

# Package: Clustering (via r-universe)

September 17, 2026

**Type** Package

**Title** Techniques for Evaluating Clustering

**Description** The design of this package allows us to run different clustering packages and compare the results between them, to determine which algorithm behaves best from the data provided. See Martos, L.A.P., García-Vico, Á.M., González, P. et al.(2023) <[doi:10.1007/s13748-022-00294-2](https://doi.org/10.1007/s13748-022-00294-2)> ``Clustering: an R library to facilitate the analysis and comparison of cluster algorithms.", Martos, L.A.P., García-Vico, Á.M., González, P. et al. ``A Multiclustering Evolutionary Hyperrectangle-Based Algorithm" <[doi:10.1007/s44196-023-00341-3](https://doi.org/10.1007/s44196-023-00341-3)> and L.A.P., García-Vico, Á.M., González, P. et al. ``An Evolutionary Fuzzy System for Multiclustering in Data Streaming" <[doi:10.1016/j.procs.2023.12.058](https://doi.org/10.1016/j.procs.2023.12.058)>.

**Version** 1.7.10

**Date** 2024-04-20

**Author** Luis Alfonso Perez Martos [aut, cre]  
(<<https://orcid.org/0000-0002-5154-6105>>)

**Maintainer** Luis Alfonso Perez Martos <lapm0001@gmail.com>

**URL** <https://github.com/laperez/clustering>

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.3.1

**Imports** amap, apcluster, cluster, ClusterR, data.table, dplyr,  
foreach, future, ggplot2, gmp, methods, pracma, pvclust, shiny,  
sqldf, stats, tools, utils, xtable, toOrdinal

**Suggests** DT, shinyalert, shinyFiles, shinyjs, shinythemes,  
shinyWidgets, tidyverse, shinycssloaders

**Config/pak/sysreqs** cmake libgmp3-dev make libuv1-dev libssl-dev zlib1g-dev

**Repository** <https://laperez.r-universe.dev>

**Date/Publication** 2024-04-22 20:31:51 UTC  
**RemoteUrl** https://github.com/laperez/clustering  
**RemoteRef** HEAD  
**RemoteSha** c6acacce9f3197343745efce2727dfdead8f3a7b

Contents

|                                                        |           |
|--------------------------------------------------------|-----------|
| [.clustering . . . . .                                 | 2         |
| appClustering . . . . .                                | 3         |
| basketball . . . . .                                   | 4         |
| best_ranked_external_metrics . . . . .                 | 4         |
| best_ranked_internal_metrics . . . . .                 | 5         |
| bolts . . . . .                                        | 6         |
| clustering . . . . .                                   | 7         |
| convert_toOrdinal . . . . .                            | 9         |
| evaluate_best_validation_external_by_metrics . . . . . | 10        |
| evaluate_best_validation_internal_by_metrics . . . . . | 11        |
| evaluate_validation_external_by_metrics . . . . .      | 12        |
| evaluate_validation_internal_by_metrics . . . . .      | 13        |
| export_file_external . . . . .                         | 14        |
| export_file_internal . . . . .                         | 15        |
| plot_clustering . . . . .                              | 16        |
| result_external_algorithm_by_metric . . . . .          | 17        |
| result_internal_algorithm_by_metric . . . . .          | 17        |
| sort.clustering . . . . .                              | 18        |
| stock . . . . .                                        | 19        |
| stulong . . . . .                                      | 20        |
| transform_dataset . . . . .                            | 21        |
| transform_dataset_internal . . . . .                   | 21        |
| weather . . . . .                                      | 22        |
| <b>Index</b>                                           | <b>23</b> |

---

|              |                                                                                 |
|--------------|---------------------------------------------------------------------------------|
| [.clustering | <i>Filter metrics in a clustering object returning a new clustering object.</i> |
|--------------|---------------------------------------------------------------------------------|

---

Description

Generates a new filtered clustering object.

Usage

```
## S3 method for class 'clustering'  
clustering[condition = TRUE]
```

**Arguments**

|            |                                             |
|------------|---------------------------------------------|
| clustering | The clustering object to filter.            |
| condition  | Expression to filter the clustering object. |

**Details**

This function allows you to filter the data set for a given evaluation metric. The evaluation metrics available are: Algorithm, Distance, Clusters, Data, Var, Time, Entropy, Variation\_information, Precision, Recall, F\_measure, Fowlkes\_mallows\_index, Connectivity, Dunn, Silhouette and TimeAtt.

**Value**

A clustering object filtered from the input parameters.

**Examples**

```
result <- Clustering::clustering(df = Clustering::basketball, algorithm = 'clara',  
min=3, max=4, metrics = c('Precision','Recall'))  
  
result[Precision > 0.14 & Recall > 0.11]
```

---

appClustering

*Clustering GUI.*

---

**Description**

Method that allows us to execute the main algorithm in graphic interface mode instead of through the console.

**Usage**

```
appClustering()
```

**Details**

The operation of this method is to generate a graphical user. interface to be able to execute the clustering algorithm without knowing the parameters. Its operation is very simple, we can change the values and see the behavior quickly.

**Value**

GUI with the parameters of the algorithm and their representation in tables and graphs.

---

|            |                                                                                                  |
|------------|--------------------------------------------------------------------------------------------------|
| basketball | <i>This data set contains a series of statistics (5 attributes) about 96 basketball players:</i> |
|------------|--------------------------------------------------------------------------------------------------|

---

### Description

This data set contains a series of statistics about basketball players:

### Usage

```
data(basketball)
```

### Format

A data frame with 96 observations on 5 variables:

This data set contains a series of statistics about basketball players:

**assists\_per\_minuteReal** average number of assistances per minute

**heightInteger** height of the player

**time\_playedReal** time played by the player

**ageInteger** number of years of the player

**points\_per\_minuteReal** average number of points per minute

### Source

KEEL, <<http://www.keel.es/>>

---

|                              |                                     |
|------------------------------|-------------------------------------|
| best_ranked_external_metrics | <i>Best rated external metrics.</i> |
|------------------------------|-------------------------------------|

---

### Description

Method in charge of searching for each algorithm those that have the best external classification.

Method that looks for those external attribute that are better classified, making use of the var column.

In this way of discard attribute and only work with those that give the best response to the algorithm in question.

### Usage

```
best_ranked_external_metrics(df)
```

### Arguments

**df** Matrix or data frame with the result of running the clustering algorithm.

**Value**

Returns a data.frame with the best classified external attribute.

**Examples**

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 4,  
  algorithm='clara',  
  metrics=c("Recall")  
)  
  
Clustering::best_ranked_external_metrics(df = result)
```

---

best\_ranked\_internal\_metrics  
*Best rated internal metrics.*

---

**Description**

Method in charge of searching for each algorithm those that have the best internal classification.

Method that looks for those internal attributes that are better classified, making use of the Var column. In this way we discard the attributes and only work with those that give the best response to the algorithm in question.

**Usage**

```
best_ranked_internal_metrics(df)
```

**Arguments**

df                      Matrix or data frame with the result of running the clustering algorithm.

**Value**

Returns a data.frame with the best classified internal attributes.

**Examples**

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='gmm',  
  metrics=c("Recall")  
)
```

```
Clustering::best_ranked_internal_metrics(df = result)
```

---

|       |                                                                                                  |
|-------|--------------------------------------------------------------------------------------------------|
| bolts | <i>Data from an experiment on the affects of machine adjustments on the time to count bolts.</i> |
|-------|--------------------------------------------------------------------------------------------------|

---

## Description

A manufacturer of automotive accessories provides hardware, e.g. nuts, bolts, washers and screws, to fasten the accessory to the car or truck. Hardware is counted and packaged automatically. Specifically, bolts are dumped into a large metal dish. A plate that forms the bottom of the dish rotates counterclockwise. This rotation forces bolts to the outside of the dish and up along a narrow ledge. Due to the vibration of the dish caused by the spinning bottom plate, some bolts fall off the ledge and back into the dish. The ledge spirals up to a point where the bolts are allowed to drop into a pan on a conveyor belt. As a bolt drops, it passes by an electronic eye that counts it. When the electronic counter reaches the preset number of bolts, the rotation is stopped and the conveyor belt is moved forward

## Usage

```
data(bolts)
```

## Format

A data frame with 40 observations on 8 variables:

A manufacturer of automotive accessories provides hardware, e.g. nuts, bolts, washers and screws, to fasten the accessory to the car or truck. Hardware is counted and packaged automatically. Specifically, bolts are dumped into a large metal dish. A plate that forms the bottom of the dish rotates counterclockwise. This rotation forces bolts to the outside of the dish and up along a narrow ledge. Due to the vibration of the dish caused by the spinning bottom plate, some bolts fall off the ledge and back into the dish. The ledge spirals up to a point where the bolts are allowed to drop into a pan on a conveyor belt. As a bolt drops, it passes by an electronic eye that counts it. When the electronic counter reaches the preset number of bolts, the rotation is stopped and the conveyor belt is moved forward

**RUNInteger** is the order in which the data were collected

**SPEED1Integer** a speed setting that controls the speed of rotation of the plate at the bottom of the dish

**TOTALInteger** total number of bolts (TOTAL) to be counted

**SPEED2Integer** a second speed setting hat is used to change the speed of rotation (usually slowing it down) for the last few bolts

**NUMBER2Integer** the number of bolts to be counted at this second speed

**SENSInteger** the sensitivity of the electronic eye

**TIMERReal** The measured response is the time, in seconds

**T20BOLTReal** In order to put times on a equal footing the response to be analyzed is the time to count 20 bolts

### Details

There are several adjustments on the machine that affect its operation. These include; a speed setting that controls the speed of rotation (SPEED1Integer) of the plate at the bottom of the dish, a total number of bolts (TOTAL) to be counted, a second speed setting (SPEED2Integer) that is used to change the speed of rotation (usually slowing it down) for the last few bolts, the number of bolts to be counted at this second speed (NUMBER2Integer), and the sensitivity of the electronic eye (SENSInteger). The sensitivity setting is to insure that the correct number of bolts are counted. Too few bolts packaged causes customer complaints. Too many bolts packaged increases costs. For each run conducted in this experiment the correct number of bolts was counted. From an engineering standpoint if the correct number of bolts is counted, the sensitivity should not affect the time to count bolts. The measured response is the time (TIMERReal), in seconds, it takes to count the desired number of bolts. In order to put times on a equal footing the response to be analyzed is the time to count 20 bolts (T20BOLTReal). Below are the data for 40 combinations of settings. RUNInteger is the order in which the data were collected.

### Source

KEEL, <<http://www.keel.es/>>

---

clustering

*Clustering algorithm.*

---

### Description

Discovering the behavior of attributes in a set of clustering packages based on evaluation metrics.

### Usage

```
clustering(
  path = NULL,
  df = NULL,
  packages = NULL,
  algorithm = NULL,
  min = 3,
  max = 4,
  metrics = NULL
)
```

**Arguments**

|                        |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>path</code>      | The path of file. NULL It is only allowed to use <code>path</code> or <code>df</code> but not both at the same time. Only files in <code>.dat</code> , <code>.csv</code> or <code>arff</code> format are allowed.                                                                                                                                                                                                                    |
| <code>df</code>        | data matrix or data frame, or dissimilarity matrix. NULL If you want to use training and test basketball attributes.                                                                                                                                                                                                                                                                                                                 |
| <code>packages</code>  | character vector with the packages running the algorithm.<br>NULL The seven packages implemented are: <code>cluster</code> , <code>ClusterR</code> , <code>amap</code> , <code>apcluster</code> , <code>pvclust</code> .<br>By default runs all packages.                                                                                                                                                                            |
| <code>algorithm</code> | character vector with the algorithms implemented within the package. NULL The algorithms implemented are: <code>hclust</code> , <code>apclusterK</code> , <code>agnes</code> , <code>clara</code> , <code>daisy</code> , <code>diana</code> , <code>fanny</code> , <code>mona</code> , <code>pam</code> , <code>gmm</code> , <code>kmeans_arma</code> , <code>kmeans_rcpp</code> , <code>mini_kmeans</code> , <code>pvclust</code> . |
| <code>min</code>       | An integer with the minimum number of clusters This data is necessary to indicate the minimum number of clusters when grouping the data. The default value is 3.                                                                                                                                                                                                                                                                     |
| <code>max</code>       | An integer with the maximum number of clusters. This data is necessary to indicate the maximum number of clusters when grouping the data. The default value is 4.                                                                                                                                                                                                                                                                    |
| <code>metrics</code>   | Character vector with the metrics implemented to evaluate the distribution of the data in clusters. NULL The metrics implemented are: <code>Entropy</code> , <code>Variation_information</code> , <code>Precision</code> , <code>Recall</code> , <code>F_measure</code> , <code>Fowlkes_mallows_index</code> , <code>Connectivity</code> , <code>Dunn</code> and <code>Silhouette</code> .                                           |

**Details**

The operation of this algorithm is to evaluate how the attributes of a dataset or a set of datasets behave in different clustering algorithms. To do this, it is necessary to indicate the type of evaluation you want to make on the distribution of the data. To be able to execute the algorithm it is necessary to indicate the number of clusters.

`min` and `max`, the algorithms `algorithm` or `packages`.

`packages` that we want to cluster and the metrics `metrics`.

**Value**

A matrix with the result of running all the metrics of the algorithms contained in the packages indicated. We also obtain information with the types of metrics, algorithms and packages executed.

- `result` It is a list with the algorithms, metrics and variables defined in the execution of the algorithm.
- `has_internal_metrics` Boolean field to indicate if there are internal metrics such as: `dunn`, `silhouette` and `connectivity`.
- `has_external_metrics` Boolean field to indicate if there are external metrics such as: `precision`, `recall`, `f-measure`, `entropy`, `variation information` and `fowlkes-mallows`.

- algorithms\_execute Character vector with the algorithms executed. These algorithms have been mentioned in the definition of the parameters.
- measures\_execute Character vector with the measures executed. These measures have been mentioned in the definition of the parameters.

## Examples

```
Clustering::clustering(  
  df = cluster::agriculture,  
  min = 3,  
  max = 3,  
  algorithm='clara',  
  metrics=c('Precision')  
)
```

---

|                   |                                              |
|-------------------|----------------------------------------------|
| convert_toOrdinal | <i>Method to convert columns to ordinal.</i> |
|-------------------|----------------------------------------------|

---

## Description

Method to convert columns to ordinal.

## Usage

```
convert_toOrdinal(df)
```

## Arguments

df                      data frame with the results.

## Value

convert data frame to Ordinal.

---

`evaluate_best_validation_external_by_metrics`*Evaluates algorithms by measures of dissimilarity based on a metric.*

---

## Description

Method that calculates which algorithm and which metric behaves best for the datasets provided.

## Usage

```
evaluate_best_validation_external_by_metrics(df, metric)
```

## Arguments

|                     |                                                                                |
|---------------------|--------------------------------------------------------------------------------|
| <code>df</code>     | Data matrix or data frame with the result of running the clustering algorithm. |
| <code>metric</code> | String with the metric.                                                        |

## Details

Method groups the data by algorithm and distance measure, instead of obtaining the best attribute from the data set.

## Value

A data.frame with the algorithms classified by measures of dissimilarity.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='kmeans_rcpp',  
  metrics=c("F_measure"))  
  
Clustering::evaluate_best_validation_external_by_metrics(result, 'F_measure')
```

---

`evaluate_best_validation_internal_by_metrics`*Evaluates algorithms by measures of dissimilarity based on a metric.*

---

## Description

Method that calculates which algorithm and which metric behaves best for the datasets provided.

## Usage

```
evaluate_best_validation_internal_by_metrics(df, metric)
```

## Arguments

|                     |                                                                                |
|---------------------|--------------------------------------------------------------------------------|
| <code>df</code>     | Data matrix or data frame with the result of running the clustering algorithm. |
| <code>metric</code> | It's a string with the metric to evaluate.                                     |

## Details

This method groups the data by algorithm and distance measure, instead of obtaining the best attribute from the data set.

## Value

A data.frame with the algorithms classified by measures of dissimilarity.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='gmm',  
  metrics=c("Precision", "Connectivity")  
)
```

```
Clustering::evaluate_best_validation_internal_by_metrics(result, "Connectivity")
```

---

`evaluate_validation_external_by_metrics`*Evaluate external validations by algorithm.*

---

## Description

Method that calculates which algorithm behaves best for the datasets provided.

## Usage

```
evaluate_validation_external_by_metrics(df)
```

## Arguments

`df` data matrix or data frame with the result of running the clustering algorithm.

## Details

It groups the results of the execution by algorithms.

## Value

A `data.frame` with all the algorithms that obtain the best results regardless of the dissimilarity measure used.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 4,  
  algorithm='kmeans_arma',  
  metrics=c("Precision")  
)
```

```
Clustering::evaluate_validation_external_by_metrics(result)
```

---

evaluate\_validation\_internal\_by\_metrics  
*Evaluate internal validations by algorithm.*

---

## Description

Method that calculates which algorithm behaves best for the datasets provided.

## Usage

```
evaluate_validation_internal_by_metrics(df)
```

## Arguments

df                      data matrix or data frame with the result of running the clustering algorithm.

## Details

It groups the results of the execution by algorithms.

## Value

A data.frame with all the algorithms that obtain the best results regardless of the dissimilarity measure used.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='kmeans_rcpp',  
  metrics=c("Recall","Silhouette")  
)  
  
Clustering::evaluate_validation_internal_by_metrics(result)  
  
Clustering::evaluate_validation_internal_by_metrics(result$result)
```

---

|                      |                                                    |
|----------------------|----------------------------------------------------|
| export_file_external | <i>Export result of external metrics in latex.</i> |
|----------------------|----------------------------------------------------|

---

## Description

Method that exports the results of external measurements in latex format to a file.

## Usage

```
export_file_external(df, path = NULL)
```

## Arguments

|      |                                                                                                                     |
|------|---------------------------------------------------------------------------------------------------------------------|
| df   | It's a dataframe that contains as a parameter a table in latex format with the results of the external validations. |
| path | It's a string with the path to a directory where a file is to be stored in latex format.                            |

## Details

When we work in latex format and we need to create a table to export the results, with this method we can export the results of the clustering algorithm to latex.

## Value

A file in Latex format with the results of the external metrics.

## Examples

```
result = Clustering::clustering(
  df = cluster::agriculture,
  min = 4,
  max = 5,
  algorithm='gmm',
  metrics=c("Precision")
)

Clustering::export_file_external(result)
file.remove("external_data.tex")
```

---

|                      |                                                    |
|----------------------|----------------------------------------------------|
| export_file_internal | <i>Export result of internal metrics in latex.</i> |
|----------------------|----------------------------------------------------|

---

## Description

Method that exports the results of internal measurements in latex format to a file.

## Usage

```
export_file_internal(df, path = NULL)
```

## Arguments

|      |                                                                                                                     |
|------|---------------------------------------------------------------------------------------------------------------------|
| df   | It's a dataframe that contains as a parameter a table in latex format with the results of the internal validations. |
| path | It's a string with the path to a directory where a file is to be stored in latex format.                            |

## Details

When we work in latex format and we need to create a table to export the results, with this method we can export the results of the clustering algorithm to latex.

## Value

A file in Latex format with the results of the internal metrics.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='gmm',  
  metrics=c("Recall", "Dunn")  
)  
  
Clustering::export_file_internal(result)  
file.remove("internal_data.tex")
```

---

|                 |                                                           |
|-----------------|-----------------------------------------------------------|
| plot_clustering | <i>Graphic representation of the evaluation measures.</i> |
|-----------------|-----------------------------------------------------------|

---

## Description

Graphical representation of the evaluation measures grouped by cluster.

## Usage

```
plot_clustering(df, metric)
```

## Arguments

|        |                                                                                |
|--------|--------------------------------------------------------------------------------|
| df     | data matrix or data frame with the result of running the clustering algorithm. |
| metric | it's a string with the name of the metric select to evaluate.                  |

## Details

In certain cases the review or filtering of the data is necessary to select the data, that is why thanks to the graphic representations this task is much easier. Therefore with this method we will be able to filter the data by metrics and see the data in a graphical way.

## Value

Generate an image with the distribution of the clusters by metrics.

## Examples

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='gmm',  
  metrics=c("Precision")  
)  
  
Clustering::plot_clustering(result,c("Precision"))
```

---

result\_external\_algorithm\_by\_metric  
*External results by algorithm.*

---

**Description**

It is used for obtaining the results of an algorithm indicated as a parameter grouped by number of clusters.

**Usage**

```
result_external_algorithm_by_metric(df, metric)
```

**Arguments**

|        |                                                                                |
|--------|--------------------------------------------------------------------------------|
| df     | data matrix or data frame with the result of running the clustering algorithm. |
| metric | It's a string with the metric to evaluate.                                     |

**Value**

A data.frame with the results of the algorithm indicated as parameter.

**Examples**

```
result = Clustering::clustering(  
  df = cluster::agriculture,  
  min = 4,  
  max = 5,  
  algorithm='gmm',  
  metrics=c("Precision")  
)  
  
Clustering::result_external_algorithm_by_metric(result, 'Precision')
```

---

result\_internal\_algorithm\_by\_metric  
*Internal results by algorithm*

---

**Description**

It is used for obtaining the results of an algorithm indicated as a parameter grouped by number of clusters.

**Usage**

```
result_internal_algorithm_by_metric(df, metric)
```

**Arguments**

|        |                                                                                |
|--------|--------------------------------------------------------------------------------|
| df     | data matrix or data frame with the result of running the clustering algorithm. |
| metric | It's a string with the metric we want to evaluate your results.                |

**Value**

A data.frame with the results of the algorithm indicated as parameter.

**Examples**

```
result = Clustering::clustering(
  df = cluster::agriculture,
  min = 4,
  max = 5,
  algorithm='gmm',
  metrics=c("Recall", "Silhouette")
)

Clustering::result_internal_algorithm_by_metric(result, 'Silhouette')
```

---

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| sort.clustering | <i>Returns the clustering result sorted by a set of metrics.</i> |
|-----------------|------------------------------------------------------------------|

---

**Description**

This function receives a clustering object and sorts the columns by parameter. By default it performs sorting by the algorithm field.

**Usage**

```
## S3 method for class 'clustering'
sort(x, decreasing = TRUE, ...)
```

**Arguments**

|            |                                                                                                                                                                                                                                                                                                 |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| x          | It's an clustering object.                                                                                                                                                                                                                                                                      |
| decreasing | A logical indicating if the sort should be increasing or decreasing. By default, decreasing.                                                                                                                                                                                                    |
| ...        | Additional parameters as "by", a String with the name of the evaluation measure to order by. Valid values are: Algorithm, Distance, Clusters, Data, Var, Time, Entropy, Variation_information, Precision, Recall, F_measure, Fowlkes_mallows_index, Connectivity, Dunn, Silhouette and TimeAtt. |

### Details

The additional argument in "... " is the 'by' argument, which is a array with the name of the evaluation measure to order by. Valid value are: Algorithm, Distance, Clusters, Data, Var, Time, Entropy, Variation\_information, Precision, Recall, F\_measure, Fowlkes\_mallows\_index, Connectivity, Dunn, Silhouette, TimeAtt.

### Value

Another clustering object with the evaluation measures sorted

### Examples

```
result <-
Clustering::clustering(df = cluster::agriculture,min = 4, max = 4,algorithm='gmm',
metrics='Recall')

sort(result, FALSE, 'Recall')
```

---

stock

*The data provided are daily stock prices from January 1988 through October 1991, for ten aerospace companies.*

---

### Description

The data provided are daily stock prices from January 1988 through October 1991, for ten aerospace companies.

### Usage

```
data(stock)
```

### Format

A data frame with 950 observations on 10 variables:

The data provided are daily stock prices from January 1988 through October 1991, for ten aerospace companies.

**Company1** company1 details

**Company2** company2 details

**Company3** company3 details

**Company4** company4 details

**Company5** company5 details

**Company6** company6 details

**Company7** company7 details

**Company8** company8 details

**Company9** company9 details

**Company10** company10 details

## Source

KEEL, <<http://www.keel.es/>>

---

stulong

*The study was performed at the 2nd Department of Medicine, 1st Faculty of Medicine of Charles University and Charles University Hospital. The data were transferred to electronic form by the European Centre of Medical Informatics, Statistics and Epidemiology of Charles University and Academy of Sciences.*

---

## Description

The study was performed at the 2nd Department of Medicine, 1st Faculty of Medicine of Charles University and Charles University Hospital. The data were transferred to electronic form by the European Centre of Medical Informatics, Statistics and Epidemiology of Charles University and Academy of Sciences.

## Usage

data(stulong)

## Format

A data frame with 1417 observations on 5 variables.

The study was performed at the 2nd Department of Medicine, 1st Faculty of Medicine of Charles University and Charles University Hospital. The data were transferred to electronic form by the European Centre of Medical Informatics, Statistics and Epidemiology of Charles University and Academy of Sciences.

**a1** Height

**a2** Weight

**a3** Blood pressure I systolic (mm Hg)

**a4** Blood pressure I diastolic (mm Hg)

**a5** percentage Cholesterol in mg

## Source

KEEL, <<http://www.keel.es/>>

---

|                   |                                                            |
|-------------------|------------------------------------------------------------|
| transform_dataset | <i>Method for filtering external columns of a dataset.</i> |
|-------------------|------------------------------------------------------------|

---

**Description**

Method for filtering external columns of a dataset.

**Usage**

```
transform_dataset(df)
```

**Arguments**

|    |                                     |
|----|-------------------------------------|
| df | Data frame with clustering results. |
|----|-------------------------------------|

**Value**

Data frame filtered with the columns of the external measurements.

Exists internal measure

---

|                            |                                                            |
|----------------------------|------------------------------------------------------------|
| transform_dataset_internal | <i>Method for filtering internal columns of a dataset.</i> |
|----------------------------|------------------------------------------------------------|

---

**Description**

Method for filtering internal columns of a dataset.

**Usage**

```
transform_dataset_internal(df)
```

**Arguments**

|    |                                     |
|----|-------------------------------------|
| df | data frame with clustering results. |
|----|-------------------------------------|

**Value**

data frame filtered with the columns of the internal measurements.

Exists internal measure

---

weather

*One of the most known testing data sets in machine learning. This data sets describes several situations where the weather is suitable or not to play sports, depending on the current outlook, temperature, humidity and wind.*

---

### Description

One of the most known testing data sets in machine learning. This data sets describes several situations where the weather is suitable or not to play sports, depending on the current outlook, temperature, humidity and wind.

### Usage

```
data(weather)
```

### Format

A data frame with 14 observations on 5 variables:

One of the most known testing data sets in machine learning. This data sets describes several situations where the weather is suitable or not to play sports, depending on the current outlook, temperature, humidity and wind.

**Outlook** sunny, overcast, rainy

**Temperature** hot, mild, cool

**Humidity** high, normal

**Windy** true, false

**Play** yes, no

### Source

KEEL, <<http://www.keel.es/>>

# Index

- \* **datasets**
  - basketball, [4](#)
  - bolts, [6](#)
  - stock, [19](#)
  - stulong, [20](#)
  - weather, [22](#)
- [.clustering, [2](#)
- appClustering, [3](#)
- basketball, [4](#)
- best\_ranked\_external\_metrics, [4](#)
- best\_ranked\_internal\_metrics, [5](#)
- bolts, [6](#)
- clustering, [7](#)
- convert\_toOrdinal, [9](#)
- evaluate\_best\_validation\_external\_by\_metrics,  
[10](#)
- evaluate\_best\_validation\_internal\_by\_metrics,  
[11](#)
- evaluate\_validation\_external\_by\_metrics,  
[12](#)
- evaluate\_validation\_internal\_by\_metrics,  
[13](#)
- export\_file\_external, [14](#)
- export\_file\_internal, [15](#)
- plot\_clustering, [16](#)
- result\_external\_algorithm\_by\_metric,  
[17](#)
- result\_internal\_algorithm\_by\_metric,  
[17](#)
- sort.clustering, [18](#)
- stock, [19](#)
- stulong, [20](#)
- transform\_dataset, [21](#)
- transform\_dataset\_internal, [21](#)
- weather, [22](#)