

Package ‘ClassificationEnsembles’

April 1, 2025

Title Automatically Builds 20 Classification Models

Version 0.5.0

Description Automatically builds 20 classification models from data. The package returns 26 plots, 5 tables and a summary report.

The package automatically builds 12 individual classification models, including error (RMSE) and predictions. That data is used to create an ensemble, which is then modeled using 8 methods.

The process is repeated as many times as the user requests. The mean of the results are presented in a summary table.

The package returns the confusion matrices for all 20 models, tables of the correlation of the numeric data, the results of the variance inflation process, the head of the ensemble and the head of the data frame.

License MIT + file LICENSE

Depends C50, car, caret, corrplot, doParallel, dplyr, e1071, ggplot2, gt, ipred, MachineShop, magrittr, parallel, pls, purrr, R (>= 2.10), randomForest, ranger, reactable, reactablefmtr, scales, tidy, tree

Encoding UTF-8

RoxygenNote 7.3.2

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 2025-04-01 16:10:05 UTC

Contents

Carseats	2
Classification	3
dry_beans_small	4
Maternal_Health_Risk	5
Index	6

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	<i>classification—function to perform classification analysis and return results to the user.</i>
----------------	---

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")),
  save_all_trained_models = c("Y", "N"),
  save_all_plots,
  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	Gives the user the opportunity to use the trained models to predict on new and untrained data
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	Converts strings to factor levels
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

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

dry_beans_small	<i>Dry Beans small</i>
-----------------	------------------------

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: $(4\pi A)/(P^2)$

Compactness Measures the roundness of an object

ShapeFactor1 Continuous value

ShapeFactor2 Continuous value

ShapeFactor3 Continuous value

ShapeFactor4 Continuous value

Class (Seker, Barbunya, Bombay, Cali, Derosan, 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

* datasets

Carseats, [2](#)

dry_beans_small, [4](#)

Maternal_Health_Risk, [5](#)

Carseats, [2](#)

Classification, [3](#)

dry_beans_small, [4](#)

Maternal_Health_Risk, [5](#)