

# Use-case-Apprentissage-Supervise - DT

Serigne Fallou MBacke NGOM

2023-12-07

## DONNEES DE TEST ET D'APPRENTISSAGE

- Chargement des donnees

```
achat <- read.csv("apprentissage-supervise/partie3-use-case-apprentissage-supervise-data-achat.csv", header = TRUE, sep = ",", dec = ".", stringsAsFactors=T)
str(achat)
```

```
## 'data.frame': 600 obs. of 12 variables:
## $ ID      : int 12101 12102 12103 12104 12105 12106 12107 12108 12109 12110 ...
## $ Age     : int 48 40 51 23 57 57 22 58 37 54 ...
## $ Sexe    : Factor w/ 2 levels "F","H": 1 2 1 1 1 1 2 2 1 2 ...
## $ Habitat : Factor w/ 4 levels "Banlieue","Centre_Ville",...: 2 3 2 3 4 3 4 3 1 3 ...
## $ Revenus : int 140368 249360 132600 163000 414024 302952 71016 199568 202432 193696 ...
## $ Marie   : Factor w/ 2 levels "Non","Oui": 1 2 2 2 2 1 2 2 2 ...
## $ Enfants : num 1 3 0 3 0 2 0 0 2 2 ...
## $ Voiture : Factor w/ 2 levels "Non","Oui": 1 2 2 1 1 1 1 2 2 2 ...
## $ Compte_Epargne: Factor w/ 2 levels "Non","Oui": 1 1 2 1 2 2 1 2 1 2 ...
## $ Compte_Courant: Factor w/ 2 levels "Non","Oui": 1 2 2 2 1 2 2 2 1 2 ...
## $ Emprunt  : Factor w/ 2 levels "Non","Oui": 1 2 1 1 1 1 1 1 1 1 ...
## $ Achat    : Factor w/ 2 levels "Non","Oui": 2 1 1 1 1 2 2 1 1 1 ...
```

```
head(achat)
```

```
##   ID Age Sexe  Habitat Revenus Marie Enfants Voiture Compte_Epargne
## 1 12101 48  F Centre_Ville 140368 Non    1   Non      Non
## 2 12102 40  H Petite_Ville 249360 Oui    3   Oui      Non
## 3 12103 51  F Centre_Ville 132600 Oui    0   Oui      Oui
## 4 12104 23  F Petite_Ville 163000 Oui    3   Non      Non
## 5 12105 57  F Rural 414024  Oui    0   Non      Oui
## 6 12106 57  F Petite_Ville 302952 Oui    2   Non      Oui
##  Compte_Courant Emprunt Achat
## 1      Non   Non  Oui
## 2      Oui   Oui  Non
## 3      Oui   Non  Non
## 4      Oui   Non  Non
## 5      Non   Non  Non
## 6      Oui   Non  Oui
```

```
table(achat$Achat)
```

```
##
## Non Oui
## 326 274
```

- Construction des ensembles d'apprentissage et de test

```
nrow(achat)
```

```
## [1] 600
```

```
achat_EA <- achat[1:400,]
achat_ET <- achat[401:600,]
nrow(achat_EA) ; nrow(achat_ET)
```

```
## [1] 400
```

```
## [1] 200
```

- Suppression variable ID

```
achat_EA <- subset(achat_EA, select = -ID)
```

- Affichages

head(achat\_EA)

```
## Age Sexe   Habitat Revenus Marie Enfants Voiture Compte_Epargne
## 1 48  F Centre_Ville 140368 Non    1    Non      Non
## 2 40  H Petite_Ville 249360 Oui    3    Oui      Non
## 3 51  F Centre_Ville 132600 Oui    0    Oui      Oui
## 4 23  F Petite_Ville 163000 Oui    3    Non      Non
## 5 57  F Rural 414024 Oui    0    Non      Oui
## 6 57  F Petite_Ville 302952 Oui    2    Non      Oui
##  Compte_Courant Emprunt Achat
## 1      Non    Non  Oui
## 2      Oui    Oui  Non
## 3      Oui    Non  Non
## 4      Oui    Non  Non
## 5      Non    Non  Non
## 6      Oui    Non  Oui
```

head(achat\_ET)

```
## ID Age Sexe   Habitat Revenus Marie Enfants Voiture Compte_Epargne
## 401 12501 34  F Rural 155848 Non    0    Non      Non
## 402 12502 26  H Rural 156504 Non    3    Oui      Non
## 403 12503 48  H Centre_Ville 308784 Oui    0    Oui      Non
## 404 12504 35  H Centre_Ville 166032 Non    0    Non      Non
## 405 12505 24  F Centre_Ville 110912 Oui    3    Non      Oui
## 406 12506 47  H Banlieue 292792 Oui    3    Oui      Oui
##  Compte_Courant Emprunt Achat
## 401      Oui    Oui  Oui
## 402      Non    Oui  Non
## 403      Oui    Non  Non
## 404      Non    Oui  Oui
## 405      Oui    Non  Oui
## 406      Oui    Non  Oui
```

summary(achat\_EA)

```
## Age      Sexe      Habitat      Revenus      Marie
## Min.   :18.00 F:198 Banlieue   : 34 Min.   :60392 Non:136
## 1st Qu.:30.00 H:202 Centre_Ville:189 1st Qu.:141684 Oui:264
## Median :43.00      Petite_Ville:115 Median :202552
## Mean   :42.95      Rural      : 62 Mean   :222100
## 3rd Qu.:56.00              3rd Qu.:288900
## Max.   :67.00              Max.   :505040
## Enfants  Voiture  Compte_Epargne Compte_Courant Emprunt  Achat
## Min.    :0.000 Non:213 Non:128 Non: 93 Non:269 Non:221
## 1st Qu.:0.000 Oui:187 Oui:272 Oui:307 Oui:131 Oui:179
## Median :1.000
## Mean    :1.042
## 3rd Qu.:2.000
## Max.    :3.000
```

summary(achat\_ET)

```
## ID Age Sexe Habitat Revenus
## Min. :12501 Min. :18.00 F:102 Banlieue :28 Min. : 50352
## 1st Qu.:12551 1st Qu.:29.75 H: 98 Centre_Ville:80 1st Qu.:133222
## Median :12600 Median :41.00 Petite_Ville:58 Median :197648
## Mean :12600 Mean :41.28 Rural :34 Mean :216898
## 3rd Qu.:12650 3rd Qu.:51.25 3rd Qu.:276556
## Max. :12700 Max. :67.00 Max. :492432
## Marie Enfants Voiture Compte_Epargne Compte_Courant Emprunt
## Non: 68 Min. :0.00 Non: 91 Non: 58 Non: 52 Non:122
## Oui:132 1st Qu.:0.00 Oui:109 Oui:142 Oui:148 Oui: 78
## Median :1.00
## Mean :0.95
## 3rd Qu.:2.00
## Max. :3.00
## Achat
## Non:105
## Oui: 95
##
##
##
##
```

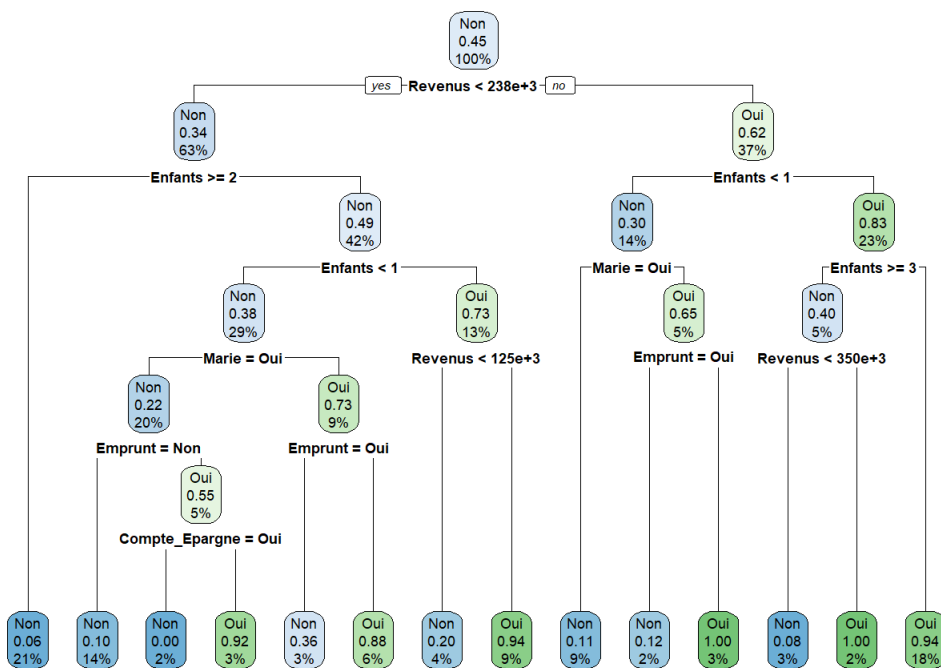
## APPRENTISSAGE DE L'ARBRE RPART

- Construction de l'arbre de décision

```
tree1 <- rpart(Achat~., achat_EA)
```

- Affichage de l'arbre par les fonctions de R Base

```
rpart.plot(tree1)
```



- Application de l'arbre de décision a l'ensemble de test 'achat\_ET'

```
test_tree1 <- predict(tree1, achat_ET, type="class")
test_tree1
```

```
## 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
## Non Non Non Non Non Non Oui Oui Oui Non Oui Non Oui Non Non Non Non Oui Non Non
## 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
## Non Oui Non Non Oui Non Non Oui Oui Non Non Oui Non Oui Non Non Oui Oui Non Non
## 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
## Oui Non Non Non Non Non Non Non Oui Oui Oui Non Non Oui Non Non Non Non Non Oui
## 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
## Oui Non Oui Non Oui Non Non Non Non Non Non Non Non Non Oui Oui Non Oui Non Non Oui
## 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
## Non Oui Non Non Oui Non Non Oui Non Non Oui Oui Non Oui Oui Non Oui Oui Non Non
## 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
## Non Oui Non Non Oui Non Oui Oui Oui Non Non Non Non Non Non Oui Oui Oui Oui Non Non
## 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
## Non Non Non Non Non Oui Non Non Non Non Oui Oui Oui Oui Non Non Non Non Oui Oui Non Non
## 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
## Oui Oui Non Oui Oui Non Non Non Non Oui Oui Non Non Non Non Non Non Non Oui Non Non Oui
## 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
## Oui Non Non Oui Oui Non Oui Non Non Non Non Non Non Non Non Non Non Non Non Non Non
## 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
## Oui Non Non Non Oui Non Oui Oui Non Non Non Non Non Non Non Non Non Non Non Non Non
## Levels: Non Oui
```

```
table(test_tree1)
```

```
## test_tree1
## Non Oui
## 126 74
```

- Ajout des prédictions comme une nouvelle colonne ‘Prediction’ dans le data frame ‘achat\_ET’

```
achat_ET$Prediction <- test_tree1
head(achat)
```

```
## ID Age Sexe Habitat Revenus Marie Enfants Voiture Compte_Epargne
## 1 12101 48 F Centre_Ville 140368 Non 1 Non Non
## 2 12102 40 H Petite_Ville 249360 Oui 3 Oui Non
## 3 12103 51 F Centre_Ville 132600 Oui 0 Oui Oui
## 4 12104 23 F Petite_Ville 163000 Oui 3 Non Non
## 5 12105 57 F Rural 414024 Oui 0 Non Oui
## 6 12106 57 F Petite_Ville 302952 Oui 2 Non Oui
## Compte_Courant Emprunt Achat
## 1 Non Non Oui
## 2 Oui Oui Non
## 3 Oui Non Non
## 4 Oui Non Non
## 5 Non Non Non
## 6 Oui Non Oui
```

- Affichage de liste des exemples de test correctement prédits

```
head(achat_ET[achat_ET$Achat==achat_ET$Prediction,])
```

```
## ID Age Sexe Habitat Revenus Marie Enfants Voiture Compte_Epargne
## 402 12502 26 H Rural 156504 Non 3 Oui Non
## 403 12503 48 H Centre_Ville 308784 Oui 0 Oui Non
## 407 12507 52 H Banlieue 366848 Non 1 Non Oui
## 408 12508 31 H Banlieue 178896 Non 0 Non Oui
## 409 12509 41 F Banlieue 175872 Oui 1 Oui Non
## 410 12510 23 H Banlieue 88584 Oui 2 Non Oui
## Compte_Courant Emprunt Achat Prediction
## 402 Non Oui Non Non
## 403 Oui Non Non Non
## 407 Oui Non Oui Oui
## 408 Oui Non Oui Oui
## 409 Non Non Oui Oui
## 410 Non Non Non Non
```

- Calcul du nombre de succès : nombre d'exemples avec classe réelle et prédiction identiques

```
nbr_succes <- nrow(achat_ET[achat_ET$Achat==achat_ET$Prediction,])
nbr_succes
```

```
## [1] 169
```

- Calcul du taux de succès : nombre de succès sur nombre d'exemples de test

```
taux_succes <- nbr_succes/nrow(achat_ET)
taux_succes
```

```
## [1] 0.845
```

- Calcul du nombre d'échecs : nombre d'exemples avec classe réelle et prédiction différentes

```
nbr_echecs <- nrow(achat_ET[achat_ET$Achat!=achat_ET$Prediction,])
nbr_echecs
```

```
## [1] 31
```

- Calcul du taux d'échecs : nombre d'échecs sur nombre d'exemples de test

```
taux_echecs <- nbr_echecs/nrow(achat_ET)
taux_echecs
```

```
## [1] 0.155
```

## PREDICTIONS PAR L'ARBRE RPART

- Chargement des exemples prospects dans un data frame 'achat\_pro'

```
achat_pro <- read.csv("apprentissage-supervise/partie3-use-case-apprentissage-supervise-data-achat-prospects.csv", header = TRUE, sep = ",", dec = "."
, stringsAsFactors=T)
```

- Application de l'arbre de décision aux prospects dans 'achat\_pro' : classe prédite

```
pred_tree1 <- predict(tree1, achat_pro, type="class")
pred_tree1
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
## Non Non Oui Oui Non Oui Non Oui Non Oui Non Non Non Non Non Oui Non Oui Non Non
## 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
## Oui Non Oui Non Non Non Non Non Oui Oui Oui Non Oui Non Non Non Non Non Non Non
## 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
## Oui Non Oui Oui Oui Oui Non Non Non Non Non Oui Oui Non Non Non Non Non Non Oui
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
## Oui Non Oui Oui Non Oui Non Non Non Oui Oui Non Non Oui Non Oui Non Oui Non Non
## 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
## Oui Oui Non Oui Non Oui Non Non Non Oui Oui Non Non Non Non Non Non Oui Oui Non Non
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
## Non Oui Non Oui Non Non Non Oui Non Non Oui Oui Non Oui Non Non Oui Oui Non Oui
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## Oui Non Oui Non Non Non Non Non Oui Oui Oui Non Non Non Oui Oui Non Oui Oui Oui
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
## Non Non Non Non Oui Oui Non Oui Oui Non Oui Non Oui Oui Oui Non Non Non Non Oui
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
## Non Non Oui Non Non Non Non Non Oui Non Non Non Non Non Oui Non Non Non Non Non Non
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
## Non Oui Non Oui Non Non Non Non Non Non Non Oui Non Non Non Non Non Non Non Non
## Levels: Non Oui
```

```
table(pred_tree1)
```

```
## pred_tree1
## Non Oui
## 126 74
```

- Ajout dans le data frame achat\_pro d'une colonne Prediction contenant la classe prédite

```
achat_pro$Prediction <- pred_tree1
```

- Création d'un data frame contenant les prédictions 'Oui'

```
achat_pro_oui <- achat_pro[achat_pro$Prediction=="Oui",]  
head(achat_pro_oui)
```

```
##      ID Age Sexe   Habitat Revenus Marie Enfants Voiture Compte_Epargne  
## 3 12703 45  F    Rural 175048  Non    0    Oui    Oui  
## 4 12704 50  H Petite_Ville 374352  Oui    2    Non    Oui  
## 6 12706 20  H Centre_Ville 133504  Non    1    Non    Oui  
## 8 12708 50  F Centre_Ville 221920  Oui    1    Oui    Oui  
## 10 12710 57  F Petite_Ville 156968  Oui    1    Oui    Non  
## 16 12716 44  H Petite_Ville 279688  Oui    1    Non    Non  
##  Compte_Courant Emprunt Prediction  
## 3      Oui    Non    Oui  
## 4      Non    Oui    Oui  
## 6      Oui    Oui    Oui  
## 8      Non    Oui    Oui  
## 10     Oui    Non    Oui  
## 16     Non    Oui    Oui
```

```
nrow(achat_pro_oui)
```

```
## [1] 74
```

- Création d'un data frame contenant les prédictions 'Non'

```
achat_pro_non <- achat_pro[achat_pro$Prediction=="Non",]  
head(achat_pro_non)
```

```
##      ID Age Sexe   Habitat Revenus Marie Enfants Voiture Compte_Epargne  
## 1 12701 23  H Centre_Ville 150128  Oui    0    Oui    Oui  
## 2 12702 30  H    Rural 79320  Non    1    Non    Oui  
## 5 12705 41  F Centre_Ville 165768  Oui    0    Oui    Oui  
## 7 12707 46  F    Rural 312544  Oui    0    Oui    Oui  
## 9 12709 42  H Centre_Ville 268672  Non    3    Oui    Oui  
## 11 12711 63  F Centre_Ville 381040  Oui    0    Non    Oui  
##  Compte_Courant Emprunt Prediction  
## 1      Non    Oui    Non  
## 2      Non    Oui    Non  
## 5      Oui    Non    Non  
## 7      Non    Oui    Non  
## 9      Non    Non    Non  
## 11     Non    Oui    Non
```

```
nrow(achat_pro_non)
```

```
## [1] 126
```