

Package ‘kmscv’

June 28, 2022

Title K-Means Based Stratified Resampling

Version 0.1.0

Description An 'rsample' extension to create cluster based stratified resamples for cross-validation.

License GPL (>= 3)

Encoding UTF-8

Imports parsnip,rsample,tibble,purrr,dplyr

RoxygenNote 7.2.0

BugReports <https://github.com/sametsoekel/kmscv/issues>

NeedsCompilation no

Author Samet Sokel [aut, cre] (<<https://orcid.org/0000-0002-4429-5125>>)

Maintainer Samet Sokel <a_s@eskisehir.edu.tr>

Repository CRAN

Date/Publication 2022-06-28 16:20:08 UTC

R topics documented:

kms_cv	1
Index	3

kms_cv	<i>K-Means Based Cross-Validation</i>
--------	---------------------------------------

Description

K-Means based cross-validation splits data into clusters and creates random splits the data into K groups of roughly equal proportions (called "folds") as the main data.

Usage

```
kms_cv(data, k = 5, centers = NULL, max_center = 20, alpha = 0.1, seed = NULL)
```

Arguments

<code>data</code>	A data frame or data structure that can be converted to a data frame.
<code>k</code>	The number of partitions of the data set.
<code>centers</code>	Number of clusters to be created. If NULL, centers are selected automatically based on scaled inertia.
<code>max_center</code>	If <code>centers</code> is NULL number of clusters to be tried.
<code>alpha</code>	Alpha coefficient to penalize inertias.
<code>seed</code>	Seed number to reproduce randomness.

Value

A tibble with classes `manual_rset`, `rset`, `tbl_df`, `tbl`, and `data.frame`. The results include a column for the data split objects and one or more identification variables. There is a column called `id` that has a character string with the fold identifier.

Examples

```
library(kmscv)
```

```
kms_cv(iris,k = 5,centers = NULL,max_center = 20,alpha = .1,seed = 1)
```

Index

kms_cv, 1