**Description**

The dal\_base class is an abstract class for all dal descendants classes. It provides both fit() and action() functions

**Usage**

```
dal_base()
```

**Value**

returns a dal\_base object

**Examples**

```
trans <- dal_base()
```

---

`dal_learner`*DAL Learner*

---

**Description**

A ancestor class for clustering, classification, regression, and time series regression. It also provides the basis for specialized evaluation of learning performance.

An example of a learner is a decision tree (`cla_dtree`)

**Usage**

```
dal_learner()
```

**Value**

returns a learner

**Examples**

```
#See ?cla_dtree for a classification example using a decision tree
```

---

`dal_transform`*DAL Transform*

---

**Description**

A transformation method applied to a dataset. If needed, the fit can be called to adjust the transform.

**Usage**

```
dal_transform()
```

**Value**

returns a `dal_transform` object.

**Examples**

```
#See ?minmax for an example of transformation
```



---

dal_tune	<i>DAL Tune</i>
----------	-----------------

---

**Description**

Creates an ancestor class for hyperparameter optimization, allowing the tuning of a base model using cross-validation.

**Usage**

```
dal_tune(base_model, folds = 10)
```

**Arguments**

base_model	base model for tuning
folds	number of folds for cross-validation

**Value**

returns a dal\_tune object

**Examples**

```
#See ?cla_tune for classification tuning  
#See ?reg_tune for regression tuning  
#See ?ts_tune for time series tuning
```

---

data_sample	<i>Data Sample</i>
-------------	--------------------

---

**Description**

The data\_sample function in R is used to randomly sample data from a given data frame. It can be used to obtain a subset of data for further analysis or modeling.

Two basic specializations of data\_sample are sample\_random and sample\_stratified. They provide random sampling and stratified sampling, respectively.

Data sample provides both training and testing partitioning (train\_test) and k-fold partitioning (k\_fold) of data.

**Usage**

```
data_sample()
```

**Value**

returns an object of class data\_sample

**Examples**

```
#using random sampling
sample <- sample_random()
tt <- train_test(sample, iris)

# distribution of train
table(tt$train$Species)

# preparing dataset into four folds
folds <- k_fold(sample, iris, 4)

# distribution of folds
tbl <- NULL
for (f in folds) {
  tbl <- rbind(tbl, table(f$Species))
}
head(tbl)
```

---

do\_fit

*Fit Time Series Model*

---

**Description**

The actual time series model fitting. This method should be override by descendants.

**Usage**

```
do_fit(obj, x, y = NULL)
```

**Arguments**

obj	an object representing the model or algorithm to be fitted
x	a matrix or data.frame containing the input features for training the model
y	a vector or matrix containing the output values to be predicted by the model

**Value**

returns a fitted object

---

do_predict	<i>Predict Time Series Model</i>
------------	----------------------------------

---

**Description**

The actual time series model prediction. This method should be override by descendants.

**Usage**

```
do_predict(obj, x)
```

**Arguments**

obj	an object representing the fitted model or algorithm
x	a matrix or data.frame containing the input features for making predictions

**Value**

returns the predicted values

---

dt_pca	<i>PCA</i>
--------	------------

---

**Description**

PCA (Principal Component Analysis) is an unsupervised dimensionality reduction technique used in data analysis and machine learning. It transforms a dataset of possibly correlated variables into a new set of uncorrelated variables called principal components.

**Usage**

```
dt_pca(attribute = NULL, components = NULL)
```

**Arguments**

attribute	target attribute to model building
components	number of components for PCA

**Value**

returns an object of class dt\_pca

**Examples**

```

mypca <- dt_pca("Species")
# Automatically fitting number of components
mypca <- fit(mypca, iris)
iris.pca <- transform(mypca, iris)
head(iris.pca)
head(mypca$pca.transf)
# Manual establishment of number of components
mypca <- dt_pca("Species", 3)
mypca <- fit(mypca, datasets::iris)
iris.pca <- transform(mypca, iris)
head(iris.pca)
head(mypca$pca.transf)

```

---

evaluate

*Evaluate*


---

**Description**

Evaluate learner performance. The actual evaluate varies according to the type of learner (clustering, classification, regression, time series regression)

**Usage**

```
evaluate(obj, ...)
```

**Arguments**

obj	object
...	optional arguments

**Value**

returns the evaluation

**Examples**

```

data(iris)
slevels <- levels(iris$Species)
model <- cla_dtree("Species", slevels)
model <- fit(model, iris)
prediction <- predict(model, iris)
predictand <- adjust_class_label(iris[, "Species"])
test_eval <- evaluate(model, predictand, prediction)
test_eval$metrics

```

---

fit	<i>Fit</i>
-----	------------

---

**Description**

Applies the `fit` method to a model object to train or configure it using the provided data and optional arguments

**Usage**

```
fit(obj, ...)
```

**Arguments**

<code>obj</code>	object
<code>...</code>	optional arguments.

**Value**

returns a object after fitting

**Examples**

```
data(iris)
# an example is minmax normalization
trans <- minmax()
trans <- fit(trans, iris)
tiris <- action(trans, iris)
```

---

<code>fit.cla_tune</code>	<i>tune hyperparameters of ml model</i>
---------------------------	---

---

**Description**

Tunes the hyperparameters of a machine learning model for classification

**Usage**

```
## S3 method for class 'cla_tune'
fit(obj, data, ranges, ...)
```

**Arguments**

<code>obj</code>	an object containing the model and tuning configuration
<code>data</code>	the dataset used for training and evaluation
<code>ranges</code>	a list of hyperparameter ranges to explore
<code>...</code>	optional arguments

**Value**

a fitted obj

---

fit.cluster\_dbscan     *fit dbscan model*

---

**Description**

Fits a DBSCAN clustering model by setting the eps parameter. If eps is not provided, it is estimated based on the k-nearest neighbor distances. It wraps dbscan library

**Usage**

```
## S3 method for class 'cluster_dbscan'
fit(obj, data, ...)
```

**Arguments**

obj	an object containing the DBSCAN model configuration, including minPts and optionally eps
data	the dataset to use for fitting the model
...	optional arguments

**Value**

returns a fitted obj with the eps parameter set

---

fit\_curvature\_max     *maximum curvature analysis*

---

**Description**

Fitting a curvature model in a sequence of observations. It extracts the the maximum curvature computed.

**Usage**

```
fit_curvature_max()
```

**Value**

returns an object of class fit\_curvature\_max, which inherits from the fit\_curvature and dal\_transform classes. The object contains a list with the following elements:

- x: The position in which the maximum curvature is reached.
- y: The value where the the maximum curvature occurs.
- yfit: The value of the maximum curvature.

**Examples**

```
x <- seq(from=1,to=10,by=0.5)
dat <- data.frame(x = x, value = -log(x), variable = "log")
myfit <- fit_curvature_max()
res <- transform(myfit, dat$value)
head(res)
```

---

fit_curvature_min	<i>minimum curvature analysis</i>
-------------------	-----------------------------------

---

**Description**

Fitting a curvature model in a sequence of observations. It extracts the the minimum curvature computed.

**Usage**

```
fit_curvature_min()
```

**Value**

Returns an object of class `fit_curvature_max`, which inherits from the `fit_curvature` and `dal_transform` classes. The object contains a list with the following elements:

- `x`: The position in which the minimum curvature is reached.
- `y`: The value where the the minimum curvature occurs.
- `yfit`: The value of the minimum curvature.

**Examples**

```
x <- seq(from=1,to=10,by=0.5)
dat <- data.frame(x = x, value = log(x), variable = "log")
myfit <- fit_curvature_min()
res <- transform(myfit, dat$value)
head(res)
```

---

inverse\_transform      *Inverse Transform*

---

**Description**

Reverses the transformation applied to data.

**Usage**

```
inverse_transform(obj, ...)
```

**Arguments**

obj                    a dal\_transform object.  
...                    optional arguments.

**Value**

dataset inverse transformed.

**Examples**

```
#See ?minmax for an example of transformation
```

---

k\_fold                    *K-fold sampling*

---

**Description**

k-fold partition of a dataset using a sampling method

**Usage**

```
k_fold(obj, data, k)
```

**Arguments**

obj                    an object representing the sampling method  
data                   dataset to be partitioned  
k                      number of folds

**Value**

returns a list of k data frames



**Examples**

```
#using random sampling
sample <- sample_random()

# preparing dataset into four folds
folds <- k_fold(sample, iris, 4)

# distribution of folds
tbl <- NULL
for (f in folds) {
  tbl <- rbind(tbl, table(f$Species))
}
head(tbl)
```

---

minmax

*Min-max normalization*

---

**Description**

The minmax performs scales data between [0,1].

$$\text{minmax} = (x - \min(x)) / (\max(x) - \min(x))$$

**Usage**

```
minmax()
```

**Value**

returns an object of class minmax

**Examples**

```
data(iris)
head(iris)

trans <- minmax()
trans <- fit(trans, iris)
tiris <- transform(trans, iris)
head(tiris)

itiris <- inverse_transform(trans, tiris)
head(itiris)
```

---

MSE.ts

*MSE*


---

**Description**

Compute the mean squared error (MSE) between actual values and forecasts of a time series

**Usage**

```
MSE.ts(actual, prediction)
```

**Arguments**

actual	real observations
prediction	predicted observations

**Value**

returns a number, which is the calculated MSE

---

outliers

*Outliers*


---

**Description**

The outliers class uses box-plot definition for outliers. An outlier is a value that is below than  $Q_1 - 1.5 \cdot IQR$  or higher than  $Q_3 + 1.5 \cdot IQR$ . The class remove outliers for numeric attributes. Users can set alpha to 3 to remove extreme values.

**Usage**

```
outliers(alpha = 1.5)
```

**Arguments**

alpha	boxplot outlier threshold (default 1.5, but can be 3.0 to remove extreme values)
-------	--

**Value**

returns an outlier object

**Examples**

```
# code for outlier removal
out_obj <- outliers() # class for outlier analysis
out_obj <- fit(out_obj, iris) # computing boundaries
iris.clean <- transform(out_obj, iris) # returning cleaned dataset

#inspection of cleaned dataset
nrow(iris.clean)

idx <- attr(iris.clean, "idx")
table(idx)
iris.outliers <- iris[idx,]
iris.outliers
```

---

plot\_bar

*Plot bar graph*


---

**Description**

this function displays a bar graph from a data frame containing x-axis categories using ggplot2.

**Usage**

```
plot_bar(data, label_x = "", label_y = "", colors = NULL, alpha = 1)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
alpha	level of transparency

**Value**

returns a ggplot graphic

**Examples**

```
#summarizing iris dataset
data <- iris |> dplyr::group_by(Species) |>
  dplyr::summarize(Sepal.Length=mean(Sepal.Length))
head(data)

#ploting data
grf <- plot_bar(data, colors="blue")
plot(grf)
```

---

plot_boxplot	<i>Plot boxplot</i>
--------------	---------------------

---

**Description**

this function displays a boxplot graph from a data frame containing x-axis categories and numeric values using ggplot2.

**Usage**

```
plot_boxplot(data, label_x = "", label_y = "", colors = NULL, barwidth = 0.25)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
barwidth	width of bar

**Value**

returns a ggplot graphic

**Examples**

```
grf <- plot_boxplot(iris, colors="white")
plot(grf)
```

---

plot_boxplot_class	<i>Boxplot per class</i>
--------------------	--------------------------

---

**Description**

This function generates boxplots grouped by a specified class label from a data frame containing numeric values using ggplot2.

**Usage**

```
plot_boxplot_class(
  data,
  class_label,
  label_x = "",
  label_y = "",
  colors = NULL
)
```

**Arguments**

data	data.frame contain x, value, and variable
class_label	name of attribute for class label
label_x	x-axis label
label_y	y-axis label
colors	color vector

**Value**

returns a ggplot graphic

**Examples**

```
grf <- plot_boxplot_class(iris |> dplyr::select(Sepal.Width, Species),
  class = "Species", colors=c("red", "green", "blue"))
plot(grf)
```

---

plot_density	<i>Plot density</i>
--------------	---------------------

---

**Description**

This function generates a density plot from a data frame containing numeric values using ggplot2. If the data frame has multiple columns, densities can be grouped and plotted.

**Usage**

```
plot_density(
  data,
  label_x = "",
  label_y = "",
  colors = NULL,
  bin = NULL,
  alpha = 0.25
)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
bin	bin width for density estimation
alpha	level of transparency

**Value**

returns a ggplot graphic

**Examples**

```
grf <- plot_density(iris |> dplyr::select(Sepal.Width), colors="blue")
plot(grf)
```

---

plot\_density\_class      *Plot density per class*

---

**Description**

This function generates density plots using ggplot2 grouped by a specified class label from a data frame containing numeric values.

**Usage**

```
plot_density_class(
  data,
  class_label,
  label_x = "",
  label_y = "",
  colors = NULL,
  bin = NULL,
  alpha = 0.5
)
```

**Arguments**

data	data.frame contain x, value, and variable
class_label	name of attribute for class label
label_x	x-axis label
label_y	y-axis label
colors	color vector
bin	bin width for density estimation
alpha	level of transparency

**Value**

returns a ggplot graphic

**Examples**

```
grf <- plot_density_class(iris |> dplyr::select(Sepal.Width, Species),
  class = "Species", colors=c("red", "green", "blue"))
plot(grf)
```

---

plot_groupedbar	<i>Plot grouped bar</i>
-----------------	-------------------------

---

**Description**

This function generates a grouped bar plot from a given data frame using ggplot2.

**Usage**

```
plot_groupedbar(data, label_x = "", label_y = "", colors = NULL, alpha = 1)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
alpha	level of transparency

**Value**

returns a ggplot graphic

**Examples**

```
#summarizing iris dataset
data <- iris |> dplyr::group_by(Species) |>
  dplyr::summarize(Sepal.Length=mean(Sepal.Length), Sepal.Width=mean(Sepal.Width))
head(data)

#ploting data
grf <- plot_groupedbar(data, colors=c("blue", "red"))
plot(grf)
```

---

plot_hist	<i>Plot histogram</i>
-----------	-----------------------

---

**Description**

This function generates a histogram from a specified data frame using ggplot2.

**Usage**

```
plot_hist(data, label_x = "", label_y = "", color = "white", alpha = 0.25)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
color	color vector
alpha	transparency level

**Value**

returns a ggplot graphic

**Examples**

```
grf <- plot_hist(iris |> dplyr::select(Sepal.Width), color=c("blue"))
plot(grf)
```

---

plot_lollipop	<i>Plot lollipop</i>
---------------	----------------------

---

**Description**

This function creates a lollipop chart using ggplot2.

**Usage**

```
plot_lollipop(  
  data,  
  label_x = "",  
  label_y = "",  
  colors = NULL,  
  color_text = "black",  
  size_text = 3,  
  size_ball = 8,  
  alpha_ball = 0.2,  
  min_value = 0,  
  max_value_gap = 1  
)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
color_text	color of text inside ball



size_text	size of text inside ball
size_ball	size of ball
alpha_ball	transparency of ball
min_value	minimum value
max_value_gap	maximum value gap

**Value**

returns a ggplot graphic

**Examples**

```
#summarizing iris dataset
data <- iris |> dplyr::group_by(Species) |>
  dplyr::summarize(Sepal.Length=mean(Sepal.Length))
head(data)

#plotting data
grf <- plot_lollipop(data, colors="blue", max_value_gap=0.2)
plot(grf)
```

---

plot_pieplot	<i>Plot pie</i>
--------------	-----------------

---

**Description**

This function creates a pie chart using ggplot2.

**Usage**

```
plot_pieplot(
  data,
  label_x = "",
  label_y = "",
  colors = NULL,
  textcolor = "white",
  bordercolor = "black"
)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
textcolor	text color
bordercolor	border color

**Value**

returns a ggplot graphic

**Examples**

```
#summarizing iris dataset
data <- iris |> dplyr::group_by(Species) |>
  dplyr::summarize(Sepal.Length=mean(Sepal.Length))
head(data)

#plotting data
grf <- plot_pieplot(data, colors=c("red", "green", "blue"))
plot(grf)
```

---

plot\_points

*Plot points*

---

**Description**

This function creates a scatter plot using ggplot2.

**Usage**

```
plot_points(data, label_x = "", label_y = "", colors = NULL)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector

**Value**

returns a ggplot graphic

**Examples**

```
x <- seq(0, 10, 0.25)
data <- data.frame(x, sin=sin(x), cosine=cos(x)+5)
head(data)

grf <- plot_points(data, colors=c("red", "green"))
plot(grf)
```

---

plot_radar	<i>Plot radar</i>
------------	-------------------

---

**Description**

This function creates a radar chart using ggplot2.

**Usage**

```
plot_radar(data, label_x = "", label_y = "", colors = NULL)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector

**Value**

returns a ggplot graphic

**Examples**

```
data <- data.frame(name = "Petal.Length", value = mean(iris$Petal.Length))
data <- rbind(data, data.frame(name = "Petal.Width", value = mean(iris$Petal.Width)))
data <- rbind(data, data.frame(name = "Sepal.Length", value = mean(iris$Sepal.Length)))
data <- rbind(data, data.frame(name = "Sepal.Width", value = mean(iris$Sepal.Width)))

grf <- plot_radar(data, colors="red") + ggplot2::ylim(0, NA)
plot(grf)
```

---

plot_scatter	<i>Scatter graph</i>
--------------	----------------------

---

**Description**

This function creates a scatter plot using ggplot2.

**Usage**

```
plot_scatter(data, label_x = "", label_y = "", colors = NULL)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector

**Value**

return a ggplot graphic

**Examples**

```
grf <- plot_scatter(iris |> dplyr::select(x = Sepal.Length,
  value = Sepal.Width, variable = Species),
  label_x = "Sepal.Length", label_y = "Sepal.Width",
  colors=c("red", "green", "blue"))
plot(grf)
```

---

plot_series	<i>Plot series</i>
-------------	--------------------

---

**Description**

This function creates a time series plot using ggplot2.

**Usage**

```
plot_series(data, label_x = "", label_y = "", colors = NULL)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector

**Value**

returns a ggplot graphic

**Examples**

```
x <- seq(0, 10, 0.25)
data <- data.frame(x, sin=sin(x))
head(data)

grf <- plot_series(data, colors=c("red"))
plot(grf)
```

---

plot_stackedbar	<i>Plot stacked bar</i>
-----------------	-------------------------

---

**Description**

this function creates a stacked bar chart using ggplot2.

**Usage**

```
plot_stackedbar(data, label_x = "", label_y = "", colors = NULL, alpha = 1)
```

**Arguments**

data	data.frame contain x, value, and variable
label_x	x-axis label
label_y	y-axis label
colors	color vector
alpha	level of transparency

**Value**

returns a ggplot graphic

**Examples**

```
#summarizing iris dataset
data <- iris |> dplyr::group_by(Species) |>
  dplyr::summarize(Sepal.Length=mean(Sepal.Length), Sepal.Width=mean(Sepal.Width))

#plotting data
grf <- plot_stackedbar(data, colors=c("blue", "red"))
plot(grf)
```

---

plot_ts	<i>Plot time series chart</i>
---------	-------------------------------

---

**Description**

This function plots a time series chart with points and a line using ggplot2.

**Usage**

```
plot_ts(x = NULL, y, label_x = "", label_y = "", color = "black")
```

**Arguments**

x	input variable
y	output variable
label_x	x-axis label
label_y	y-axis label
color	color for time series

**Value**

returns a ggplot graphic

**Examples**

```
x <- seq(0, 10, 0.25)
data <- data.frame(x, sin=sin(x))
head(data)

grf <- plot_ts(x = data$x, y = data$sin, color=c("red"))
plot(grf)
```

---

plot\_ts\_pred

*Plot a time series chart with predictions*

---

**Description**

This function plots a time series chart with three lines: the original series, the adjusted series, and the predicted series using ggplot2.

**Usage**

```
plot_ts_pred(
  x = NULL,
  y,
  yadj,
  ypred = NULL,
  label_x = "",
  label_y = "",
  color = "black",
  color_adjust = "blue",
  color_prediction = "green"
)
```

**Arguments**

x	time index
y	time series
yadj	adjustment of time series
ypred	prediction of the time series
label_x	x-axis title
label_y	y-axis title
color	color for the time series
color_adjust	color for the adjusted values
color_prediction	color for the predictions

**Value**

returns a ggplot graphic

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 0)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size= 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_arima()
model <- fit(model, x=io_train$input, y=io_train$output)
adjust <- predict(model, io_train$input)

prediction <- predict(model, x=io_test$input, steps_ahead=5)
prediction <- as.vector(prediction)

yvalues <- c(io_train$output, io_test$output)
grf <- plot_ts_pred(y=yvalues, yadj=adjust, ypre=prediction)
plot(grf)
```

---

predictor

*DAL Predict*

---

**Description**

Ancestor class for regression and classification. It provides basis for fit and predict methods. Besides, action method proxies to predict.

An example of learner is a decision tree (cla\_dtree)

**Usage**

```
predictor()
```

**Value**

returns a predictor object

**Examples**

```
#See ?cla_dtree for a classification example using a decision tree
```

---

R2.ts	<i>R2</i>
-------	-----------

---

**Description**

Compute the R-squared ( $R^2$ ) between actual values and forecasts of a time series

**Usage**

```
R2.ts(actual, prediction)
```

**Arguments**

actual	real observations
prediction	predicted observations

**Value**

returns a number, which is the calculated  $R^2$

---

regression	<i>Regression</i>
------------	-------------------

---

**Description**

Ancestor class for regression problems. This ancestor class is used to define and manage the target attribute for regression tasks.

**Usage**

```
regression(attribute)
```

**Arguments**

attribute	attribute target to model building
-----------	------------------------------------



**Value**

returns a regression object

**Examples**

```
#See ?reg_dtree for a regression example using a decision tree
```

---

reg\_dtree

*Decision Tree for regression*

---

**Description**

Creates a regression object that uses the Decision Tree method for regression It wraps the tree library.

**Usage**

```
reg_dtree(attribute)
```

**Arguments**

attribute      attribute target to model building.

**Value**

returns a decision tree regression object

**Examples**

```
data(Boston)
model <- reg_dtree("medv")

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

model <- fit(model, train)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

reg_knn	<i>knn regression</i>
---------	-----------------------

---

**Description**

Creates a regression object that uses the K-Nearest Neighbors (knn) method for regression

**Usage**

```
reg_knn(attribute, k)
```

**Arguments**

attribute	attribute target to model building
k	number of k neighbors

**Value**

returns a knn regression object

**Examples**

```
data(Boston)
model <- reg_knn("medv", k=3)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

model <- fit(model, train)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

reg_mlp	<i>MLP for regression</i>
---------	---------------------------

---

**Description**

Creates a regression object that uses the Multi-Layer Perceptron (MLP) method. It wraps the nnet library.

**Usage**

```
reg_mlp(attribute, size = NULL, decay = 0.05, maxit = 1000)
```

**Arguments**

attribute	attribute target to model building
size	number of neurons in hidden layers
decay	decay learning rate
maxit	number of maximum iterations for training

**Value**

returns a object of class reg\_mlp

**Examples**

```
data(Boston)
model <- reg_mlp("medv", size=5, decay=0.54)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

model <- fit(model, train)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

 reg\_rf

*Random Forest for regression*


---

**Description**

Creates a regression object that uses the Random Forest method. It wraps the randomForest library.

**Usage**

```
reg_rf(attribute, nodesize = 1, ntree = 10, mtry = NULL)
```

**Arguments**

attribute	attribute target to model building
nodesize	node size
ntree	number of trees
mtry	number of attributes to build tree

**Value**

returns an object of class `reg_rfobj`

**Examples**

```
data(Boston)
model <- reg_rf("medv", ntree=10)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

model <- fit(model, train)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

reg\_svm

*SVM for regression*

---

**Description**

Creates a regression object that uses the Support Vector Machine (SVM) method for regression. It wraps the `e1071` and `svm` library.

**Usage**

```
reg_svm(attribute, epsilon = 0.1, cost = 10, kernel = "radial")
```

**Arguments**

attribute	attribute target to model building
epsilon	parameter that controls the width of the margin around the separating hyperplane
cost	parameter that controls the trade-off between having a wide margin and correctly classifying training data points
kernel	the type of kernel function to be used in the SVM algorithm (linear, radial, polynomial, sigmoid)

**Value**

returns a SVM regression object

**Examples**

```
data(Boston)
model <- reg_svm("medv", epsilon=0.2, cost=40.000)

# preparing dataset for random sampling
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

model <- fit(model, train)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

reg\_tune

*Regression Tune*

---

**Description**

Creates an object for tuning regression models

**Usage**

```
reg_tune(base_model, folds = 10)
```

**Arguments**

base_model	base model for tuning
folds	number of folds for cross-validation

**Value**

returns a reg\_tune object.

**Examples**

```
# preparing dataset for random sampling
data(Boston)
sr <- sample_random()
sr <- train_test(sr, Boston)
train <- sr$train
test <- sr$test

# hyper parameter setup
tune <- reg_tune(reg_mlp("medv"))
ranges <- list(size=c(3), decay=c(0.1, 0.5))
```

```
# hyper parameter optimization
model <- fit(tune, train, ranges)

test_prediction <- predict(model, test)
test_predictand <- test[, "medv"]
test_eval <- evaluate(model, test_predictand, test_prediction)
test_eval$metrics
```

---

sample\_random

*Sample Random*

---

### Description

The `sample_random` function in R is used to generate a random sample of specified size from a given data set.

### Usage

```
sample_random()
```

### Value

returns an object of class 'sample\_random'

### Examples

```
#using random sampling
sample <- sample_random()
tt <- train_test(sample, iris)

# distribution of train
table(tt$train$Species)

# preparing dataset into four folds
folds <- k_fold(sample, iris, 4)

# distribution of folds
tbl <- NULL
for (f in folds) {
  tbl <- rbind(tbl, table(f$Species))
}
head(tbl)
```

---

sample_stratified	<i>Stratified Random Sampling</i>
-------------------	-----------------------------------

---

### Description

The `sample_stratified` function in R is used to generate a stratified random sample from a given dataset. Stratified sampling is a statistical method that is used when the population is divided into non-overlapping subgroups or strata, and a sample is selected from each stratum to represent the entire population. In stratified sampling, the sample is selected in such a way that it is representative of the entire population and the variability within each stratum is minimized.

### Usage

```
sample_stratified(attribute)
```

### Arguments

`attribute`      attribute target to model building

### Value

returns an object of class `sample_stratified`

### Examples

```
#using stratified sampling
sample <- sample_stratified("Species")
tt <- train_test(sample, iris)

# distribution of train
table(tt$train$Species)

# preparing dataset into four folds
folds <- k_fold(sample, iris, 4)

# distribution of folds
tbl <- NULL
for (f in folds) {
  tbl <- rbind(tbl, table(f$Species))
}
head(tbl)
```

---

select_hyper	<i>Selection hyper parameters</i>
--------------	-----------------------------------

---

**Description**

Selects the optimal hyperparameters from a dataset resulting from k-fold cross-validation

**Usage**

```
select_hyper(obj, hyperparameters)
```

**Arguments**

obj	the object or model used for hyperparameter selection.
hyperparameters	data set with hyper parameters and quality measure from execution

**Value**

returns the index of selected hyper parameter

---

select_hyper.cla_tune	<i>selection of hyperparameters</i>
-----------------------	-------------------------------------

---

**Description**

Selects the optimal hyperparameter by maximizing the average classification metric. It wraps dplyr library.

**Usage**

```
## S3 method for class 'cla_tune'
select_hyper(obj, hyperparameters)
```

**Arguments**

obj	an object representing the model or tuning process
hyperparameters	a dataframe with columns key (hyperparameter configuration) and metric (classification metric)

**Value**

returns a optimized key number of hyperparameters



---

select\_hyper.ts\_tune    *Select Optimal Hyperparameters for Time Series Models*

---

### Description

Identifies the optimal hyperparameters by minimizing the error from a dataset of hyperparameters. The function selects the hyperparameter configuration that results in the lowest average error. It wraps the dplyr library.

### Usage

```
## S3 method for class 'ts_tune'
select_hyper(obj, hyperparameters)
```

### Arguments

obj                    a ts\_tune object containing the model and tuning settings  
hyperparameters       hyperparameters dataset

### Value

returns the optimized key number of hyperparameters

---

set\_params            *Assign parameters*

---

### Description

set\_params function assigns all parameters to the attributes presented in the object.

### Usage

```
set_params(obj, params)
```

### Arguments

obj                    object of class dal\_base  
params                 parameters to set obj

### Value

returns an object with parameters set

### Examples

```
obj <- set_params(dal_base(), list(x = 0))
```

---

set_params.default	<i>Default Assign parameters</i>
--------------------	----------------------------------

---

**Description**

Default method for set\_params which returns the object unchanged

**Usage**

```
## Default S3 method:  
set_params(obj, params)
```

**Arguments**

obj	object
params	parameters

**Value**

returns the object unchanged

---

sin_data	<i>Time series example dataset</i>
----------	------------------------------------

---

**Description**

Synthetic dataset of sine function.

- x: correspond time from 0 to 10.
- y: dependent variable for time series modeling.

**Usage**

```
data(sin_data)
```

**Format**

data.frame.

**Source**

This dataset was generated for examples.

**Examples**

```
data(sin_data)  
head(sin_data)
```

---

`sMAPE.ts`*sMAPE*

---

**Description**

Compute the symmetric mean absolute percent error (sMAPE)

**Usage**

```
sMAPE.ts(actual, prediction)
```

**Arguments**

actual	real observations
prediction	predicted observations

**Value**

returns the sMAPE between the actual and prediction vectors

---

`smoothing`*Smoothing*

---

**Description**

Smoothing is a statistical technique used to reduce the noise in a signal or a dataset by removing the high-frequency components. The smoothing level is associated with the number of bins used. There are alternative methods to establish the smoothing: equal interval, equal frequency, and clustering.

**Usage**

```
smoothing(n)
```

**Arguments**

n	number of bins
---	----------------

**Value**

returns an object of class `smoothing`

**Examples**

```
data(iris)
obj <- smoothing_inter(n = 2)
obj <- fit(obj, iris$Sepal.Length)
sl.bi <- transform(obj, iris$Sepal.Length)
table(sl.bi)
obj$interval

entro <- evaluate(obj, as.factor(names(sl.bi)), iris$Species)
entro$entropy
```

---

smoothing\_cluster      *Smoothing by cluster*

---

**Description**

Uses clustering method to perform data smoothing. The input vector is divided into clusters using the k-means algorithm. The mean of each cluster is then calculated and used as the smoothed value for all observations within that cluster.

**Usage**

```
smoothing_cluster(n)
```

**Arguments**

n                      number of bins

**Value**

returns an object of class `smoothing_cluster`

**Examples**

```
data(iris)
obj <- smoothing_cluster(n = 2)
obj <- fit(obj, iris$Sepal.Length)
sl.bi <- transform(obj, iris$Sepal.Length)
table(sl.bi)
obj$interval

entro <- evaluate(obj, as.factor(names(sl.bi)), iris$Species)
entro$entropy
```

---

smoothing_freq	<i>Smoothing by Freq</i>
----------------	--------------------------

---

**Description**

The 'smoothing\_freq' function is used to smooth a given time series data by aggregating observations within a fixed frequency.

**Usage**

```
smoothing_freq(n)
```

**Arguments**

n	number of bins
---	----------------

**Value**

returns an object of class smoothing\_freq

**Examples**

```
data(iris)
obj <- smoothing_freq(n = 2)
obj <- fit(obj, iris$Sepal.Length)
sl.bi <- transform(obj, iris$Sepal.Length)
table(sl.bi)
obj$interval

entro <- evaluate(obj, as.factor(names(sl.bi)), iris$Species)
entro$entropy
```

---

smoothing_inter	<i>Smoothing by interval</i>
-----------------	------------------------------

---

**Description**

The "smoothing by interval" function is used to apply a smoothing technique to a vector or time series data using a moving window approach.

**Usage**

```
smoothing_inter(n)
```

**Arguments**

n	number of bins
---	----------------

**Value**

returns an object of class `smoothing_inter`

**Examples**

```
data(iris)
obj <- smoothing_inter(n = 2)
obj <- fit(obj, iris$Sepal.Length)
sl.bi <- transform(obj, iris$Sepal.Length)
table(sl.bi)
obj$interval

entro <- evaluate(obj, as.factor(names(sl.bi)), iris$Species)
entro$entropy
```

---

train\_test

*Train-Test Partition*

---

**Description**

Partitions a dataset into training and test sets using a specified sampling method

**Usage**

```
train_test(obj, data, perc = 0.8, ...)
```

**Arguments**

obj	an object of a class that supports the <code>train_test</code> method
data	dataset to be partitioned
perc	a numeric value between 0 and 1 specifying the proportion of data to be used for training
...	additional optional arguments passed to specific methods.

**Value**

returns a list with two elements:

- train: A data frame containing the training set
- test: A data frame containing the test set

**Examples**

```
#using random sampling
sample <- sample_random()
tt <- train_test(sample, iris)

# distribution of train
table(tt$train$Species)
```

---

train\_test\_from\_folds *k-fold training and test partition object*

---

### Description

Splits a dataset into training and test sets based on k-fold cross-validation. The function takes a list of data partitions (folds) and a specified fold index k. It returns the data corresponding to the k-th fold as the test set, and combines all other folds to form the training set.

### Usage

```
train_test_from_folds(folds, k)
```

### Arguments

folds	data partitioned into folds
k	k-fold for test set, all reminder for training set

### Value

returns a list with two elements:

- train: A data frame containing the combined data from all folds except the k-th fold, used as the training set.
- test: A data frame corresponding to the k-th fold, used as the test set.

### Examples

```
# Create k-fold partitions of a dataset (e.g., iris)
folds <- k_fold(sample_random(), iris, k = 5)

# Use the first fold as the test set and combine the remaining folds for the training set
train_test_split <- train_test_from_folds(folds, k = 1)

# Display the training set
head(train_test_split$train)

# Display the test set
head(train_test_split$test)
```

---

transform	<i>Transform</i>
-----------	------------------

---

**Description**

Defines a transformation method.

**Usage**

```
transform(obj, ...)
```

**Arguments**

obj	a dal_transform object.
...	optional arguments.

**Value**

returns a transformed data.

**Examples**

```
#See ?minmax for an example of transformation
```

---

ts_arima	<i>ARIMA</i>
----------	--------------

---

**Description**

Creates a time series prediction object that uses the AutoRegressive Integrated Moving Average (ARIMA). It wraps the forecast library.

**Usage**

```
ts_arima()
```

**Value**

returns a ts\_arima object.



**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 0)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_arima()
model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test
```

---

ts_conv1d	<i>Conv1D</i>
-----------	---------------

---

**Description**

Creates a time series prediction object that uses the Conv1D. It wraps the pytorch library.

**Usage**

```
ts_conv1d(preprocess = NA, input_size = NA, epochs = 10000L)
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
epochs	maximum number of epochs

**Value**

returns a ts\_conv1d object.

**Examples**

```
#See an example of using `ts_conv1d` at this
#[link](https://github.com/cefet-rj-dal/daltoolbox/blob/main/timeseries/ts_conv1d.ipynb)
```

---

ts_data	<i>ts_data</i>
---------	----------------

---

### Description

Time series data structure used in DAL Toolbox. It receives a vector (representing a time series) or a matrix *y* (representing a sliding windows). Internal *ts\_data* is matrix of sliding windows with size *sw*. If *sw* equals to zero, it store a time series as a single matrix column.

### Usage

```
ts_data(y, sw = 1)
```

### Arguments

<i>y</i>	output variable
<i>sw</i>	integer: sliding window size.

### Value

returns a *ts\_data* object.

### Examples

```
data(sin_data)
head(sin_data)

data <- ts_data(sin_data$y)
ts_head(data)

data10 <- ts_data(sin_data$y, 10)
ts_head(data10)
```

---

ts_elm	<i>ELM</i>
--------	------------

---

### Description

Creates a time series prediction object that uses the Extreme Learning Machine (ELM). It wraps the *elmNNRcpp* library.

### Usage

```
ts_elm(preprocess = NA, input_size = NA, nhid = NA, actfun = "purelin")
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
nhid	ensemble size
actfun	defines the type to use, possible values: 'sig', 'radbas', 'tribas', 'relu', 'purelin' (default).

**Value**

returns a ts\_elm object.

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_elm(ts_norm_gminmax(), input_size=4, nhid=3, actfun="purelin")
model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test
```

---

ts\_head

*Extract the First Observations from a ts\_data Object*


---

**Description**

Returns the first n observations from a ts\_data

**Usage**

```
ts_head(x, n = 6L, ...)
```

**Arguments**

x	ts_data object
n	number of rows to return
...	optional arguments

**Value**

returns the first n observations of a ts\_data

**Examples**

```
data(sin_data)
data10 <- ts_data(sin_data$y, 10)
ts_head(data10)
```

---

ts_knn	<i>KNN time series prediction</i>
--------	-----------------------------------

---

**Description**

Creates a prediction object that uses the K-Nearest Neighbors (KNN) method for time series regression

**Usage**

```
ts_knn(preprocess = NA, input_size = NA, k = NA)
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
k	number of k neighbors

**Value**

returns a ts\_knn object.

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_knn(ts_norm_gminmax(), input_size=4, k=3)
model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test
```

---

ts_lstm	<i>LSTM</i>
---------	-------------

---

**Description**

Creates a time series prediction object that uses the LSTM. It wraps the pytorch library.

**Usage**

```
ts_lstm(preprocess = NA, input_size = NA, epochs = 10000L)
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
epochs	maximum number of epochs

**Value**

returns a ts\_lstm object.

**Examples**

```
#See an example of using `ts_ts_lstmconv1d` at this
#[link](https://github.com/cefet-rj-dal/daltoolbox/blob/main/timeseries/ts_lstm.ipynb)
```

---

ts_mlp	<i>MLP</i>
--------	------------

---

**Description**

Creates a time series prediction object that uses the Multilayer Perceptron (MLP). It wraps the nnet library.

**Usage**

```
ts_mlp(preprocess = NA, input_size = NA, size = NA, decay = 0.01, maxit = 1000)
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
size	number of neurons inside hidden layer
decay	decay parameter for MLP
maxit	maximum number of iterations

**Value**

returns a ts\_mlp object.

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_mlp(ts_norm_gminmax(), input_size=4, size=4, decay=0)
model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test
```

---

ts\_norm\_an

*Time Series Adaptive Normalization*

---

**Description**

Transform data to a common scale while taking into account the changes in the statistical properties of the data over time.

**Usage**

```
ts_norm_an(remove_outliers = TRUE, nw = 0)
```

**Arguments**

remove\_outliers  
logical: if TRUE (default) outliers will be removed.

nw  
integer: window size.

**Value**

returns a ts\_norm\_an object.

**Examples**

```
# time series to normalize
data(sin_data)

# convert to sliding windows
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)
summary(ts[,10])

# normalization
preproc <- ts_norm_an()
preproc <- fit(preproc, ts)
tst <- transform(preproc, ts)
ts_head(tst, 3)
summary(tst[,10])
```

---

ts\_norm\_diff

*Time Series Diff*

---

**Description**

This function calculates the difference between the values of a time series.

**Usage**

```
ts_norm_diff(remove_outliers = TRUE)
```

**Arguments**

remove\_outliers  
logical: if TRUE (default) outliers will be removed.

**Value**

returns a ts\_norm\_diff object.

**Examples**

```
# time series to normalize
data(sin_data)

# convert to sliding windows
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)
summary(ts[,10])

# normalization
preproc <- ts_norm_diff()
preproc <- fit(preproc, ts)
```

```
tst <- transform(preproc, ts)
ts_head(tst, 3)
summary(tst[,9])
```

---

ts_norm_ean	<i>Time Series Adaptive Normalization (Exponential Moving Average - EMA)</i>
-------------	--

---

### Description

Creates a normalization object for time series data using an Exponential Moving Average (EMA) method. This normalization approach adapts to changes in the time series and optionally removes outliers.

### Usage

```
ts_norm_ean(remove_outliers = TRUE, nw = 0)
```

### Arguments

remove_outliers	logical: if TRUE (default) outliers will be removed.
nw	windows size

### Value

returns a ts\_norm\_ean object.

### Examples

```
# time series to normalize
data(sin_data)

# convert to sliding windows
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)
summary(ts[,10])

# normalization
preproc <- ts_norm_ean()
preproc <- fit(preproc, ts)
tst <- transform(preproc, ts)
ts_head(tst, 3)
summary(tst[,10])
```



---

ts_norm_gminmax	<i>Time Series Global Min-Max</i>
-----------------	-----------------------------------

---

**Description**

Rescales data, so the minimum value is mapped to 0 and the maximum value is mapped to 1.

**Usage**

```
ts_norm_gminmax(remove_outliers = TRUE)
```

**Arguments**

remove\_outliers  
logical: if TRUE (default) outliers will be removed.

**Value**

returns a ts\_norm\_gminmax object.

**Examples**

```
# time series to normalize
data(sin_data)

# convert to sliding windows
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)
summary(ts[,10])

# normalization
preproc <- ts_norm_gminmax()
preproc <- fit(preproc, ts)
tst <- transform(preproc, ts)
ts_head(tst, 3)
summary(tst[,10])
```

---

ts_norm_swminmax	<i>Time Series Sliding Window Min-Max</i>
------------------	---

---

**Description**

The ts\_norm\_swminmax function creates an object for normalizing a time series based on the "sliding window min-max scaling" method

**Usage**

```
ts_norm_swminmax(remove_outliers = TRUE)
```

**Arguments**

remove\_outliers  
logical: if TRUE (default) outliers will be removed.

**Value**

returns a ts\_norm\_swminmax object.

**Examples**

```
# time series to normalize
data(sin_data)

# convert to sliding windows
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)
summary(ts[,10])

# normalization
preproc <- ts_norm_swminmax()
preproc <- fit(preproc, ts)
tst <- transform(preproc, ts)
ts_head(tst, 3)
summary(tst[,10])
```

---

ts\_projection

*Time Series Projection*

---

**Description**

Separates a ts\_data object into input and output components for time series analysis. This function is useful for preparing data for modeling, where the input and output variables are extracted from a time series dataset.

**Usage**

```
ts_projection(ts)
```

**Arguments**

ts                   matrix or data.frame containing the time series.

**Value**

returns a ts\_projection object.

**Examples**

```
#setting up a ts_data
data(sin_data)
ts <- ts_data(sin_data$y, 10)

io <- ts_projection(ts)

#input data
ts_head(io$input)

#output data
ts_head(io$output)
```

---

ts_reg	<i>TSReg</i>
--------	--------------

---

**Description**

Time Series Regression directly from time series Ancestral class for non-sliding windows implementation.

**Usage**

```
ts_reg()
```

**Value**

returns ts\_reg object

**Examples**

```
#See ?ts_arma for an example using Auto-regressive Integrated Moving Average
```

---

ts_regsw	<i>TSRegSW</i>
----------	----------------

---

**Description**

Time Series Regression from Sliding Windows. Ancestral class for Machine Learning Implementation.

**Usage**

```
ts_regsw(preprocess = NA, input_size = NA)
```

**Arguments**

preprocess      normalization  
input\_size      input size for machine learning model

**Value**

returns a ts\_regrsw object

**Examples**

#See ?ts\_elm for an example using Extreme Learning Machine

---

ts_rf	<i>Random Forest</i>
-------	----------------------

---

**Description**

Creates a time series prediction object that uses the Random Forest. It wraps the randomForest library.

**Usage**

```
ts_rf(preprocess = NA, input_size = NA, nodesize = 1, ntree = 10, mtry = NULL)
```

**Arguments**

preprocess      normalization  
input\_size      input size for machine learning model  
nodesize        node size  
ntree            number of trees  
mtry            number of attributes to build tree

**Value**

returns a ts\_rf object.

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_rf(ts_norm_gminmax(), input_size=4, nodesize=3, ntree=50)
```

```

model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test

```

---

ts\_sample

*Time Series Sample*


---

## Description

Separates the `ts_data` into training and test. It separates the test size from the last observations minus an offset. The offset is important to allow replication under different recent origins. The data for train uses the number of rows of a `ts_data` minus the test size and offset.

## Usage

```
ts_sample(ts, test_size = 1, offset = 0)
```

## Arguments

<code>ts</code>	time series.
<code>test_size</code>	integer: size of test data (default = 1).
<code>offset</code>	integer: starting point (default = 0).

## Value

returns a list with the two samples

## Examples

```

#setting up a ts_data
data(sin_data)
ts <- ts_data(sin_data$y, 10)

#separating into train and test
test_size <- 3
samp <- ts_sample(ts, test_size)

#first five rows from training data
ts_head(samp$train, 5)

#last five rows from training data
ts_head(samp$train[-c(1:(nrow(samp$train)-5)),])

#testing data
ts_head(samp$test)

```

ts\_svm

*SVM***Description**

Creates a time series prediction object that uses the Support Vector Machine (SVM). It wraps the e1071 library.

**Usage**

```
ts_svm(
  preprocess = NA,
  input_size = NA,
  kernel = "radial",
  epsilon = 0,
  cost = 10
)
```

**Arguments**

preprocess	normalization
input_size	input size for machine learning model
kernel	SVM kernel (linear, radial, polynomial, sigmoid)
epsilon	error threshold
cost	this parameter controls the trade-off between achieving a low error on the training data and minimizing the model complexity

**Value**

returns a ts\_svm object.

**Examples**

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

model <- ts_svm(ts_norm_gminmax(), input_size=4)
model <- fit(model, x=io_train$input, y=io_train$output)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)
```

```
ev_test <- evaluate(model, output, prediction)
ev_test
```

---

ts\_tune

*Time Series Tune*


---

## Description

Creates a `ts_tune` object for tuning hyperparameters of a time series model. This function sets up a tuning process for the specified base model by exploring different configurations of hyperparameters using cross-validation.

## Usage

```
ts_tune(input_size, base_model, folds = 10)
```

## Arguments

<code>input_size</code>	input size for machine learning model
<code>base_model</code>	base model for tuning
<code>folds</code>	number of folds for cross-validation

## Value

returns a `ts_tune` object

## Examples

```
data(sin_data)
ts <- ts_data(sin_data$y, 10)
ts_head(ts, 3)

samp <- ts_sample(ts, test_size = 5)
io_train <- ts_projection(samp$train)
io_test <- ts_projection(samp$test)

tune <- ts_tune(input_size=c(3:5), base_model = ts_elm(ts_norm_gminmax()))
ranges <- list(nhid = 1:5, actfun=c('purelin'))

# Generic model tuning
model <- fit(tune, x=io_train$input, y=io_train$output, ranges)

prediction <- predict(model, x=io_test$input[1,], steps_ahead=5)
prediction <- as.vector(prediction)
output <- as.vector(io_test$output)

ev_test <- evaluate(model, output, prediction)
ev_test
```

---

zscore	<i>Z-score normalization</i>
--------	------------------------------

---

**Description**

Scale data using z-score normalization.

$$zscore = (x - mean(x))/sd(x)$$

**Usage**

```
zscore(nmean = 0, nsd = 1)
```

**Arguments**

nmean	new mean for normalized data
nsd	new standard deviation for normalized data

**Value**

returns the z-score transformation object

**Examples**

```
data(iris)
head(iris)

trans <- zscore()
trans <- fit(trans, iris)
tiris <- transform(trans, iris)
head(tiris)

itiris <- inverse_transform(trans, tiris)
head(itiris)
```

---

[.ts_data	<i>Subset Extraction for Time Series Data</i>
-----------	---

---

**Description**

Extracts a subset of a time series object based on specified rows and columns. The function allows for flexible indexing and subsetting of time series data.

**Usage**

```
## S3 method for class 'ts_data'
x[i, j, ...]
```



**Arguments**

x	ts_data object
i	row i
j	column j
...	optional arguments

**Value**

returns a new ts\_data object

**Examples**

```
data(sin_data)
data10 <- ts_data(sin_data$y, 10)
ts_head(data10)
#single line
data10[12,]

#range of lines
data10[12:13,]

#single column
data10[,1]

#range of columns
data10[,1:2]

#range of rows and columns
data10[12:13,1:2]

#single line and a range of columns
#'data10[12,1:2]

#range of lines and a single column
data10[12:13,1]

#single observation
data10[12,1]
```

# Index

## \* datasets

- Boston, 9
- sin\_data, 58
- [.ts\_data, 80

action, 4

action.dal\_transform, 5

adjust\_class\_label, 5

adjust\_data.frame, 6

adjust\_factor, 6

adjust\_matrix, 7

adjust\_ts\_data, 7

autoenc\_encode, 8

autoenc\_encode\_decode, 9

Boston, 9

categ\_mapping, 11

cla\_dtree, 12

cla\_knn, 13

cla\_majority, 14

cla\_mlp, 15

cla\_nb, 16

cla\_rf, 16

cla\_svm, 17

cla\_tune, 18

classification, 11

clu\_tune, 22

cluster, 19

cluster\_dbscan, 20

cluster\_kmeans, 21

cluster\_pam, 22

clusterer, 20

dal\_base, 23

dal\_learner, 24

dal\_transform, 24

dal\_tune, 25

data\_sample, 25

do\_fit, 26

do\_predict, 27

dt\_pca, 27

evaluate, 28

fit, 29

fit.cla\_tune, 29

fit.cluster\_dbscan, 30

fit\_curvature\_max, 30

fit\_curvature\_min, 31

inverse\_transform, 32

k\_fold, 32

minmax, 33

MSE.ts, 34

outliers, 34

plot\_bar, 35

plot\_boxplot, 36

plot\_boxplot\_class, 36

plot\_density, 37

plot\_density\_class, 38

plot\_groupedbar, 39

plot\_hist, 39

plot\_lollipop, 40

plot\_pieplot, 41

plot\_points, 42

plot\_radar, 43

plot\_scatter, 43

plot\_series, 44

plot\_stackedbar, 45

plot\_ts, 45

plot\_ts\_pred, 46

predictor, 47

R2.ts, 48

reg\_dtree, 49

reg\_knn, 50

reg\_mlp, 50  
reg\_rf, 51  
reg\_svm, 52  
reg\_tune, 53  
regression, 48

sample\_random, 54  
sample\_stratified, 55  
select\_hyper, 56  
select\_hyper.cla\_tune, 56  
select\_hyper.ts\_tune, 57  
set\_params, 57  
set\_params.default, 58  
sin\_data, 58  
sMAPE.ts, 59  
smoothing, 59  
smoothing\_cluster, 60  
smoothing\_freq, 61  
smoothing\_inter, 61

train\_test, 62  
train\_test\_from\_folds, 63  
transform, 64  
ts\_arima, 64  
ts\_conv1d, 65  
ts\_data, 66  
ts\_elm, 66  
ts\_head, 67  
ts\_knn, 68  
ts\_lstm, 69  
ts\_mlp, 69  
ts\_norm\_an, 70  
ts\_norm\_diff, 71  
ts\_norm\_ean, 72  
ts\_norm\_gminmax, 73  
ts\_norm\_swminmax, 73  
ts\_projection, 74  
ts\_reg, 75  
ts\_regsw, 75  
ts\_rf, 76  
ts\_sample, 77  
ts\_svm, 78  
ts\_tune, 79

zscore, 80