rate

2023-05-19

```
data<-read.csv("D://dataformodel.csv")
question<-read.csv("C://yasuo//questions.csv")
question$ClosedByAttorneyUno <- ifelse(question$ClosedByAttorneyUno=="NULL",0,1)
#琛ㄦ牸鍚堝苟
questiondf<-question[,c("QuestionUno","ClosedByAttorneyUno")]
# 浣跨敤merge鍑芥暟鏍规嵁QuestionUno鍒楄繘閰嶆暟鎹?
data <- merge(questiondf, data, by = "QuestionUno", all.x = FALSE)
class(data$Category)#鏌ョ湅鍘熷□绫诲瀷
```

```
## [1] "character"
```

```
class(data$StateAbbr)
```

```
## [1] "character"
```

```
data$Category<-as.factor(data$Category)#杞□寲涓哄洜瀛愬瀷鍙橀噺
data$StateAbbr<-as.factor(data$StateAbbr)
#璋冩暣鏃堕棿鏍煎紡
data$Time_m <- as.numeric(as.POSIXlt(data$Time_m, format = "%H:%M"))
summary(data$StateAbbr)#杞□寲鍚庢暟鎹□綋琛ㄧ幇鎯呭喌
```

```
##    AK    AL    AR    AZ    CA    CT    FL    GA    HI    IA    IL    IN    KS
##   370   498  1606  1176   953   712 12958  1255  1225   588  7275  6244    37
##    LA    MA    MD    ME    MI    MO    MS    NC    NE    NH    NJ    NM    NY
##  2225  5508  1476  2563   189  5331   751  3780  1733   894   558   634  4172
##    OK    PA    SC    SD    TN    TX    US    UT    VA    VT    WI    WV    WY
##  2310   118  5848   805  8905  8322    93  1241  3950   577  4578  1387   913
```

```
summary(data$Time_m)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 1.684e+09 1.684e+09 1.684e+09 1.684e+09 1.684e+09 1.685e+09
```

```
column_counts <-sapply(data, function(x) sum(!is.na(x), na.rm = TRUE))
column_counts_df <- data.frame(Column = names(column_counts), Count = column_counts)
data <- na.omit(data)
summary_stats <- summary(data)
# 灏哠ummary_stats杞□崲涓烘暟鎹□□
summary_table <- as.data.frame(summary_stats)
# 浣跨敤kable()鍑芥暟鐢熸垚琛ㄦ牸
#杩涜□鍒嗘被妯″瀷娴嬭瘯
#閫昏緫鍥炲綊妯″瀷
library(stats)
model_glm_rate <- glm(ClosedByAttorneyUno ~ SRate+length+AQRate+Time_m+CRate, data = data, fami
ly = binomial())
model_glm_factor <- glm(ClosedByAttorneyUno ~ StateAbbr+length+AQRate+Time_m+Category, data = d
ata, family = binomial())
summary(model_glm_rate)
```

```
##
## Call:
## glm(formula = ClosedByAttorneyUno ~ SRate + length + AQRate +
##     Time_m + CRate, family = binomial(), data = data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.497e+03  7.877e+02   3.169  0.00153 **
## SRate        1.978e+00  7.662e-02  25.821  < 2e-16 ***
## length       1.038e-03  5.671e-04   1.830  0.06723 .
## AQRate       1.004e-01  3.466e-02   2.895  0.00379 **
## Time_m      -1.483e-06  4.676e-07  -3.170  0.00152 **
## CRate        2.102e+00  2.569e-01   8.183 2.78e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61953  on 103757  degrees of freedom
## Residual deviance: 61265  on 103752  degrees of freedom
## AIC: 61277
##
## Number of Fisher Scoring iterations: 5
```

```
summary(model_glm_factor)
```

```
##
## Call:
## glm(formula = ClosedByAttorneyUno ~ StateAbbr + length + AQRate +
##     Time_m + Category, family = binomial(), data = data)
##
## Coefficients:
##                                          Estimate Std. Error z value
## (Intercept)                             1.954e+03  7.941e+02   2.461
## StateAbbrAL                             3.153e-01  2.213e-01   1.425
## StateAbbrAR                             4.256e-01  1.838e-01   2.315
## StateAbbrAZ                            -4.984e-01  1.771e-01  -2.814
## StateAbbrCA                            -3.910e-02  1.860e-01  -0.210
## StateAbbrCT                            -2.287e-01  1.903e-01  -1.202
## StateAbbrFL                             8.269e-01  1.638e-01   5.049
## StateAbbrGA                            -8.179e-01  1.725e-01  -4.741
## StateAbbrHI                             3.718e-01  1.900e-01   1.957
## StateAbbrIA                            -1.882e-01  1.977e-01  -0.952
## StateAbbrIL                             4.169e-01  1.650e-01   2.526
## StateAbbrIN                             6.473e-01  1.674e-01   3.867
## StateAbbrKS                            -1.296e+00  3.806e-01  -3.406
## StateAbbrLA                             4.480e-01  1.775e-01   2.524
## StateAbbrMA                             3.407e-01  1.663e-01   2.049
## StateAbbrMD                             2.080e-01  1.808e-01   1.151
## StateAbbrME                             4.952e-01  1.762e-01   2.810
## StateAbbrMI                             6.897e-01  3.292e-01   2.095
## StateAbbrMO                             3.270e-01  1.664e-01   1.966
## StateAbbrMS                             2.633e-01  2.035e-01   1.294
## StateAbbrNC                             3.618e-01  1.692e-01   2.138
## StateAbbrNE                             1.614e-01  1.777e-01   0.908
## StateAbbrNH                             1.959e-01  1.936e-01   1.012
## StateAbbrNJ                             2.428e-01  2.108e-01   1.151
## StateAbbrNM                             8.196e-02  2.029e-01   0.404
## StateAbbrNY                             5.482e-01  1.694e-01   3.236
## StateAbbrOK                             2.723e-01  1.745e-01   1.561
## StateAbbrPA                            -6.806e-01  2.738e-01  -2.485
## StateAbbrSC                             1.615e+00  1.785e-01   9.050
## StateAbbrSD                             2.420e-01  1.995e-01   1.213
## StateAbbrTN                             3.805e-01  1.639e-01   2.322
## StateAbbrTX                             3.673e-01  1.640e-01   2.240
## StateAbbrUS                            -3.750e-01  3.079e-01  -1.218
## StateAbbrUT                             1.804e-01  1.843e-01   0.978
## StateAbbrVA                             4.156e-01  1.691e-01   2.458
## StateAbbrVT                             3.424e-01  2.143e-01   1.598
## StateAbbrWI                             4.229e-01  1.681e-01   2.516
## StateAbbrWV                            -1.325e+00  1.694e-01  -7.823
## StateAbbrWY                             1.010e-01  1.921e-01   0.526
## length                                  1.239e-03  5.751e-04   2.155
## AQRate                                  7.778e-02  3.523e-02   2.208
## Time_m                                 -1.159e-06  4.714e-07  -2.459
## CategoryEducation                       1.254e-02  1.987e-01   0.063
## CategoryFamily and Children             3.829e-01  3.926e-02   9.754
## CategoryHealth and Disability          -1.303e-01  9.489e-02  -1.373
## CategoryHousing and Homelessness        1.038e-01  4.279e-02   2.425
## CategoryIncome Maintenance             -5.863e-02  9.699e-02  -0.604
## CategoryIndividual Rights              -5.951e-02  6.360e-02  -0.936
## CategoryJuvenile                        6.006e-01  3.466e-01   1.733
## CategoryOther                           3.111e-03  4.319e-02   0.072
```

```
## CategoryWork, Employment and Unemployment  4.295e-02  5.827e-02   0.737
##                                           Pr(>|z|)
## (Intercept)                               0.013857 *
## StateAbbrAL                               0.154213
## StateAbbrAR                               0.020612 *
## StateAbbrAZ                               0.004891 **
## StateAbbrCA                               0.833557
## StateAbbrCT                               0.229412
## StateAbbrFL                               4.44e-07 ***
## StateAbbrGA                               2.13e-06 ***
## StateAbbrHI                               0.050350 .
## StateAbbrIA                               0.341053
## StateAbbrIL                               0.011550 *
## StateAbbrIN                               0.000110 ***
## StateAbbrKS                               0.000659 ***
## StateAbbrLA                               0.011602 *
## StateAbbrMA                               0.040456 *
## StateAbbrMD                               0.249873
## StateAbbrME                               0.004950 **
## StateAbbrMI                               0.036182 *
## StateAbbrMO                               0.049352 *
## StateAbbrMS                               0.195807
## StateAbbrNC                               0.032520 *
## StateAbbrNE                               0.363775
## StateAbbrNH                               0.311574
## StateAbbrNJ                               0.249529
## StateAbbrNM                               0.686290
## StateAbbrNY                               0.001214 **
## StateAbbrOK                               0.118608
## StateAbbrPA                               0.012940 *
## StateAbbrSC                               < 2e-16 ***
## StateAbbrSD                               0.225130
## StateAbbrTN                               0.020229 *
## StateAbbrTX                               0.025067 *
## StateAbbrUS                               0.223216
## StateAbbrUT                               0.327828
## StateAbbrVA                               0.013967 *
## StateAbbrVT                               0.110133
## StateAbbrWI                               0.011863 *
## StateAbbrWV                               5.16e-15 ***
## StateAbbrWY                               0.598857
## length                                    0.031176 *
## AQRate                                    0.027275 *
## Time_m                                    0.013933 *
## CategoryEducation                         0.949696
## CategoryFamily and Children               < 2e-16 ***
## CategoryHealth and Disability             0.169825
## CategoryHousing and Homelessness          0.015290 *
## CategoryIncome Maintenance                0.545513
## CategoryIndividual Rights                 0.349403
## CategoryJuvenile                          0.083126 .
## CategoryOther                             0.942570
## CategoryWork, Employment and Unemployment 0.461083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 61953  on 103757  degrees of freedom
## Residual deviance: 59920  on 103707  degrees of freedom
## AIC: 60022
##
## Number of Fisher Scoring iterations: 6
```

```
#鍗℃柟妫€楠?
result_time_m <- chisq.test(table(data$ClosedByAttorneyUno,data$Time_m))
```

```
## Warning in chisq.test(table(data$ClosedByAttorneyUno, data$Time_m)):
## Chi-squared近似算法有可能不准
```

```
result_state<-chisq.test(table(data$ClosedByAttorneyUno,data$StateAbbr))
```

```
## Warning in chisq.test(table(data$ClosedByAttorneyUno, data$StateAbbr)):
## Chi-squared近似算法有可能不准
```

```
result_length<-chisq.test(table(data$ClosedByAttorneyUno,data$length))
```

```
## Warning in chisq.test(table(data$ClosedByAttorneyUno, data$length)):
## Chi-squared近似算法有可能不准
```

```
result_Cate<-chisq.test(table(data$ClosedByAttorneyUno,data$Category))
result_AQrate<-chisq.test(table(data$ClosedByAttorneyUno,data$AQRate))
#璁□粌闆嗘祴璇曢泦鍒掑垎
data$ClosedByAttorneyUno<-as.factor(data$ClosedByAttorneyUno)
library(caret)
```

```
## 载入需要的程辑包：ggplot2
```

```
## 载入需要的程辑包：lattice
```

```
index <- createDataPartition(data$ClosedByAttorneyUno, p = 0.7, list = FALSE)
# 鏍规嵁绱㈠紩鍒掑垎璁□粌闆嗗拰娴嬭瘯闆?
train_data <- data[index, ]
test_data <- data[-index, ]
#閫昏緫鍥炲綊妯″瀷
model_glm_factor <- glm(ClosedByAttorneyUno ~ StateAbbr+length+AQRate+Time_m+Category, data = t
rain_data, family = binomial())
predicted_glm <- predict(model_glm_factor, newdata = test_data)
predicted_glm <- ifelse(predicted_glm >= 0.5, "1", "0")
# 鏍规嵁棰勬祴缁撴灉鍜屽疄闄呮爣绛剧敓鎴愭贩娣嗙煩闃?
confusion_matrix_glm <- table(predicted_glm, test_data$ClosedByAttorneyUno)
# 璁＄畻姝ｇ‘鐜?
accuracy_glm <- sum(diag(confusion_matrix_glm)) / sum(confusion_matrix_glm)
#鍐崇瓥鏍?
library(rpart)
# 鏋勫缓鍐崇瓥鏍戞ā鍨?
model_rtree <- rpart(ClosedByAttorneyUno ~ StateAbbr+length+AQRate+Time_m+Category,data=data)
summary(model_rtree)
```

```
## Call:
## rpart(formula = ClosedByAttorneyUno ~ StateAbbr + length + AQRate +
##     Time_m + Category, data = data)
##   n= 103758
##
##   CP nsplit rel error xerror xstd
## 1  0      0         1      0    0
##
## Node number 1: 103758 observations
##   predicted class=1  expected loss=0.08828235  P(node) =1
##     class counts:  9160 94598
##    probabilities: 0.088 0.912
```

```
predicted_rtree <- predict(model_rtree, newdata = test_data,type="class")
confusion_matrix_rtree <- table(predicted_rtree, test_data$ClosedByAttorneyUno)
# 璁＄畻鍑嗙‘鐜?
accuracy_rtree <- sum(diag(confusion_matrix_rtree)) / sum(confusion_matrix_rtree)
#闄忔満妫□灄
library(randomForest)
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## 载入程辑包：'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
model_forest <- randomForest(ClosedByAttorneyUno ~ StateAbbr+length+AQRate+Time_m+Category, dat
a = train_data)
predicted_forest <- predict(model_forest, newdata = test_data,type="class")
confusion_matrix_forest <- table(predicted_forest, test_data$ClosedByAttorneyUno)
# 璁＄畻鍑嗙‘鐜?
accuracy_forest <- sum(diag(confusion_matrix_forest)) / sum(confusion_matrix_forest)
#SVM
library(e1071)
model_svm <- svm(ClosedByAttorneyUno ~ StateAbbr+length+AQRate+Time_m+Category, data = train_da
ta)
predicted_svm <- predict(model_svm, newdata = test_data,type="class")
confusion_matrix_svm <- table(predicted_svm, test_data$ClosedByAttorneyUno)
# 璁＄畻鍑嗙‘鐜?
accuracy_svm <- sum(diag(confusion_matrix_svm)) / sum(confusion_matrix_svm)
```