Detailed Machine Learning Workflow Report

thiloshonnagarajah

Mon Aug 12 00:10:51 2019

ML Plan

Type: classification

Target Variable: Species

Meta of Data

Number of classes: 3

Size of majority class: 50

Size of minority class: 50

Number of features: 5

Number of numeric features: 4

Number of symbolic features: 0

Number of records: 150

Number of records with missing values: 150

Number of total missing values: 0

Data Highlight:

```
## # A tibble: 150 x 1
##    `dataStore$mlPlan$data~ $Sepal.Width $Petal.Length $Petal.Width $Species
##                      <dbl>        <dbl>         <dbl>        <dbl> <chr>
##  1                     5.1          3.5           1.4          0.2 setosa
##  2                     4.9          3             1.4          0.2 setosa
##  3                     4.7          3.2           1.3          0.2 setosa
##  4                     4.6          3.1           1.5          0.2 setosa
##  5                     5            3.6           1.4          0.2 setosa
##  6                     5.4          3.9           1.7          0.4 setosa
##  7                     4.6          3.4           1.4          0.3 setosa
##  8                     5            3.4           1.5          0.2 setosa
##  9                     4.4          2.9           1.4          0.2 setosa
## 10                     4.9          3.1           1.5          0.1 setosa
## # ... with 140 more rows
```

ML Pipes

```
######################## xgboost ########################

Machine Learning Pipeline Object

ID: xgboost
```

```
Learning Algorithm: [1] "classif.xgboost"
Preprocessing List:             id           label
1    factorPre   Factorization
2   outlierPre Outlier Removal
3 normalizePre  Normalization
                                           applied_on pre_split
1                                             Species      TRUE
2                             Sepal.Length, Sepal.Width      TRUE
3 Sepal.Length, Sepal.Width, Petal.Length, Petal.Width      TRUE
Train, Test, Cross Validation Split: [[1]]
Resample description: holdout with 0.90 split rate.
Predict: test
Stratification: FALSE

#####  MLR Task: ##### [[1]]
Supervised task: xgboost
Type: classif
Target: Species
Observations: 150
Features:
   numerics     factors     ordered functionals
          4           0           0           0
Missings: FALSE
Has weights: FALSE
Has blocking: FALSE
Has coordinates: FALSE
Classes: 3
    setosa versicolor  virginica
        50         50         50
Positive class: NA

#####  MLR Learner: ##### [[1]]
Learner classif.xgboost from package xgboost
Type: classif
Name: eXtreme Gradient Boosting; Short name: xgboost
Class: classif.xgboost
Properties: twoclass,multiclass,numerics,prob,weights,missings,featimp
Predict-Type: response
Hyperparameters: nrounds=1,verbose=0


#####  MLR Model: ##### [[1]]
Resample Result
Task: xgboost
Learner: classif.xgboost
Aggr perf: ber.test.mean=0.0000000,acc.test.mean=1.0000000,timetrain.test.mean=0.0010000
Runtime: 0.00690317
```

References

Thiloshon Nagarajah and Guhanathan Poravi (2019). automlr: Automated Machine Learning in R. R package version 0.0.009.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical

Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/