Package 'ClassificationEnsembles'

February 26, 2026

Title Automatically Builds 12 Classification Models (6 Individual and 6 Ensembles of Models) from Classification Data

Version 1.0.2

Description Automatically builds 12 classification models from data. The package also returns 25 plots, 5 tables and a summary report.

License MIT + file LICENSE

Depends C50, car, caret, corrplot, doParallel, dplyr, e1071, ggplot2, gt, htmltools, htmlwidgets, ipred, MachineShop, magrittr, parallel, pls, purrr, R (>= 3.5), randomForest, ranger, reactable, scales, tidyr, tree

Encoding UTF-8

RoxygenNote 7.3.3

LazyData true

Suggests knitr, rmarkdown

VignetteBuilder knitr

URL https://github.com/InfiniteCuriosity/ClassificationEnsembles

BugReports https://github.com/InfiniteCuriosity/ClassificationEnsembles/issues

NeedsCompilation no

Author Russ Conte [aut, cre, cph]

Maintainer Russ Conte <russconte@mac.com>

Repository CRAN

Date/Publication 2026-02-26 21:00:02 UTC

Contents

Carseats   *Carseats data*

Description

This is the Carseats data as shown in the ISLR package.

Usage

Carseats

Format

Carseats A simulated data set with 400 observations and 11 rows

Sales Unit sales (in thousands) at each location

CompPrice Price charged by competitor at each location

Income Community income level (in thousands of dollars)

Advertising Local advertising budget for company at each location (in thousands of dollars)

Population Population size in region (in thousands)

Price Price company charges for car seats at each site

ShelveLoc A factor with levels Bad, Good and Medium indicating the quality of the shelving location for the car seats at each site

Age Average age of the local population

Urban A factor with levels No and Yes to indicate whether the store is in an urban or rural location

US A factor with levels No and Yes to indicate whether the store is in the US or not

Source

ISLR data set, https://www.rdocumentation.org/packages/ISLR/versions/1.4/topics/Carseats

---

Classification   *classification—function to perform classification analysis and return results to the user.*

Description

classification—function to perform classification analysis and return results to the user.

Usage

```
Classification(
  data,
  colnum,
  numresamples,
  predict_on_new_data = c("Y", "N"),
  remove_VIF_above,
  scale_all_numeric_predictors_in_data,
  how_to_handle_strings = c(0("No strings"), 1("Strings as factors")),
  set_seed = c("Y", "N"),
  save_all_trained_models = c("Y", "N"),
  save_all_plots,
  stratified_random_column,
  use_parallel = c("Y", "N"),
  train_amount,
  test_amount,
  validation_amount
)
```

Arguments

`data`	a data set that includes classification data. For example, the Carseats data in the ISLR package
`colnum`	the number of the column. For example, in the Carseats data this is column 7, ShelveLoc with three values, Good, Medium and Bad
`numresamples`	the number of times to resample the analysis
`predict_on_new_data`	asks if the user has new data to be analyzed using the trained models that were just developed
`remove_VIF_above`	Removes columns with Variance Inflation Factors above the level chosen by the user
`scale_all_numeric_predictors_in_data`	Scales all numeric predictors in the original data
`how_to_handle_strings`	Option to convert strings to factor levels
`set_seed`	Option to set a seed
`save_all_trained_models`	Gives the user the option to save all trained models in the Environment
`save_all_plots`	Saves all plots in the user's chosen format
`stratified_random_column`	0 if no stratified random sampling, or column number for stratified random sampling
`use_parallel`	"Y" or "N" for parallel processing
`train_amount`	set the amount for the training data

`test_amount`   set the amount for the testing data

`validation_amount`

        Set the amount for the validation data

Value

a full analysis, including data visualizations, statistical summaries, and a full report on the results of 35 models on the data

---

`Cleveland_heart`   *Cleveland Heart*

Description

Posted by John Gennari, 3/13/90, This is Dr. Detrano's database modified to be a real MIXED dataset.

Usage

`Cleveland_heart`

Format

Cleveland_heart These are the original attributes: Attributes: 8 symbolic, 6 numeric. Age; sex; chest pain type (angina, abnang, notang, asympt) Trestbps (resting blood pres); cholesteral; fasting blood sugar < 120 (true or false); resting ecg (norm, abn, hyper); max heart rate; exercise induced angina (true or false); oldpeak; slope (up, flat, down) number of vessels colored (???); thal (norm, fixed, rever). Finally, the class is either healthy (buff) or with heart-disease (sick).

Source

https://archive-beta.ics.uci.edu/dataset/45/heart+disease/files?path=cleve.mod

The column names were corrected to be usable by R ('Chest_pain_type' instead of 'chest pain type'), removed the '<' symbol from a column name because the '<' symbol causes some errors reading the column names (such as tree models), removed three columns due to very high number of missing cells (noted as '?' in the original file.) Those three columns are 'Number_of_vessels_colored', 'Thal', and 'Resting_ECG'.

Age Age of the subject

Sex Sex of the subject, either male or female

Chest_pain_type One of angina, abnang, notang, or asympt

Resting_blood_pressure The resting blood pressure for the subject

Cholesteral The patient's cholesterol

Fasting_blood_sugar Binary, whether the fasting blood sugar is <120

Max_heart_rate The maximum measured heart rate for the patient

Exercise_induced_angina Binary, whether angina was induced due to exercise

Old_peak Numeric value

Slope Three levels: Down, flat, up

Sick_or_buff Binary, is the patient sick or buff

Class 0 is healthy, 1,2,3,4 is sick.

---

dry_beans_small *Dry Beans small*

Description

This is a stratified version of the full dry beans data set. This is about 7 percent of the full data set

Usage

```
dry_beans_small
```

Format

dry_beans_small A reduced version with 813 rows and 17 columns of the full data set available on UCI: https://archive.ics.uci.edu/dataset/602/dry+bean+dataset

Area The area of a bean zone and the number of pixels within its boundaries

Perimeter Bean circumference is defined as the length of its border

MajorAxisLength The distance between the ends of the longest line that can be drawn from a bean

MinorAxisLength The longest line that can be drawn from the bean while standing perpendicular to the main axis

AspectRatio Defines the relationship between MajorAxisLength and MinorAxisLength

Eccentricity Eccentricity of the ellipse having the same moments as the region

ConvexArea Number of pixels in the smallest convex polygon that can contain the area of a bean seed

EquivDiameter Equivalent diameter: The diameter of a circle having the same area as a bean seed area

Extent The ratio of the pixels in the bounding box to the bean area

Solidity Also known as convexity. The ratio of the pixels in the convex shell to those found in beans.

Roundness Calculated with the following formula: (4piA)/(P^2)

Compactness Measures the roundness of an object

ShapeFactor1 Continuous value

ShapeFactor2 Continuous value

ShapeFactor3 Continuous value

ShapeFactor4 Continuous value

Class (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira)

@source https://archive.ics.uci.edu/dataset/602/dry+bean+dataset

---

Maternal_Health_Risk   *Maternal Health Risk*

Description

Data has been collected from different hospitals, community clinics, maternal health cares from the rural areas of Bangladesh through the IoT based risk monitoring system.

Usage

```
Maternal_Health_Risk
```

Format

Maternal_Health_Risk Age, Systolic Blood Pressure as SystolicBP, Diastolic BP as DiastolicBP, Blood Sugar as BS, Body Temperature as BodyTemp, HeartRate and RiskLevel. All these are the responsible and significant risk factors for maternal mortality, that is one of the main concern of SDG of UN.

Age Any ages in years when a women during pregnant.

SystolicBP Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.

DiastolicBP Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.

BS Blood glucose levels is in terms of a molar concentration

BodyTemp Body temperature in Farenheit

HeartRate A normal resting heart rate

RiskLevel Predicted Risk Intensity Level during pregnancy considering the previous attribute.

Index