Package 'reidentify'

July 23, 2019

Type Package

Title Reidentify and check the safety of Anonymized data

Version 0.0.0.900

Author @gg_hatano

Maintainer @gg_hatano <adad0405+reidentify@gmail.com>

Description Check the safety of Anonymized data by estimating reid-attack success rate.

License MIT + file LICENSE

Encoding UTF-8

Imports dplyr,
magrittr,
openssl,
stringi,
tibble,
philentropy

LazyData true

RoxygenNote 6.1.1

R topics documented:

| calc_KL | *calc KL divergence from 2 character vector which has distribution expression (A:B:C:...)* |

Description

calc KL divergence from 2 character vector which has distribution expression (A:B:C:...)

Usage

```
calc_KL(x, y, split = ":")
```

Arguments

x	vector
y	vector
split	split (default: ":")

---

create_dummy_master_data
create dummy master data

Description

create dummy master data

Usage

```
create_dummy_master_data(people = 100)
```

Arguments

| people | number of people |

---

create_dummy_transaction_data
create dummy transaction data

Description

create dummy transaction data

Usage

```
create_dummy_transaction_data(people = 100, size = 2)
```

Arguments

people	number of people
size	mean record number

Examples

```
data_tran = create_dummy_transaction_data(people = 10, size = 4)
```

---

distribution_distance

calculation distributin distance(L2) from 2 character vector which has distribution expression (A:B:C:...)

Description

calculation distributin distance(L2) from 2 character vector which has distribution expression (A:B:C:...)

Usage

```
distribution_distance(x, y, split = ":")
```

Arguments

x	vector
y	vector
split	split (default: ":")

---

join_raw_anon_data *create reid-format data from raw and anon data frame*

Description

create reid-format data from raw and anon data frame

Usage

```
join_raw_anon_data(raw, anon, raw_header = "RAW_",
  anon_header = "ANON_")
```

Arguments

raw	raw data frame
anon	anonymized data frame
raw_header	strings which is added for columns from raw data
anon_header	strings which is added for columns from anon data

---

reid_by_char   *reidentify by character static column*

Description

reidentify by character static column

Usage

```
reid_by_char(dat_raw_anon, target, row_number = "ROW_NUMBER")
```

Arguments

dat_raw_anon   dataframe of raw_anon form

target   target column

row_number   row number column name(default: "ROW_NUMBER")

---

reid_by_dist   *reidentify by single list column (list A, B, C is expressed by "A:B:C")*

Description

reidentify by single list column (list A, B, C is expressed by "A:B:C")

Usage

```
reid_by_dist(dat_raw_anon, target, row_number = "ROW_NUMBER",
  split = ":")
```

Arguments

dat_raw_anon   dataframe of raw_anon form

target   target column

row_number   row number column name(default: "ROW_NUMBER")

split   character for split _DIST value (default: ":")

| `reid_by_num` | *reidentify by single num static column by using L2 norm* |

Description

reidentify by single num static column by using L2 norm

Usage

```
reid_by_num(dat_raw_anon, target, row_number = "ROW_NUMBER")
```

Arguments

`dat_raw_anon`	dataframe of raw_anon form
`target`	target column
`row_number`	row number column name(default: "ROW_NUMBER")

| `reid_by_num_rank` | *reidentify by single num static by using rank* |

Description

reidentify by single num static by using rank

Usage

```
reid_by_num_rank(dat_raw_anon, target, row_number = "ROW_NUMBER")
```

Arguments

`dat_raw_anon`	dataframe of raw_anon form
`target`	target column
`row_number`	row number column name(default: "ROW_NUMBER")

---

reid_result	*craete text of reidentify result (method: ******, success/trial : ***** / ******)*

Description

craete text of reidentify result (method: ******, success/trial : ***** / ******)

Usage

```
reid_result(dat_reid_result, raw_row_number = "RAW_ROW_NUMBER",
  anon_row_number = "ANON_ROW_NUMBER", result = "RESULT",
  method = NULL)
```

Arguments

dat_reid_result	
	reid result data frame (RAW_ROW_NUMBER, ANON_ANON_NUMBER, RESULT)
raw_row_number	
	column name of row number in RAW data
anon_row_number	
	column name of row number in ANON data
result	true or false
method	reid method name

---

transform_transaction_to_master

transform transaction -> master

Description

transform transaction -> master

Usage

```
transform_transaction_to_master(dat, ROW_NUMBER = "ROW_NUMBER",
  ID = "ID", collapse = ":", STATIC_NUM = NULL, STATIC_CHAR = NULL,
  DYNAMIC_NUM = NULL, DYNAMIC_CHAR = NULL)
```

Arguments

dat	transaction data frame
ROW_NUMBER	column name for row number in the create data frame
ID	identifier name
collapse	separation character (defalt ":")
STATIC_NUM	list of column name which shows STATIC NUMBER attribute
STATIC_CHAR	list of column name which shows STATIC CHARACTER attribute
DYNAMIC_NUM	list of column name which shows DYNAMIC NUMBER attribute
DYNAMIC_CHAR	list of column name which shows DYNAMIC CHARACTER attribute

Index