

# Le travail de Jean SOURIS est disponible ici

Théo Marié

20/12/2020

## SYNTHESE DU TRAVAIL EN QUESTION

Le package Caret est important car il nous introduit à l'analyse prédictive, très intéressante en programmation. La notion de seed doit être expliquée, ainsi que l'essence même de ce qu'est l'analyse prédictive, et quelques étapes importantes, comme l'entraînement et le test. Malheureusement, le travail ici n'est pas complet. Il s'arrête seulement à la partie de l'entraînement, ou il entraîne ses données, mais il effectue cette manipulation 2 fois. Aussi, on ne comprend pas ce qu'est le seed : ce n'est pas un chiffre aléatoire généré, c'est la référence de départ que je vais donner à l'algorithme pour générer de l'aléa. Il ne contient pas non plus de fonctionnalité externe qui pourraient habiller l'algorithme, pour le rendre plus efficace.

## EXTRAIT COMMENTE DES PARTIES DU CODE

```
library(lattice)
library(ggplot2)
library(caret)
set.seed(2020)
Essaie_1 <- createDataPartition(iris$Species, p = .6,
                                list = FALSE,
                                times = 3)
head(Essaie_1)
```

```
##      Resample1 Resample2 Resample3
## [1,]         1         2         1
## [2,]         2         4         2
## [3,]         3         7         4
## [4,]         4         8         7
## [5,]         6         9         8
## [6,]         7        13         9
```

Il crée ici un tableau comportant les données d'Iris, grâce à la fonction "createDataPartition". Ensuite, il la paramètre, en disant qu'il veut évaluer 60% des données, et ceci dans 3 colonnes, en ajoutant list = FALSE, ce qui évite que les données soient sous forme de liste.

```
Essaie_2 <- createDataPartition(iris$Species, p = .6,
                                list = TRUE,
                                times = 2)
head(Essaie_2)
```

```
## $Resample1
## [1]  3  4  6  7 10 11 12 14 15 17 18 19 20 25 26 28 30 31 32
## [20] 33 36 37 38 42 43 44 45 47 48 50 52 55 57 58 60 64 65 67
```

```
## [39] 69 71 72 73 74 76 77 78 79 80 81 82 84 86 89 90 92 93 96
## [58] 97 98 99 101 105 106 108 110 112 113 115 116 118 119 120 121 122 123 124
## [77] 125 128 129 130 132 134 135 137 138 141 144 147 148 150
##
## $Resample2
## [1] 1 3 4 6 7 9 10 11 12 14 15 17 19 20 21 24 26 27 28
## [20] 29 30 33 36 37 38 42 43 44 45 47 52 53 56 57 58 60 62 65
## [39] 69 70 72 76 77 80 81 82 83 85 87 89 90 91 92 93 94 95 96
## [58] 98 99 100 102 103 104 106 107 108 109 110 111 113 115 116 118 119 122 123
## [77] 125 126 127 128 129 134 136 138 139 141 143 145 146 150
```

Même chose que plus aussi, sauf pour 2 colonnes.

## EVALUATION DU TRAVAIL EN QUESTION

Critère 1 : Visuellement appréciable sur pdf 1/4 Peu d'efforts

Critère 2 : idées pour faire le code 1/4 Peu recherché, pas complet

Critère 3 : Fonctionnalité du code 2/4 le code marche mais est incomplet

Critère 4 : lisibilité du code 3/4 Le code est bien lisible

Critère 5 : explications données 1/4 Des informations parfois fausses et insuffisantes.

## CONCLUSION

Le travail est basique, et même si il marche, il ne dévoile en rien les particularités du package Caret, c'est dommage. On aurait préféré des résultats, et des fonctionnalités autres, comme le temps que l'algorithme prend pour fonctionner, ou un graphique.