

Package ‘finnts’

July 14, 2022

Title Microsoft Finance Time Series Forecasting Framework

Version 0.2.0

Description

Automated time series forecasting developed by Microsoft Finance. The Microsoft Finance Time Series Forecasting Framework, aka Finn, can be used to forecast any component of the income statement, balance sheet, or any other area of interest by finance. Any numerical quantity over time,

Finn can be used to forecast it. While it can be applied outside of the finance domain, Finn was built

to meet the needs of financial analysts to better forecast their businesses within a company, and has a lot of built in features that are specific to the needs of financial forecasters. Happy forecasting!

URL <https://microsoft.github.io/finnts/>,
<https://github.com/microsoft/finnts>

BugReports <https://github.com/microsoft/finnts/issues>

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

Imports cli, Cubist, dials, doParallel, dplyr, earth, foreach, generics, glmnet, gtools, hts, kernlab, lubridate, magrittr, methods, modeltime.gluons, modeltime.resample, parallel, parsnip, plyr, purrr, recipes, rsample, rules, stringr, tibble, tidyr, tidyselect, timetk, tune, workflows

Suggests knitr, reactable, rmarkdown, sparklyr, testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 3.6.0), modeltime

VignetteBuilder knitr

NeedsCompilation no

Author Mike Tokic [aut, cre] (<<https://orcid.org/0000-0002-7630-7055>>),
Aadharsh Kannan [aut] (<<https://orcid.org/0000-0002-6475-8211>>)

Maintainer Mike Tokic <mftokic@gmail.com>

Repository CRAN

Date/Publication 2022-07-14 15:40:05 UTC

R topics documented:

forecast_time_series 2

Index 7

forecast_time_series *Finn Forecast Framework*

Description

Calls the Finn forecast framework to automatically forecast any historical time series.

Usage

```
forecast_time_series(
  input_data,
  combo_variables,
  target_variable,
  date_type,
  forecast_horizon,
  external_regressors = NULL,
  run_name = "finnts_forecast",
  hist_start_date = NULL,
  hist_end_date = NULL,
  combo_cleanup_date = NULL,
  fiscal_year_start = 1,
  clean_missing_values = TRUE,
  clean_outliers = FALSE,
  back_test_scenarios = NULL,
  back_test_spacing = NULL,
  modeling_approach = "accuracy",
  forecast_approach = "bottoms_up",
  parallel_processing = NULL,
  run_model_parallel = FALSE,
  num_cores = NULL,
  target_log_transformation = FALSE,
  negative_forecast = FALSE,
  fourier_periods = NULL,
  lag_periods = NULL,
  rolling_window_periods = NULL,
  recipes_to_run = NULL,
  pca = NULL,
  reticulate_environment = NULL,
  models_to_run = NULL,
  models_not_to_run = NULL,
  run_deep_learning = FALSE,
  run_global_models = NULL,
```

```

run_local_models = TRUE,
run_ensemble_models = NULL,
average_models = TRUE,
max_model_average = 3,
weekly_to_daily = TRUE,
seed = 123
)

```

Arguments

input_data	A data frame or tibble of historical time series data. Can also include external regressors for both historical and future data.
combo_variables	List of column headers within input data to be used to separate individual time series.
target_variable	The column header formatted as a character value within input data you want to forecast.
date_type	The date granularity of the input data. Finn accepts the following as a character string day, week, month, quarter, year.
forecast_horizon	Number of periods to forecast into the future.
external_regressors	List of column headers within input data to be used as features in multivariate models.
run_name	Name used when submitting jobs to external compute like Azure Batch. Formatted as a character string.
hist_start_date	Date value of when your input_data starts. Default of NULL is to use earliest date value in input_data.
hist_end_date	Date value of when your input_data ends. Default of NULL is to use the latest date value in input_data.
combo_cleanup_date	Date value to remove individual time series that don't contain non-zero values after that specified date. Default of NULL is to not remove any time series and attempt to forecast all of them.
fiscal_year_start	Month number of start of fiscal year of input data, aids in building out date features. Formatted as a numeric value. Default of 1 assumes fiscal year starts in January.
clean_missing_values	If TRUE, cleans missing values. Only impute values for missing data within an existing series, and does not add new values onto the beginning or end, but does provide a value of 0 for said values. Turned off when running hierarchical forecasts.
clean_outliers	If TRUE, outliers are cleaned and inputted with values more in line with historical data

back_test_scenarios	Number of specific back test folds to run when determining the best model. Default of NULL will automatically choose the number of back tests to run based on historical data size, which tries to always use a minimum of 80% of the data when training a model.
back_test_spacing	Number of periods to move back for each back test scenario. Default of NULL moves back 1 period at a time for year, quarter, and month data. Moves back 4 for week and 7 for day data.
modeling_approach	How Finn should approach your data. Current default and only option is 'accuracy'. In the future this could evolve to other areas like optimizing for interpretability over accuracy.
forecast_approach	How the forecast is created. The default of 'bottoms_up' trains models for each individual time series. 'grouped_hierarchy' creates a grouped time series to forecast at while 'standard_hierarchy' creates a more traditional hierarchical time series to forecast, both based on the hts package.
parallel_processing	Default of NULL runs no parallel processing and forecasts each individual time series one after another. 'local_machine' leverages all cores on current machine Finn is running on. 'azure_batch' runs time series in parallel on a remote compute cluster in Azure Batch. 'spark' runs time series in parallel on a spark cluster in Azure Databricks/Synapse.
run_model_parallel	If TRUE, runs specific components like hyperparameter tuning or model refitting in parallel, only works when parallel_processing is not set to 'local_machine'.
num_cores	Number of cores to run when parallel processing is set up. Used when running parallel computations on local machine or within Azure. Default of NULL uses total amount of cores on machine minus one. Can't be greater than number of cores on machine minus 1.
target_log_transformation	If TRUE, log transform target variable before training models.
negative_forecast	If TRUE, allow forecasts to dip below zero.
fourier_periods	List of values to use in creating fourier series as features. Default of NULL automatically chooses these values based on the date_type.
lag_periods	List of values to use in creating lag features. Default of NULL automatically chooses these values based on date_type.
rolling_window_periods	List of values to use in creating rolling window features. Default of NULL automatically chooses these values based on date type.
recipes_to_run	List of recipes to run on multivariate models that can run different recipes. A value of NULL runs all recipes, but only runs the R1 recipe for weekly and daily date types, and also for global models to prevent memory issues. A value of "all"

	runs all recipes, regardless of date type or if it's a local/global model. A list like <code>c("R1")</code> or <code>c("R2")</code> would only run models with the R1 or R2 recipe.
<code>pca</code>	If TRUE, run principle component analysis on any lagged features to speed up model run time. Default of NULL runs PCA on day and week date types across all local multivariate models, and also for global models across all date types.
<code>reticulate_environment</code>	File path to python environment to use when training gluonts deep learning models. Only important when <code>parallel_processing</code> is not set to <code>'azure_batch'</code> or <code>'spark'</code> . Azure options should use their own docker image that has python environment already installed.
<code>models_to_run</code>	List of models to run. Default of NULL runs all models.
<code>models_not_to_run</code>	List of models not to run, overrides values in <code>models_to_run</code> . Default of NULL doesn't turn off any model.
<code>run_deep_learning</code>	If TRUE, run deep learning models from gluonts (deepar and nbeats). Overrides <code>models_to_run</code> and <code>models_not_to_run</code> .
<code>run_global_models</code>	If TRUE, run multivariate models on the entire data set (across all time series) as a global model. Can be override by <code>models_not_to_run</code> . Default of NULL runs global models for all date types except week and day.
<code>run_local_models</code>	If TRUE, run models by individual time series as local models.
<code>run_ensemble_models</code>	If TRUE, run ensemble models. Default of NULL runs ensemble models only for quarter and month date types.
<code>average_models</code>	If TRUE, create simple averages of individual models.
<code>max_model_average</code>	Max number of models to average together. Will create model averages for 2 models up until input value or max number of models ran.
<code>weekly_to_daily</code>	If TRUE, convert a week forecast down to day by evenly splitting across each day of week. Helps when aggregating up to higher temporal levels like month or quarter.
<code>seed</code>	Set seed for random number generator. Numeric value.

Value

A list of three separate data sets: the future forecast, the back test results, and the best model per time series.

Examples

```
finn_forecast <- forecast_time_series(
  input_data = m750 %>% dplyr::rename(Date = date),
  combo_variables = c("id"),
```

```
target_variable = "value",  
date_type = "month",  
forecast_horizon = 3,  
run_model_parallel = FALSE,  
models_to_run = c("arima", "ets", "snaive"))
```

Index

forecast_time_series, [2](#)