Package ‘tidypredict’

February 10, 2020

Version 0.4.5

Title Run Predictions Inside the Database

Description It parses a fitted ‘R’ model object, and returns a formula in 'Tidy Eval' code that calculates the predictions.

It works with several databases back-ends because it leverages 'dplyr' and 'dbplyr' for the final 'SQL' translation of the algorithm. It currently supports lm(), glm(), randomForest(), ranger(), earth(), xgb.Booster.complete(), cubist(), and ctree() models.

Depends R (>= 3.1)

Imports dplyr (>= 0.7), rlang, purrr, knitr, generics, tibble

Suggests dbplyr, testthat (>= 2.1.0), randomForest, ranger, earth, markdown, nycflights13, RSQLite, methods, DBI, covr, xgboost, Cubist, mlbench, partykit, yaml, parsnip

License GPL-3

URL https://tidymodels.github.io/tidypredict

BugReports https://github.com/tidymodels/tidypredict/issues

RoxygenNote 7.0.2

Encoding UTF-8

VignetteBuilder knitr

NeedsCompilation no

Author Max Kuhn [aut, cre]

Maintainer Max Kuhn <max@r.studio.com>

Repository CRAN

Date/Publication 2020-02-10 21:20:02 UTC

R topics documented:

acceptable_formula ................................................................. 2
as_parsed_model .................................................................. 2
parse_model ........................................................................ 3
acceptable_formula

Checks that the formula can be parsed

Description

Uses an S3 method to check that a given formula can be parsed based on its class. It currently scans for contrasts that are not supported and in-line functions. (e.g: `lm(wt ~ as.factor(am)))`. Since this function is meant for function interaction, as opposed to human interaction, a successful check is silent.

Usage

acceptable_formula(model)

Arguments

model An R model object

Examples

model <- lm(mpg ~ wt, mtcars)
acceptable_formula(model)

as_parsed_model

Prepares parsed model object

Description

Prepares parsed model object

Usage

as_parsed_model(x)

Arguments

x A parsed model object
parse_model

Converts an R model object into a table

Description
It parses a fitted R model’s structure and extracts the components needed to create a dplyr formula for prediction. The function also creates a data frame using an specific format so that other functions in the future can also pass parsed tables to a given formula creating function.

Usage
parse_model(model)

Arguments
model An R model object.

Examples
library(dplyr)
df <- mutate(mtcars, cyl = paste0("cyl", cyl))
model <- lm(mpg ~ wt + cyl * disp, offset = am, data = df)
parse_model(model)

tidy.pm_regression Tidy the parsed model results

Description
Tidy the parsed model results

Usage
## S3 method for class 'pm_regression'
tidy(x, ...)

Arguments
x A parsed_model object
... Reserved for future use
tidypredict_fit | Returns a Tidy Eval formula to calculate fitted values

Description
It parses a model or uses an already parsed model to return a Tidy Eval formula that can then be used inside a dplyr command.

Usage
```
tidypredict_fit(model)
```

Arguments
- `model`: An R model or a list with a parsed model.

Examples
```
model <- lm(mpg ~ wt + cyl * disp, offset = am, data = mtcars)
tidypredict_fit(model)
```

---

tidypredict_interval | Returns a Tidy Eval formula to calculate prediction interval

Description
It parses a model or uses an already parsed model to return a Tidy Eval formula that can then be used inside a dplyr command.

Usage
```
tidypredict_interval(model, interval = 0.95)
```

Arguments
- `model`: An R model or a list with a parsed model
- `interval`: The prediction interval, defaults to 0.95

Details
The result still has to be added to the fit to obtain the upper bound, and subtracted from fit to obtain the lower bound.
Examples

```r
model <- lm(mpg ~ wt + cyl * disp, offset = am, data = mtcars)
tidypredict_interval(model)
```

Description

Compares the results of predict() and tidypredict_to_column() functions.

Usage

```r
tidypredict_test(
  model,
  df = model$model,
  threshold = 1e-12,
  include_intervals = FALSE,
  max_rows = NULL,
  xg_df = NULL
)
```

Arguments

- **model**: An R model or a list with a parsed model. It currently supports `lm()`, `glm()` and `randomForest()` models.
- **df**: A data frame that contains all of the needed fields to run the prediction. It defaults to the "model" data frame object inside the model object.
- **threshold**: The number that a given result difference, between predict() and tidypredict_to_column() should not exceed. For continuous predictions, the default value is 0.000000000001 (1e-12), for categorical predictions, the default value is 0.
- **include_intervals**: Switch to indicate if the prediction intervals should be included in the test. It defaults to `FALSE`.
- **max_rows**: The number of rows in the object passed in the df argument. Highly recommended for large data sets.
- **xg_df**: A xgb.DMatrix object, required only for XGBoost models. It defaults to `NULL` recommended for large data sets.

Examples

```r
model <- lm(mpg ~ wt + cyl * disp, offset = am, data = mtcars)
tidypredict_test(model)
```
tidypredict_to_column  

Description

Adds a new column with the results form tidypredict_fit() to a piped command set. If add_interval is set to TRUE, then it will add two additional columns, one for the lower and another for the upper prediction interval bounds.

Usage

```r
	tidypredict_to_column(
    df,
    model,
    add_interval = FALSE,
    interval = 0.95,
    vars = c("fit", "upper", "lower")
)
```

Arguments

- `df`  
  A data.frame or tibble
- `model`  
  An R model or a parsed model inside a data frame
- `add_interval`  
  Switch that indicates if the prediction interval columns should be added. Defaults to FALSE
- `interval`  
  The prediction interval, defaults to 0.95. It is ignored if add_interval is set to FALSE
- `vars`  
  The name of the variables that this function will produce. It defaults to "fit", "upper", and "lower".
Index

acceptable_formula, 2
as_parsed_model, 2

parse_model, 3

tidy.pm_regression, 3
tidypredict_fit, 4
tidypredict_interval, 4
tidypredict_test, 5
tidypredict_to_column, 6