Package 'rethnicity'

July 23, 2025

Type Package

Title Predicting Ethnic Group from Names

Version 0.2.7

Maintainer Fangzhou Xie <fangzhou.xie@rutgers.edu>

Description Implementation of the race/ethnicity prediction method, described in ``rethnicity: An R package for predicting ethnicity from names'' by Fangzhou Xie (2022) <doi:10.1016/j.softx.2021.100965> and ``Rethnicity: Predicting Ethnicity from Names'' by Fangzhou Xie (2021) <doi:10.48550/arXiv.2109.09228>.

License CC BY-NC-SA 4.0

Encoding UTF-8

RoxygenNote 7.3.2

URL https://github.com/fangzhou-xie/rethnicity

BugReports https://github.com/fangzhou-xie/rethnicity/issues

Depends R (>= 3.4.0)

LinkingTo Rcpp, RcppEigen, RcppThread (>= 2.1.3)

Imports Rcpp, cli, rlang

Suggests pak, knitr, rmarkdown, testthat (>= 3.0.0), magrittr, parallel

VignetteBuilder knitr

Language en-US

Config/testthat/edition 3

UseLTO TRUE

NeedsCompilation yes

Author Fangzhou Xie [aut, cre, cph] (ORCID: <https://orcid.org/0000-0001-7702-093X>)

Repository CRAN

Date/Publication 2025-03-12 20:10:02 UTC

Contents

`predict_ethnicity` *Predict ethnicity from names.*

Description

Predict ethnicity either by last names or both first and last names. This is the default and recommended method for prediction.

Usage

```
predict_ethnicity(
  firstnames = NULL,
  lastnames = NULL,
  method = "fullname",
  threads = 0,
  na.rm = FALSE
)
```

Arguments

`firstnames`	A character vector of first names. Default to NULL. Only use this if you are using 'method' = 'fullname'.
`lastnames`	A character vector of last names. Default to NULL. Use this in both 'fullname' and 'lastname' methods.
`method`	"fullname" or "lastname". Inference method to choose from.
`threads`	single integer. Number of threads to use for multi-threading.
`na.rm`	TRUE or FALSE (bool). If TRUE, then the NAs will be removed; if FALSE, then return error if there is NA in the arguments.

Value

data.frame with probability of being each ethnic group and the predicted group (one with highest probability)

Examples

```
predict_ethnicity(firstnames = "Alan", lastnames = "Turing")
```

predict_fullname *Predict ethnicity from full name*

Description

Predicts ethnicity from first names and last names, using self-trained model with customized labels. This is designed for advanced users who wish to use their own models. For most use cases, use `predict_ethnicity()` for prediction.

Usage

```
predict_fullname(
  firstnames,
  lastnames,
  na.rm = FALSE,
  threads = 0L,
  labels = NULL,
  model_path = NULL
)
```

Arguments

`firstnames`	character vector, first names
`lastnames`	character vector, last names
`na.rm`	bool, default to FALSE, whether to remove the na in the `lastnames`
`threads`	int, number of threads for multi-threading
`labels`	character vector, labels of the classification model, needs to be in the same order as the trained model
`model_path`	character file path, the path to the trained model in .json format (converted from Keras by frugally-deep)

Value

data.frame with predicted probability and predicted ethnicity

predict_lastname *Predict ethnicity from last name*

Description

Predicts ethnicity from last names, using self-trained model with customized labels. This is designed for advanced users who wish to use their own models. For most use cases, use `predict_ethnicity()` for prediction.

Usage

```
predict_lastname(
  lastnames,
  na.rm = FALSE,
  threads = 0L,
  labels = NULL,
  model_path = NULL
)
```

Arguments

lastnames	character vector, last names
na.rm	bool, default to FALSE, whether to remove the na in the lastnames
threads	int, number of threads for multi-threading
labels	character vector, labels of the classification model, needs to be in the same order as the trained model
model_path	character file path, the path to the trained model in .json format (converted from Keras by frugally-deep)

Value

data.frame with predicted probability and predicted ethnicity

Index