Package 'featurizer'

October 13, 2022

Title Some Helper Functions that Help Create Features from Data

Version 0.2

Date 2017-06-11

Author Chhavi Choudhury [aut, cre]

Maintainer Chhavi Choudhury <chhavi.choudhury@gmail.com>

Description A collection of functions that would help one to build features based on external data. Very useful for Data Scientists in data to day work. Many functions create features using parallel computation. Since the nitty gritty of parallel computation is hidden under the hood, the user need not worry about creating clusters and shutting them down.

Depends R (>= 3.2.1), utils, parallel

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-06-11 23:12:50 UTC

R topics documented:

holiday_lst *Fedral Holidays DataFrame*

Description

This is a simple function that returns a data frame of federal holidays and date on which they were observed. It has only 2 columns, Date and Holiday for years 2012-2020.The functions does not take any parameter but returns a data frame.

Usage

```
holiday_lst()
```

Value

Returns a data frame if fedral holidays and when they were observed from 2012 - 2020

Examples

```
holiday_lst()
```

infer_gender *Infer gender for list of names fast*

Description

This function used a database of names and corresponding gender to lookup the gender of a name. It takes in a vector of names and returns a vector indicating gender for the name. The gender is denoted by "m"/"f"/"u" for male/female/unknown respectively.

Usage

```
infer_gender(x, detect_cores)
```

Arguments

x	The vector of names to perform paralleized inference on
detect_cores	If False (default), 2 cores are used. If True, half the number of cores are used on Mac OS else 2 cores are used.

Value

The a vector with m/f/u if the name in the original vector was male female or Unknown respectively

Examples

```
infer_gender(x=c("abby", "Nick", "abc"))
```

`infer_holiday` *Infer if a date is a fedral holiday*

Description

A function that returns the federal holiday observed on a given date. If there was no holiday observed, it returns None. The function takes in a list of dates, where each date is in the format YYYY-MM-DD and checks them against a database of federal holidays (2012-2020). It returns the holiday that was/will be observed on the date. If no holiday is observed on that date, it returns "None" corresponding to that date. This computation happens in parallel and hence is very fast.

Usage

```
infer_holiday(date_lst, detect_cores)
```

Arguments

`date_lst`	a list of dates strings of the format YYYY-MM-DD
`detect_cores`	If False (default), 2 cores are used. If True, half the number of cores are used on Mac OS else 2 cores are used.

Value

Returns the kind of fedral holiday that was/will be observed on that date. If there was no holiday observed, it returns None. It does the lookup in parallel and hence is very fast.

Examples

```
infer_holiday(date_lst = c("2020-12-25", "2020-09-07", "2020-09-08", "2016-01-18",
 "2016-01-19", "2016-01-15"))
```

`par_sapply` *A friendlier parSapply.*

Description

A sleek wrapper on the function parSapply. This function does not require user to allocate clusters and then stop them. This function uses parSapply across half the available cores and after the computation is done, it stops the clusters as well. The inputs are same os parSapply from parallel package

Usage

```
par_sapply(X, FUNC, detect_cores)
```

Arguments

X	The vector of value to perform paralleized sapply on.
FUNC	Function to be applied to every value of X, similar to sapply
detect_cores	If False (default), 2 cores are used. If True, half the number of cores are used on Mac OS else 2 cores are used.

Value

The a vector just like sapply. It does applies the function on X in parallel and hence is very fast.

Examples

```
par_sapply(1:100, function(x) x*100)
```

Index