Package 'MSclassifier'

June 23, 2020

Type Package

Title Median-Supplement model based Classification tool for automated knowledge discovery

Version 1.0.0

Maintainer Emmanuel S. Adabor <emmanuelsadabor@gimpa.edu.gh>

Description Median-Supplement model based Classification implements a median-supplement approach to machine learning, supporting complete compliance efforts by never missing sensitive sub-datasets or allowing some sub-datasets to escape the classification process when balancing overall dataset as required in traditional classification models, for an automated and effective binary classification for optimal decision making.

Depends R (>= 3.3.3)

Imports e1071, randomForest (>= 4.6-12)

ByteCompile true

License GNU General Public Licence (Refer to <http://www.gnu.org/licenses/>)

Encoding UTF-8

LazyData True

RoxygenNote 6.0.1

URL https://nweb.gimpa.edu.gh/schools/school-of-technology/software/MSclassifier/

BugReports Email to Emmanuel S. Adabor <emmanuelsadabor@gimpa.edu.gh>

NeedsCompilation yes

Author Emmanuel S. Adabor
George K. Acquaah-Mensah
Gaston K. Mazandu

R topics documented:

MSclassifier-package Median-Supplement model based Classification

Description

 Implementation of the median-supplement model based classification tool

Details

 Median-Supplement model based Classification implements a median-supplement approach to machine learning, supporting complete compliance efforts by never missing sensitive sub-datasets or allowing some sub-datasets to escape the classification process when balancing overall dataset as required in traditional classification models, for an automated and effective binary classification for optimal decision making.

Package : MSclassifier
Version : 1.0.0
Date : 23 June 2020
Depends: R (>= 3.3.3)
Imports : e1071, randomForest (>= 4.6-12)
License : GNU General Public Licence (GPL >= 3), Refer to <http://www.gnu.org/licenses/>)

 Installation of the MSclassifier-package

Procedures

 Installation from the terminal – Note that the directory to the package is designated.
 Unzip the package and use the following command for installation:
 R CMD INSTALL 'path_to/Rpackage/MSclassifier'

 Alternative installation in R interphase
 library(devtools)
 install("path_to/Rpackage/MSclassifier")

 MSclassifier method

Description

 Given a data frame of values and set of new instances to be classified (see arguments below), this function classifies these new instances.

Usage

 MSclassifier(X, testset=NULL, method="MSRandomForest")

Arguments (Input parameters)

X   A data frame of values of attributes (e.g. gene expression levels) and classes (e.g. receptor status phenotypes in breast cancer). Samples are in rows while attributes are in columns. The last column of X should have the classes for all instances in X (e.g. receptor status phenotypes of samples). This form the training set.

testset   This is the set of new instances to be classified. The Default is NULL. When set to NULL, the function returns only the model. To classify new instances, specify the data frame of the new instances as the test set. It should have the form (and attributes) of X.

method   It specifies whether to determine a median-supplement Random Forest or median-supplement Naive Bayes. "MSRandomForest" infers median-supplement Random Forest. "MSNaiveBayes" applies the median-supplement Naive Bayes. The default is median-supplement Random Forest.

Value (Output)
Classification decision

Illustration (Application)
Comprehensive description of function and application can be found in the help file after loading the package and getting the full description of the package:
> library(MSclassifier)
> ?MSclassifier

There are sample data sets that are used to illustrate the package in this vignette. These are included in the package that is loaded.
> data(her2)
> data(testset)
> her2[1:3,1:3]

```
    NPTXR_23467 DOCK3_1795 LOC400927_400927
1   266.0075    38.1356    12.7119
2   461.8575    34.9231    6.5481
3   199.3335    11.8146    6.9676
```

> testset[1:3,1:3]

```
    NPTXR_23467 DOCK3_1795 LOC400927_400927
1   304.5058    21.0756    0.7267
2   453.4778    10.1620    7.5072
3   510.8080     3.5776    4.9292
```

To classify instances using median-supplement Random Forest the following apply:
> Predictions <- MSclassifier(her2,testset = testset, method ="MSRandomForest")
> head(Predictions)
Sample1 Sample2 Sample3 Sample4 Sample5 Sample6
Negative Negative Positive Negative Negative Negative

Levels: Negative Positive
To analyse median-supplement Random Forest for error matrix, here is a sample:
```
> Model <- MSclassifier(her2, testset = NULL, method = "MSRandomForest")
> predictions <- predict(Model, newdata = testset)
> head(predictions)
1    2    3    4    5    6
Negative Negative Positive Negative Negative Negative
Levels: Negative Positive

> table(predictions, testset$her2_status)
 predictions Negative Positive
 Negative     47       10
 Positive      4        1
```

To classify instances using median-supplement Naive Bayes, here is a sample:
```
> Predictions <- MSclassifier(her2,testset = testset, method = "MSNaiveBayes")
> head(Predictions)
Sample1  Sample2  Sample3  Sample4  Sample5  Sample6
Negative Negative Positive Negative Negative Negative
Levels: Negative Positive
```

To analyse median-supplement Naive Bayes for error matrix, the following is an example:
```
> Model <- MSclassifier(her2, testset = NULL, method = "MSNaiveBayes")
> predictions <- predict(Model, newdata = testset[,-ncol(testset)])
> head(predictions)
[1] Negative Negative Positive Negative Negative Negative
Levels: Negative Positive

> table(predictions, testset$her2_status)
predictions Negative Positive
Negative     51       8
Positive      0       3
```

Package test datasets

Description
The package contains a test dataset with training and test samples as described below:

her2   This is a sample training data consisting of 86 HER2 receptor-negative and 14 HER2 receptor-positive instances.

testset   This is a sample test data for testing models developed from the her2 training data. It consists of 51 HER2 receptor-negative and 11 HER2 receptor-positive instances.