

# Package ‘tetragon’

October 14, 2022

**Type** Package

**Title** Automatic Sequence Prediction by Expansion of the Distance Matrix

**Version** 1.3.0

**Author** Giancarlo Vercellino

**Maintainer** Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

**Description** Each sequence is predicted by expanding the distance matrix. The compact set of hyper-parameters is tuned through random search.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Depends** R (>= 4.1)

**Imports** purrr (>= 0.3.4), abind (>= 1.4-5), ggplot2 (>= 3.3.5), readr (>= 2.0.1), stringr (>= 1.4.0), lubridate (>= 1.7.10), narray (>= 0.4.1.1), imputeTS (>= 3.2), scales (>= 1.1.1), tictoc (>= 1.0.1), modeest (>= 2.4.0), moments (>= 0.14), greybox (>= 1.0.1), dqrng (>= 0.3.0), entropy (>= 1.3.1), Rfast (>= 2.0.6), philentropy (>= 0.5.0), fastDummies (>= 1.6.3), fANCOVA (>= 0.6-1)

**URL** [https://rpubs.com/giancarlo\\_vercellino/tetragon](https://rpubs.com/giancarlo_vercellino/tetragon)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-08-13 17:30:02 UTC

## R topics documented:

covid_in_europe	2
tetragon	2

<b>Index</b>	<b>5</b>
--------------	----------

---

covid_in_europe	<i>covid_in_europe data set</i>
-----------------	---------------------------------

---

**Description**

A data frame with with daily and cumulative cases of Covid infections and deaths in Europe since March 2021.

**Usage**

```
covid_in_europe
```

**Format**

A data frame with 5 columns and 163 rows.

**Source**

[www.ecdc.europa.eu](http://www.ecdc.europa.eu)

---

tetragon	<i>tetragon</i>
----------	-----------------

---

**Description**

Each sequence is predicted by expanding the distance matrix. The compact set of hyper-parameters is tuned via grid or random search.

**Usage**

```
tetragon(  
  df,  
  seq_len = NULL,  
  smoother = F,  
  ci = 0.8,  
  method = NULL,  
  distr = NULL,  
  n_windows = 3,  
  n_sample = 30,  
  dates = NULL,  
  error_scale = "naive",  
  error_benchmark = "naive",  
  seed = 42  
)
```

**Arguments**

<code>df</code>	A data frame with time features as columns. They could be continuous variables or not.
<code>seq_len</code>	Positive integer. Time-step number of the projected sequence. Default: NULL (random selection between maximum boundaries).
<code>smoother</code>	Logical. Perform optimal smoothing using standard loess. Default: FALSE
<code>ci</code>	Confidence interval. Default: 0.8.
<code>method</code>	String. Distance method for calculating distance matrix among sequences. Options are: "euclidean", "manhattan", "maximum", "minkowski". Default: NULL (random selection among all possible options).
<code>distr</code>	String. Distribution used to expand the distance matrix. Options are: "norm", "logis", "t", "exp", "chisq". Default: NULL (random selection among all possible options).
<code>n_windows</code>	Positive integer. Number of validation tests to measure/sample error. Default: 3 (but a larger value is strongly suggested to really understand your accuracy).
<code>n_sample</code>	Positive integer. Number of samples for random search. Default: 30.
<code>dates</code>	Date. Vector with dates for time features.
<code>error_scale</code>	String. Scale for the scaled error metrics (only for continuous variables). Two options: "naive" (average of naive one-step absolute error for the historical series) or "deviation" (standard error of the historical series). Default: "naive".
<code>error_benchmark</code>	String. Benchmark for the relative error metrics (only for continuous variables). Two options: "naive" (sequential extension of last value) or "average" (mean value of true sequence). Default: "naive".
<code>seed</code>	Positive integer. Random seed. Default: 42.

**Value**

This function returns a list including:

- `exploration`: list of all explored models, complete with predictions, testing metrics and plots
- `history`: a table with the sampled models, hyper-parameters, validation errors
- `best`: results for the best model including:
  - `predictions`: min, max, q25, q50, q75, quantiles at selected `ci`, and a bunch of specific measures for each point fo predicted sequences
  - `testing_errors`: testing errors for one-step and sequence for each `ts` feature
  - `plots`: confidence interval plot for each time feature
- `time_log`

**Author(s)**

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

**See Also**

Useful links:

- [https://rpubs.com/giancarlo\\_vercellino/tetragon](https://rpubs.com/giancarlo_vercellino/tetragon)

**Examples**

```
tetragon(covid_in_europe[, c(2, 4)], seq_len = 40, n_sample = 2)
```

# Index

## \* datasets

covid\_in\_europe, [2](#)

covid\_in\_europe, [2](#)

tetragon, [2](#)

tetragon-package (tetragon), [2](#)