Package: barn (via r-universe)

June 7, 2026

Title Preprocessing and Feature Engineering Steps before Modeling

Version 0.0.0.9000

Description The package provides pipeable functions to simplify preprocessing of tabular data prior to machine learning modeling. Users can combine multiple datasets, define feature engineering steps (such as creating new predictors from nominal or numeric columns), and then split the data back into preprocessed datasets ready to be used in machine learning workflows.

License MIT + file LICENSE

URL https://jrosell.github.io/barn

BugReports https://github.com/jrosell/barn/issues

Encoding UTF-8

Language en

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2.9000

Depends R (>= 4.3.0)

Suggests devtools, here, pak, pkgdown, purrr, rhub, roxygen2, testthat, usethis

Imports cli, crayon, dplyr, ellmer, glue, rlang, stringr, tibble (>= 3.1.0), tidyr, utils

Config/pak/sysreqs libicu-dev libssl-dev

Repository https://jrosell.r-universe.dev

Date/Publication 2025-09-10 23:00:06 UTC

RemoteUrl https://github.com/jrosell/barn

RemoteRef HEAD

RemoteSha cea7b0d85af4ea960b6be1c170a340bbf0677837

Contents

barn 2
harvest 3
plant_count_encode 3
plant_decimals_extract 4
plant_decimals_round 5
plant_label_encode 5
plant_new_nominal_pairs 6
plant_new_numeric_factors 6
plant_summarize 7

Index **8**

`barn` *Combine datasets to preprocess*

Description

Combine multiple data frames based on their common columns. That's the first step for preprocessing with the `barn` package. When printing it shows the characteristics of the combined datasets.

Usage

```
barn(..., nominal_sufix = "_cat", numeric_sufix = "_num")

## S3 method for class 'barn'
print(x, form_width = 30, ...)
```

Arguments

`...`	Extra arguments.
`nominal_sufix`	An optional string for dealing with nominal variables. Defaults to "_cat".
`numeric_sufix`	An optional string for dealing with numeric variables. Defaults to "_num".
`x`	An object of class "barn".
`form_width`	An integer specifying the minimum column width (in characters). Default is 30.

Value

A barn object containing the combined data frame, row counts,

Examples

```
full <- data.frame(id = 1:3, p1 = c("A", "B", "C"), p2 = 10:12, y = 1:3)
holdout <- data.frame(id = 4:5, p1 = c("D", "E"), p2 = 1:2)
original <- data.frame(id = 1:2, p1 = c("F", "G"), p2 = 3:4, y = 4:5)
print(barn(full, holdout, original))
```

harvest *Split the combined dataset*

Description

Splits the `combined` data frame from a `barn` object back into a named list containing the preprocessed predictors.

Usage

```
harvest(barn_obj)
```

Arguments

`barn_obj`	An object of class `"barn"`, created by `barn()`.

Value

A named list of data frames, one for each dataset originally passed to `barn()`.

Examples

```
full <- data.frame(id = 1:3, p1 = c("A", "B", "C"), p2 = 10:12, y = 1:3)
holdout <- data.frame(id = 4:5, p1 = c("D", "E"), p2 = 1:2)
original <- data.frame(id = 1:2, p1 = c("F", "G"), p2 = 3:4, y = 4:5)
harvested <- barn(full, holdout) |> harvest()
names(harvested)
harvested[["full"]]
```

plant_count_encode *Encode categorical columns with counts*

Description

Frequency encoding of nominal variables.

Usage

```
plant_count_encode(barn_obj, nominal_suffix = "_cat")
```

Arguments

`barn_obj`	A Barn object, created by `barn()`.
`nominal_suffix`	The suffix applied to column names. Defaults to "_cat".

Value

The modified `barn_obj` with the transformed combined data frame.

`plant_decimals_extract` *Extract decimals in numeric features*

Description

Creates new integer columns by extracting specific digits from numeric columns. This function emulates a feature engineering technique often used in machine learning.

Usage

```
plant_decimals_extract(barn_obj, numeric_sufix = "_num", from = 1, to = 10)
```

Arguments

`barn_obj`	A `barn` object, created by `barn()`.
`numeric_sufix`	The suffix used to identify numeric columns to process. Defaults to "_num".
`from`	The starting digit position to extract (e.g., 1 for the first decimal place). Defaults to 1.
`to`	The ending digit position to extract (e.g., 9 for the ninth decimal place). Defaults to 9.

Value

The modified `barn_obj` with the transformed combined data frame.

Examples

```
df <- tibble::tibble(x_num = c(1.234, 5.678, NA))
b <- barn(df) |> plant_decimals_extract(from = 1, to = 3)
harvest(b)[[1]]
```

plant_decimals_round *Round numeric features to specified precisions*

Description

Creates new numeric columns by rounding existing numeric columns at specified decimal precisions. This is useful for feature engineering, where different rounding granularities may capture meaningful patterns.

Usage

```
plant_decimals_round(barn_obj, numeric_sufix = "_num", precisions = c(9, 8))
```

Arguments

barn_obj	A barn object, created by barn().
numeric_sufix	The suffix used to identify numeric columns to process. Defaults to "_num".
precisions	A numeric vector specifying the number of decimal places to round to (e.g., c(9, 8)).

Value

The modified barn_obj with the transformed combined data frame.

Examples

```
df <- tibble::tibble(x_num = c(1.23456789))
b <- barn(df) |> plant_decimals_round(precisions = c(2, 3))
harvest(b)[[1]]
harvest(b)[[1]]$x_r2_num
harvest(b)[[1]]$x_r3_num
```

plant_label_encode *Encode labels in a barn object*

Description

Transform nominal columns from factors to integers.

Usage

```
plant_label_encode(barn_obj)
```

Arguments

barn_obj	An instance of class "barn".

Value

The modified `barn_obj` with the transformed combined data frame.

`plant_new_nominal_pairs`

Create new nominal pairs

Description

A function to create new features based on combinations of categorical columns in a barn object.

Usage

```
plant_new_nominal_pairs(barn_obj, nominal_suffix = "_cat")
```

Arguments

`barn_obj`	An object inheriting from the "barn" class.
`nominal_suffix`	A character string that specifies the suffix for the newly created columns. Optional, default is "_cat".

Value

The modified `barn_obj` with the transformed combined data frame.

`plant_new_numeric_factors`

New factors from numerical columns

Description

A function to transform numeric and character columns in a `barn` object into new factor columns. It appends "_num" for numeric columns, "_cat" for character columns, and renames both to factors. Original columns are deleted from the combined data frame within the barn object.

Usage

```
plant_new_numeric_factors(
  barn_obj,
  numeric_suffx = "_num",
  nominal_suffix = "_cat"
)
```

Arguments

barn_obj	A barn object, created by barn().
numeric_suffx	The suffix for new numeric factor columns. Default is "_num".
nominal_suffix	The suffix for new nominal factor columns. Default is "_cat".

Value

The modified barn_obj with the transformed combined data frame.

---

plant_summarize *Summarize Barn Object*

Description

A function to group and summarize to add aggregations to a barn_obj using specified variables and expressions. WARNING: Risk of overfitting and bad generalization if not done when resampling.

Usage

```
plant_summarize(barn_obj, .by = NULL, ...)
```

Arguments

barn_obj	An object of class 'barn'.
.by	Varible(s) to group by. Currently unused; must be empty.
...	Expressions to compute summarizing values.

Value

A modified barn_obj with summarized data in the combined slot.

Index

barn, 2
barn(), *3–5*, *7*

harvest, 3

plant_count_encode, 3
plant_decimals_extract, 4
plant_decimals_round, 5
plant_label_encode, 5
plant_new_nominal_pairs, 6
plant_new_numeric_factors, 6
plant_summarize, 7
print.barn *(*barn*)*, 2