DecisionTree

Flo

2024-04-05

Remaqrque

Les jambes ne servent a rien pour les decision tree

```
data=read.csv("preprocessing_data_UFC.csv")
head(data)
```

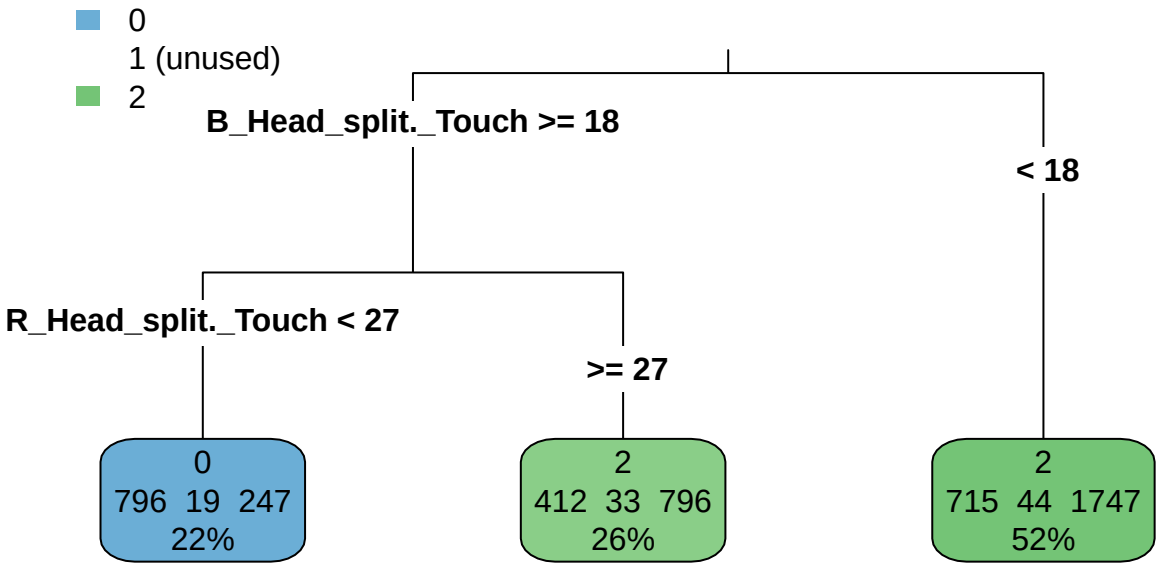
```
##         R_fighter         B_fighter R_KD B_KD R_SUB_ATT B_SUB_ATT R_REV B_REV
## 1    Adrian Yanez     Gustavo Lopez    2    0         0         0     0     0
## 2    Trevin Giles     Roman Dolidze    0    0         1         2     0     1
## 3     Tai Tuivasa   Harry Hunsucker    1    0         0         0     0     0
## 4   Cheyanne Buys Montserrat Conejo    0    0         0         2     3     1
## 5   Marion Reneau     Macy Chiasson    0    0         0         0     0     1
## 6 Leonardo Santos      Grant Dawson    0    0         0         0     0     0
##   R_CTRL B_CTRL              win_by last_round last_round_time           date
## 1   0:03   0:00               KO/TKO          3            0:27 March 20, 2021
## 2   1:15   4:15 Decision - Unanimous          3            5:00 March 20, 2021
## 3   0:10   0:00               KO/TKO          1            0:49 March 20, 2021
## 4   1:04   9:53 Decision - Unanimous          3            5:00 March 20, 2021
## 5   2:15   3:48 Decision - Unanimous          3            5:00 March 20, 2021
## 6   1:21   8:18               KO/TKO          3            4:59 March 20, 2021
##                  Fight_type            Winner R_Head_split._Touch
## 1         Bantamweight Bout      Adrian Yanez                  32
## 2         Middleweight Bout      Trevin Giles                  22
## 3          Heavyweight Bout       Tai Tuivasa                  10
## 4  Women's Strawweight Bout Montserrat Conejo                  26
## 5 Women's Bantamweight Bout     Macy Chiasson                  14
## 6          Lightweight Bout      Grant Dawson                  14
##   R_Head_split._NoTouch B_Head_split._Touch B_Head_split._NoTouch
## 1                    83                  14                    40
## 2                    51                  10                    37
## 3                    14                   1                     5
## 4                    60                  10                    35
## 5                    40                  29                   110
## 6                    45                  16                    48
##   R_Body_split._Touch R_Body_split._NoTouch B_Body_split._Touch
## 1                   8                    19                   5
## 2                   4                     4                   7
## 3                   0                     0                   0
## 4                   5                     5                   0
## 5                   7                    13                  15
## 6                   6                    10                  23
##   B_Body_split._NoTouch R_Leg_split._Touch R_Leg_split._NoTouch
```

```
## 1                     7                  1                    1
## 2                    14                  1                    2
## 3                     0                  4                    4
## 4                     1                  0                    0
## 5                    20                  9                   10
## 6                    27                 10                   12
##   B_Leg_split._Touch B_Leg_split._NoTouch R_Distance_split._Touch
## 1                  4                    4                     41
## 2                 15                   16                     15
## 3                  1                    1                      9
## 4                  5                    5                     26
## 5                  7                    8                     25
## 6                  7                    9                     28
##   R_Distance_split._NoTouch B_Distance_split._Touch B_Distance_split._NoTouch
## 1                       102                     23                       51
## 2                        42                     28                       59
## 3                        10                      2                        6
## 4                        56                     15                       41
## 5                        54                     36                      119
## 6                        65                     33                       68
##   R_SIG_STR_split._Touch R_SIG_STR_split._NoTouch B_SIG_STR_split._Touch
## 1                     41                     103                     23
## 2                     27                      57                     32
## 3                     14                      18                      2
## 4                     31                      65                     15
## 5                     30                      63                     51
## 6                     30                      67                     46
##   B_SIG_STR_split._NoTouch R_TOTAL_STR_split._Touch R_TOTAL_STR_split._NoTouch
## 1                       51                      41                       103
## 2                       67                      43                        73
## 3                        6                      14                        18
## 4                       41                      49                        87
## 5                      138                      59                        93
## 6                       84                      74                       115
##   B_TOTAL_STR_split._Touch B_TOTAL_STR_split._NoTouch R_TD_split._Touch
## 1                       23                        51                0
## 2                       75                       110                1
## 3                        2                         6                0
## 4                      136                       168                0
## 5                       92                       184                2
## 6                       75                       132                1
##   R_TD_split._NoTouch B_TD_split._Touch B_TD_split._NoTouch
## 1                   0                0                   1
## 2                   2                1                   3
## 3                   0                0                   0
## 4                   0                4                   5
## 5                   4                1                   1
## 6                   2                1                  13
##   R_CLINCH_split._Touch R_CLINCH_split._NoTouch B_CLINCH_split._Touch
## 1                     0                      0                    0
## 2                     4                      5                    3
## 3                     0                      0                    0
## 4                     2                      2                    0
## 5                     5                      9                   13
```

```
## 6                     2                       2                     9
##   B_CLINCH_split._NoTouch R_GROUND_split._Touch R_GROUND_split._NoTouch
## 1                       0                     0                       1
## 2                       6                     8                      10
## 3                       0                     5                       8
## 4                       0                     3                       7
## 5                      15                     0                       0
## 6                      11                     0                       0
##   B_GROUND_split._Touch B_GROUND_split._NoTouch B_TOTAL_ROUNDS R_TOTAL_ROUNDS
## 1                     0                       0              4               1
## 2                     1                       2              4              17
## 3                     0                       0              0              13
## 4                     0                       0              0               0
## 5                     2                       4             11              30
## 6                     4                       5             10              18
##   B_CURRENT_WIN_STREAK R_CURRENT_WIN_STREAK B_LONGEST_WIN_STREAK
## 1                    0                    1                     1
## 2                    2                    2                     2
## 3                    0                    3                     0
## 4                    0                    0                     0
## 5                    3                    2                     3
## 6                    4                    1                     4
##   R_LONGEST_WIN_STREAK B_WINS R_WINS B_LOSSES R_LOSSES B_HEIGHT R_HEIGHT
## 1                    1      1      1        1        0   165.10   170.18
## 2                    2      2      4        0        2   187.96   182.88
## 3                    3      0      4        0        3   187.96   187.96
## 4                    0      0      0        0        0   152.40   160.02
## 5                    2      4      5        1        6   180.34   167.64
## 6                    6      4      7        0        1   177.80   182.88
##   B_WEIGHT R_WEIGHT B_REACH R_REACH R_AGE B_AGE R_WIN_DECISION_MAJORITY
## 1      135      135  170.18  177.80    27    31                       0
## 2      205      185  193.04  187.96    28    32                       0
## 3      241      264  190.50  190.50    28    32                       0
## 4      115      115  154.94  160.02    25    28                       0
## 5      135      135  182.88  172.72    43    29                       0
## 6      145      155  182.88  190.50    41    27                       0
##   B_WIN_DECISION_MAJORITY R_WIN_DECISION_SPLIT B_WIN_DECISION_SPLIT
## 1                       0                    0                   0
## 2                       0                    1                   1
## 3                       0                    0                   0
## 4                       0                    0                   0
## 5                       0                    0                   0
## 6                       0                    1                   0
##   R_WIN_DECISION_UNANIMOUS B_WIN_DECISION_UNANIMOUS R_WIN_KO_TKO B_WIN_KO_TKO
## 1                        0                        0           1            0
## 2                        0                        0           3            1
## 3                        1                        0           3            0
## 4                        0                        0           0            0
## 5                        1                        1           2            2
## 6                        2                        2           2            0
##   R_WIN_SUB B_WIN_SUB R_WIN_TKO_STOP_DOCTORS B_WIN_TKO_STOP_DOCTORS WINNER
## 1         0         1                      0                      0      2
## 2         0         0                      0                      0      2
## 3         0         0                      0                      0      2
```

```
## 4         0         0                     0                     0      0
## 5         2         1                     0                     0      0
## 6         2         2                     0                     0      0
##   WEIGHT_CLASS R_STANCE B_STANCE
## 1            1        2        2
## 2            8        2        2
## 3            5        4        2
## 4           14        5        4
## 5           11        2        2
## 6            7        2        5
```

```r
# Charger le package nécessaire
library(rpart)
library(rpart.plot)

# Charger les données
df <- read.csv("preprocessing_data_UFC.csv")
create_train_test <- function(data, size = 0.8, train = TRUE) {
    n_row = nrow(data)
    total_row = size * n_row
    train_sample = 1: total_row
    if (train == TRUE) {
        return (data[train_sample, ])
    } else {
        return (data[-train_sample, ])
    }
}

data_train <- create_train_test(df, 0.8, train = TRUE)
data_test <- create_train_test(df, 0.8, train = FALSE)
# a voir comment faire j'ai mit cette ligne car sinon matrice rectanguaire et donc pb dans le calcul de
data_test$WINNER[1]=0




# Définir les paramètres de contrôle
control <- rpart.control(cp = 0.12, # complexité du modèle
                         minbucket = round(5 / 3),
                         minsplit = 4, # nombre minimum d'observations pour diviser un nœud
                         maxdepth = 3) # profondeur maximale de l'arbre
# Fit a decision tree using the Gini index
treeFitted_1 <- rpart(WINNER ~ R_Head_split._Touch + B_Head_split._Touch, data = data_train, method = "

# Plot the decision tree
rpart.plot(treeFitted_1, type = 3, extra = 101)
```

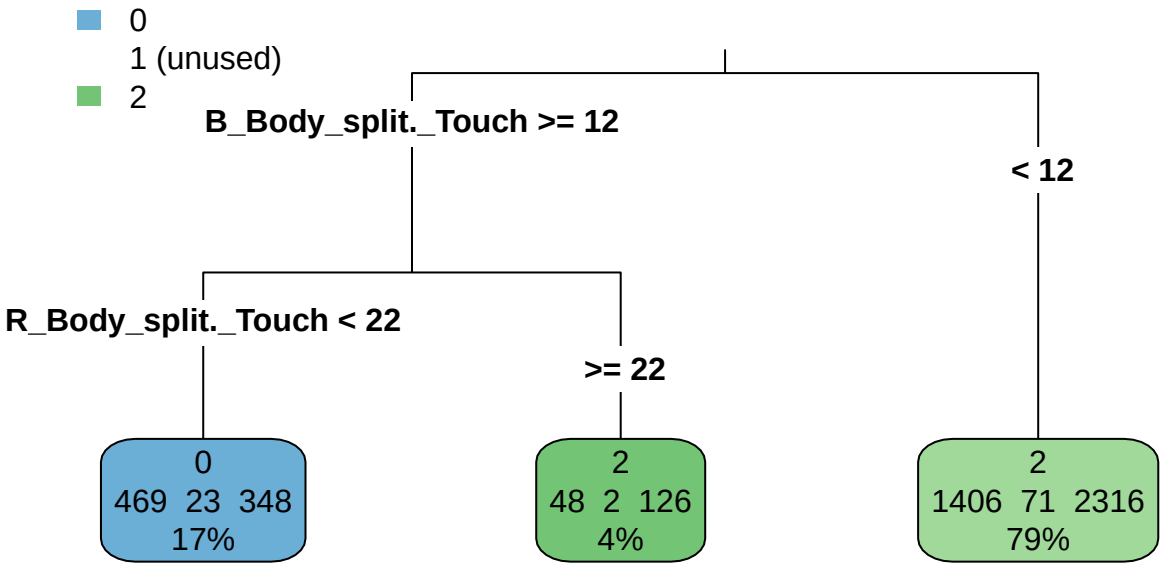



```
# Make prediction
predictionWinner_1 <- predict(treeFitted_1, data_test, type = "class")

# Print the prediction
#print(paste("The predicted winner is: ", predictionWinner_1))
table_mat_1 <- table(data_test$WINNER, predictionWinner_1)
table_mat_1
```

```
##    predictionWinner_1
##        0    1    2
##   0    0    0    1
##   1    0    0   14
##   2   68    0 1120
```

```
accuracy_Test_1 <- sum(diag(table_mat_1)) / sum(table_mat_1)
```

```
# Définir les paramètres de contrôle
control <- rpart.control(cp = 0.02, # complexité du modèle
                         minbucket = round(5 / 3),
                         minsplit = 4, # nombre minimum d'observations pour diviser un nœud
                         maxdepth = 3) # profondeur maximale de l'arbre
# Fit a decision tree using the Gini index
treeFitted_2 <- rpart(WINNER ~ R_Body_split._Touch + B_Body_split._Touch , data = data_train, method =

# Plot the decision tree
rpart.plot(treeFitted_2, type = 3, extra = 101)
```

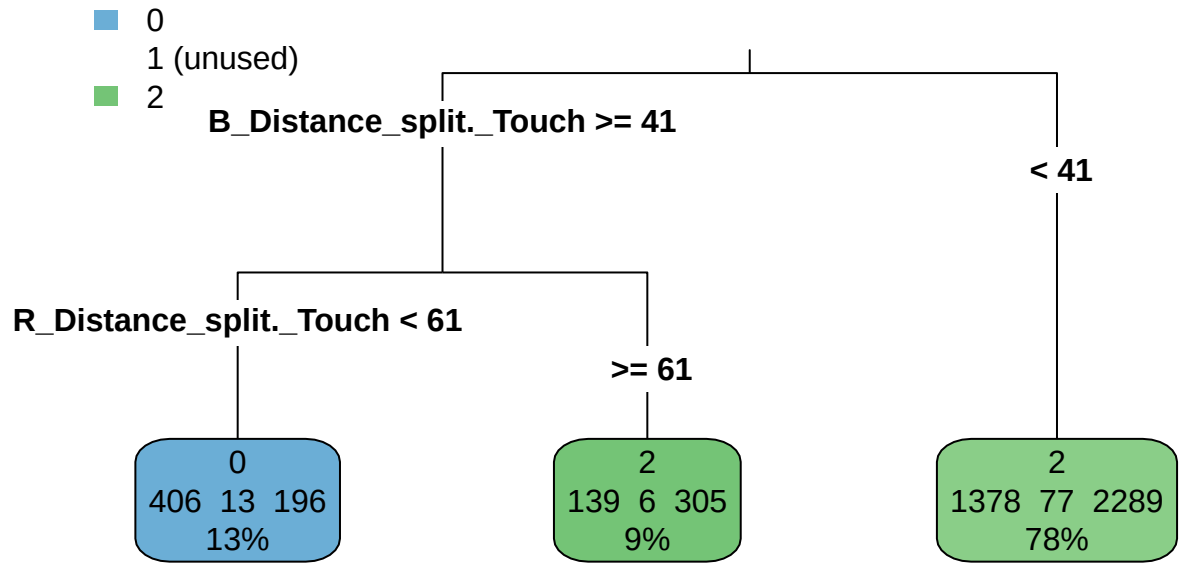



```
# Make prediction
predictionWinner_2 <- predict(treeFitted_2, data_test, type = "class")

# Print the prediction
#print(paste("The predicted winner is: ", predictionWinner_1))
table_mat_2 <- table(data_test$WINNER, predictionWinner_2)
table_mat_2
```

```
##    predictionWinner_2
##        0    1    2
##   0    0    0    1
##   1    2    0   12
##   2   48    0 1140
```

```
accuracy_Test_2 <- sum(diag(table_mat_2)) / sum(table_mat_2)
```

```
# Définir les paramètres de contrôle
control <- rpart.control(cp = 0.02, # complexité du modèle
                         minbucket = round(5 / 3),
                         minsplit = 4, # nombre minimum d'observations pour diviser un nœud
                         maxdepth = 3) # profondeur maximale de l'arbre
# Fit a decision tree using the Gini index
treeFitted_3 <- rpart(WINNER ~ R_Distance_split._Touch + B_Distance_split._Touch , data = data_train, m

# Plot the decision tree
rpart.plot(treeFitted_3, type = 3, extra = 101)
```




```
# Make prediction
predictionWinner_3 <- predict(treeFitted_3, data_test, type = "class")

# Print the prediction
#print(paste("The predicted winner is: ", predictionWinner_1))
table_mat_3 <- table(data_test$WINNER, predictionWinner_3)
table_mat_3
```

```
##    predictionWinner_3
##        0    1    2
##   0    0    0    1
##   1    0    0   14
##   2   23    0 1165
```

```
accuracy_Test_3 <- sum(diag(table_mat_3)) / sum(table_mat_3)
```

```
# Vérifier si les trois prédictions sont différentes
diff1 <- predictionWinner_1 != predictionWinner_2
diff2 <- predictionWinner_1 != predictionWinner_3
diff3 <- predictionWinner_2 != predictionWinner_3

# Trouver les indices des observations pour lesquelles les prédictions sont différentes
indices <- which(diff1 | diff2 | diff3)

# Afficher les indices
print(indices)
```

```
##   [1]    3   27   31   32   37   39   46   47   49   56   61   63   67   77   84
```

```
##  [16]  101  103  105  107  112  113  117  121  124  130  132  147  155  156  161
##  [31]  163  165  167  174  175  183  198  200  211  215  223  227  229  231  232
##  [46]  236  237  241  248  260  272  278  283  284  285  295  300  303  325  326
##  [61]  332  333  351  352  355  392  400  401  402  406  423  434  455  470  480
##  [76]  490  494  502  518  533  556  566  569  577  578  600  605  624  630  647
##  [91]  665  666  685  699  748  760  768  790  793  818  823  825  843  851  859
## [106]  861  864  869  884  896  952  977  979 1004 1013 1017 1026 1042 1044 1046
## [121] 1055 1107 1112
```

Pseudo Random Forest

```
numberOfTree=3
indicesDiff=4809+825
RealWinner=df$WINNER[indicesDiff]

predictionWinner_RF_T1 <- predict(treeFitted_1, data.frame(R_Head_split._Touch = df$R_Head_split._Touch
resT1=print(paste("The predicted winner is: ", predictionWinner_RF_T1))
```

```
## [1] "The predicted winner is:  0"
```

```
predictionWinner_RF_T2 <- predict(treeFitted_2, data.frame(R_Body_split._Touch = df$R_Body_split._Touch
resT2=print(paste("The predicted winner is: ", predictionWinner_RF_T2))
```

```
## [1] "The predicted winner is:  2"
```

```
predictionWinner_RF_T3 <- predict(treeFitted_3, data.frame(R_Distance_split._Touch = df$R_Distance_spli
resT3=print(paste("The predicted winner is: ", predictionWinner_RF_T3))
```

```
## [1] "The predicted winner is:  2"
```

```
res <- numeric(numberOfTree)
for (i in 1:numberOfTree) {
  split_string <- strsplit(get(paste0("resT", i)), split = " ")[[1]]
  res[i] <- as.integer(split_string[length(split_string)])
}

#res
# Compter les occurrences de chaque valeur dans le tableau
occurrences <- table(res)

# Trouver le nombre maximal d'occurrences
max_occurrences <- max(occurrences)

# Trouver les valeurs qui ont le nombre maximal d'occurrences
majority_vote <- names(occurrences[occurrences == max_occurrences])


# Afficher le résultat
print(RealWinner)
```

```
## [1] 2
```

```
print(paste0("On décide que le winner est dans la catégorie : ", majority_vote))
```

```
## [1] "On décide que le winner est dans la catégorie : 2"
```

```r
# Création d'un exemple de tableau
tableau <- res

# Compter les occurrences de chaque valeur dans le tableau
occurrences <- table(tableau)

# Trouver le nombre maximal d'occurrences
max_occurrences <- max(occurrences)

# Trouver les valeurs qui ont le nombre maximal d'occurrences
max_value <- names(occurrences[occurrences == max_occurrences])

print(max_value)
```

```
## [1] "2"
```