

# RMiner : Data Mining Classification et Méthode de Régression

Zakaria RIDADARAJAT

02/02/2021

## Entête :

Nous allons explorer dans ce document le package Rminer faite par ARNAUD-BRUEL-YANKO

Voici le lien vers son GitHub est <https://github.com/ARNAUD-BRUEL-YANKO/PSBX/blob/main/Package%20RMiner.pdf>

## Synthèse :

Le package Rminer facilite l'utilisation des méthodes du dataming qui œuvre sur les algorithmes d'apprentissage tel que la classification, la régression et même l'analyse des séries temporelles.

## Extrait commenté :

```
library(rminer)
```

```
## Warning: package 'rminer' was built under R version 3.6.2
```

```
t=1:20
```

```
d=CasesSeries(1:10,c(1,3,4))
```

```
print(d)
```

```
##   lag4 lag3 lag1 y
## 1    1    2    4 5
## 2    2    3    5 6
## 3    3    4    6 7
## 4    4    5    7 8
## 5    5    6    8 9
## 6    6    7    9 10
```

```
d=CasesSeries(1:10,c(1,2,3))
```

```
print(d)
```

```
##   lag3 lag2 lag1 y
## 1    1    2    3 4
## 2    2    3    4 5
## 3    3    4    5 6
## 4    4    5    6 7
## 5    5    6    7 8
## 6    6    7    8 9
## 7    7    8    9 10
```

La fonction `CasesSeries()` appartient à l'étape data préparation de la donnée temporelles, elle crée un data frame d'une série temporelle en utilisant des retards (lags).

Pour la phase de modélisation, l'auteur a utilisé deux fonctions : `fit`, `predict`.

-Fit : la fonction ajuste un modèle sélectionné de jeux de données et peut automatiquement ajuster les hyperparamètres. Les hyperparamètres sont des paramètres réglables qui vous permettent de contrôler le processus d'entraînement du modèle.

-Predict : la fonction donne un modèle ajusté et calcule les prédictions pour un jeu de données.

Les fonctions `fit` et `predict` proposent plusieurs modèles tels que :

- `rpart` (arbre de décision)
- `randomForest`
- `lm` (régression linéaire ou multiple)
- `cv.glmnet` : modèle linéaire généralisé

## Evaluation :

1) Exécution du fichier `rmd`:

Le code R est reproductible.

2) Qualité de rédaction :

Le travail est bien fait et bien présenté.

3) Aspect didactique

Manque d'explication au niveau des choix des paramètres mais globalement simple à comprendre.

4) Lisibilité du `rmarkdown`:

Le code est bien organisé.

5) Bibliographie :

Non, il y'a pas bibliographie.

## Conclusion :

À mon sens, ce travail est intéressant. L'auteur fournit un dossier documenté et facile à lire, malgré le manque d'explication sur les fonctions et leurs utilités.