project_eng

July 20, 2023

[22]:
```
data = read.table('/content/adult.csv', sep = ',', header = TRUE)

head(data)
```

A data.frame: 6 × 15

	age	workclass	fnlwgt	education	educational.num	marital.status	o
	<int>	<chr>	<int>	<chr>	<int>	<chr>	<
1	25	Private	226802	11th	7	Never-married	M
2	38	Private	89814	HS-grad	9	Married-civ-spouse	F
3	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	P
4	44	Private	160323	Some-college	10	Married-civ-spouse	M
5	18	?	103497	Some-college	10	Never-married	?
6	34	Private	198693	10th	6	Never-married	C

Listing of attributes:

50K, <=50K.

age: continuous. workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked. fnlwgt: continuous. education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool. education-num: continuous. marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse. occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces. relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried. race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black. sex: Female, Male. capital-gain: continuous. capital-loss: continuous. hours-per-week: continuous. native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

0.1 Zad 1

0.1.1 In the type_employer column:

1. replace the entries "Federal-gov" and "Local-gov" with "SL-gov"

2. replace the entries "Self-emp-inc" and "Self-emp-not-inc" with "self-emp"

[7]:
```
install.packages('caret')
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘listenv’, ‘parallelly’, ‘future’, ‘globals’,
‘shape’, ‘future.apply’, ‘numDeriv’, ‘progressr’, ‘SQUAREM’, ‘diagram’, ‘lava’,
‘prodlim’, ‘proxy’, ‘iterators’, ‘Rcpp’, ‘clock’, ‘gower’, ‘hardhat’, ‘ipred’,
‘timeDate’, ‘e1071’, ‘foreach’, ‘ModelMetrics’, ‘plyr’, ‘pROC’, ‘recipes’,
‘reshape2’
```

[5]:
```
library(tidyverse)
```

```
  Attaching packages                                    tidyverse
1.3.1

  ggplot2 3.4.2       purrr   1.0.1
  tibble  3.2.1       dplyr   1.1.2
  tidyr   1.3.0       stringr 1.5.0
  readr   2.1.4       forcats 1.0.0

  Conflicts
tidyverse_conflicts()
  dplyr::filter() masks stats::filter()
  dplyr::lag()    masks stats::lag()
```

[8]:
```
library(caret)
```

```
Loading required package: lattice


Attaching package: ‘caret’


The following object is masked from ‘package:purrr’:

    lift
```

[23]:
```
table(data$workclass)
```

```
                 ?      Federal-gov        Local-gov     Never-worked
              2799             1432             3136               10
           Private     Self-emp-inc Self-emp-not-inc        State-gov
```

```
          33906             1695             3862             1981
     Without-pay
              21
```

[24]:
```
data$workclass = str_replace(data$workclass, c("Local-gov|Federal-gov"),␣
 ↪"SL-gov")
data$workclass = str_replace(data$workclass,␣
 ↪c("Self-emp-inc|Self-emp-not-inc"), "self-emp")
table(data$workclass)
```

```
          ? Never-worked      Private     self-emp       SL-gov    State-gov
       2799           10        33906         5557         4568         1981
Without-pay
         21
```

0.2 Zad 2

0.2.1 In the marital column, reduce the number of entries to three (Married;Not-Married;Never-Married)

[25]:
```
table(data$marital.status)
```

```
             Divorced     Married-AF-spouse    Married-civ-spouse
                 6633                    37                 22379
Married-spouse-absent         Never-married             Separated
                  628                 16117                  1530
              Widowed
                 1518
```

[26]:
```
data$marital.status = str_replace(data$marital.status,␣
 ↪c("Married-AF-spouse|Married-civ-spouse|Married-spouse-absent|Separated"),␣
 ↪"Married")
data$marital.status = str_replace(data$marital.status, c("Divorced|Widowed"),␣
 ↪"Not-Married")
table(data$marital.status)
```

```
      Married Never-married   Not-Married
        24574         16117          8151
```

0.3 Zad 3

0.3.1 Reduce the number of entries in the country column (e.g. grouping by continents? other approach?).

[27]:
```
data %>%
    group_by(native.country) %>%
    summarise(disp = mean(age))
```

A tibble: 42 × 2

native.country <chr>	disp <dbl>
?	38.77246
Cambodia	36.89286
Canada	44.04945
China	41.85246
Columbia	39.45882
Cuba	46.35507
Dominican-Republic	37.97087
Ecuador	37.66667
El-Salvador	33.38065
England	40.52756
France	40.31579
Germany	38.60194
Greece	45.83673
Guatemala	32.09091
Haiti	38.60000
Holand-Netherlands	32.00000
Honduras	35.05000
Hong	34.23333
Hungary	50.36842
India	38.36424
Iran	38.37288
Ireland	38.48649
Italy	45.41905
Jamaica	37.14151
Japan	37.35870
Laos	35.21739
Mexico	33.63512
Nicaragua	36.28571
Outlying-US(Guam-USVI-etc)	38.82609
Peru	36.43478
Philippines	39.63390
Poland	42.75862
Portugal	41.23881
Puerto-Rico	39.86413
Scotland	46.76190
South	38.09565
Taiwan	34.18462
Thailand	37.66667
Trinadad&Tobago	39.25926
United-States	38.69869
Vietnam	34.61628
Yugoslavia	40.47826

0.4 Zad 4

0.4.1 Replace "?" entries with NA values

[28]:
```
table(data$workclass)
```

```
           ? Never-worked      Private     self-emp       SL-gov    State-gov
        2799           10        33906         5557         4568         1981
 Without-pay
          21
```

[29]:
```
data[data == '?'] = NA
table(data$workclass)
```

```
Never-worked      Private     self-emp       SL-gov    State-gov  Without-pay
          10        33906         5557         4568         1981           21
```

0.5 Zad 5

0.5.1 Delete lines containing NA entries

[30]:
```
data = drop_na(data)
```

[31]:
```
table(data$workclass)
```

```
    Private    self-emp      SL-gov   State-gov Without-pay
      33307        5442        4506        1946          21
```

[32]:
```
head(data)
```

A data.frame: 6 × 15

	age <int>	workclass <chr>	fnlwgt <int>	education <chr>	educational.num <int>	marital.status <chr>	occupa <chr>
1	25	Private	226802	11th	7	Never-married	Machi
2	38	Private	89814	HS-grad	9	Married	Farmi
3	28	SL-gov	336951	Assoc-acdm	12	Married	Protec
4	44	Private	160323	Some-college	10	Married	Machi
5	34	Private	198693	10th	6	Never-married	Other-
6	63	self-emp	104626	Prof-school	15	Married	Prof-s

[33]:
```
data$income = as.character(data$income)
data$income[data$income == "<=50K"] = 0
data$income[data$income == ">50K"] = 1
data['income'] = as.factor(data$income)
```

[34]:
```
data$age = as.numeric(data$age)
data$capital.gain = as.numeric(data$capital.gain)
data$fnlwgt = as.numeric(data$fnlwgt)
data$capital.loss = as.numeric(data$capital.loss)
data$hours.per.week = as.numeric(data$hours.per.week)
```

```
data$educational.num = as.numeric(data$educational.num)
```

[35]:
```
data$gender = as.factor(data$gender)
data$race = as.factor(data$race)
data$education = as.factor(data$education)
data$marital.status = as.factor(data$marital.status)
data$workclass = as.factor(data$workclass)
```

0.6 Zad 6

0.6.1 Divide the data into test and teaching sets

[36]:
```
y = data$income
```

[37]:
```
data_index = createDataPartition(
  y,
  times = 1,
  p = 0.7,
  list = FALSE
)
```

[38]:
```
train_set = data[data_index, ]
test_set = data[-data_index, ]
```

0.7 Zad 7

0.7.1 Build the model using functions (glm)

[39]:
```
formula = income ~ .
```

[40]:
```
model = glm(formula, data = train_set, family = "binomial")
```

```
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
```

[41]:
```
summary(model)
```

```

Call:
glm(formula = formula, family = "binomial", data = train_set)

Coefficients: (1 not defined because of singularities)
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -5.327e+00  6.585e-01  -8.089 6.01e-16
age                                  2.499e-02  1.648e-03  15.168  < 2e-16
workclassself-emp                   -3.038e-01  5.330e-02  -5.699 1.20e-08
workclassSL-gov                      8.265e-02  5.949e-02   1.389 0.164723
workclassState-gov                  -2.972e-01  8.924e-02  -3.330 0.000867
workclassWithout-pay                -4.273e-01  8.134e-01  -0.525 0.599363
```

fnlwgt	7.832e-07	1.699e-07	4.611	4.01e-06
education11th	2.522e-03	2.083e-01	0.012	0.990341
education12th	3.269e-01	2.714e-01	1.205	0.228249
education1st-4th	-1.161e+00	5.987e-01	-1.940	0.052414
education5th-6th	-3.155e-01	3.413e-01	-0.924	0.355245
education7th-8th	-6.499e-01	2.349e-01	-2.767	0.005654
education9th	-4.086e-01	2.691e-01	-1.518	0.128955
educationAssoc-acdm	1.296e+00	1.744e-01	7.433	1.06e-13
educationAssoc-voc	1.148e+00	1.696e-01	6.772	1.27e-11
educationBachelors	1.847e+00	1.576e-01	11.723	< 2e-16
educationDoctorate	2.851e+00	2.190e-01	13.019	< 2e-16
educationHS-grad	7.035e-01	1.532e-01	4.590	4.43e-06
educationMasters	2.152e+00	1.680e-01	12.807	< 2e-16
educationPreschool	-1.525e+00	1.203e+00	-1.268	0.204844
educationProf-school	2.813e+00	2.077e-01	13.544	< 2e-16
educationSome-college	1.061e+00	1.555e-01	6.821	9.04e-12
educational.num	NA	NA	NA	NA
marital.statusNever-married	-7.121e-01	1.236e-01	-5.760	8.40e-09
marital.statusNot-Married	-2.800e-01	1.217e-01	-2.300	0.021437
occupationArmed-Forces	6.079e-02	1.330e+00	0.046	0.963549
occupationCraft-repair	6.250e-03	7.813e-02	0.080	0.936250
occupationExec-managerial	7.754e-01	7.568e-02	10.246	< 2e-16
occupationFarming-fishing	-1.069e+00	1.347e-01	-7.935	2.11e-15
occupationHandlers-cleaners	-7.002e-01	1.347e-01	-5.198	2.01e-07
occupationMachine-op-inspct	-3.528e-01	1.001e-01	-3.524	0.000425
occupationOther-service	-9.444e-01	1.156e-01	-8.169	3.10e-16
occupationPriv-house-serv	-1.411e+00	7.518e-01	-1.877	0.060521
occupationProf-specialty	5.092e-01	7.969e-02	6.390	1.66e-10
occupationProtective-serv	4.161e-01	1.202e-01	3.460	0.000540
occupationSales	2.105e-01	8.099e-02	2.599	0.009349
occupationTech-support	4.884e-01	1.083e-01	4.508	6.55e-06
occupationTransport-moving	-1.809e-01	9.726e-02	-1.860	0.062887
relationshipNot-in-family	-1.487e+00	1.180e-01	-12.599	< 2e-16
relationshipOther-relative	-1.704e+00	2.138e-01	-7.974	1.54e-15
relationshipOwn-child	-2.424e+00	1.737e-01	-13.956	< 2e-16
relationshipUnmarried	-1.634e+00	1.367e-01	-11.950	< 2e-16
relationshipWife	1.243e+00	1.022e-01	12.164	< 2e-16
raceAsian-Pac-Islander	1.139e+00	2.734e-01	4.167	3.08e-05
raceBlack	3.638e-01	2.314e-01	1.572	0.115954
raceOther	4.568e-01	3.476e-01	1.314	0.188779
raceWhite	5.495e-01	2.198e-01	2.500	0.012405
genderMale	8.014e-01	7.728e-02	10.370	< 2e-16
capital.gain	3.207e-04	1.037e-05	30.929	< 2e-16
capital.loss	6.598e-04	3.746e-05	17.612	< 2e-16
hours.per.week	3.109e-02	1.649e-03	18.855	< 2e-16
native.countryCanada	1.537e-01	6.368e-01	0.241	0.809218
native.countryChina	-1.420e+00	6.485e-01	-2.190	0.028509
native.countryColumbia	-2.698e+00	9.913e-01	-2.722	0.006489

```
native.countryCuba                          -3.605e-01  6.702e-01  -0.538 0.590681
native.countryDominican-Republic            -2.416e+00  1.199e+00  -2.015 0.043929
native.countryEcuador                       -2.499e-01  8.523e-01  -0.293 0.769338
native.countryEl-Salvador                   -8.067e-01  7.455e-01  -1.082 0.279228
native.countryEngland                       -1.369e-01  6.606e-01  -0.207 0.835845
native.countryFrance                         2.721e-01  8.042e-01   0.338 0.735125
native.countryGermany                       -2.740e-01  6.350e-01  -0.432 0.666097
native.countryGreece                        -1.030e+00  7.558e-01  -1.363 0.172948
native.countryGuatemala                     -7.545e-01  9.319e-01  -0.810 0.418204
native.countryHaiti                          2.580e-01  8.096e-01   0.319 0.749916
native.countryHoland-Netherlands            -1.060e+01  5.354e+02  -0.020 0.984204
native.countryHonduras                      -9.739e+00  1.305e+02  -0.075 0.940513
native.countryHong                          -1.372e+00  9.064e-01  -1.514 0.130068
native.countryHungary                       -7.293e-02  9.015e-01  -0.081 0.935528
native.countryIndia                         -1.163e+00  6.314e-01  -1.842 0.065420
native.countryIran                          -3.888e-01  7.243e-01  -0.537 0.591431
native.countryIreland                        2.562e-01  8.877e-01   0.289 0.772874
native.countryItaly                          2.599e-01  6.646e-01   0.391 0.695774
native.countryJamaica                       -6.788e-02  7.204e-01  -0.094 0.924929
native.countryJapan                         -1.023e+00  6.807e-01  -1.502 0.132981
native.countryLaos                          -1.381e+00  1.038e+00  -1.330 0.183365
native.countryMexico                        -1.104e+00  6.213e-01  -1.778 0.075451
native.countryNicaragua                     -4.521e-01  9.958e-01  -0.454 0.649830
native.countryOutlying-US(Guam-USVI-etc)    -1.172e+00  1.247e+00  -0.940 0.347233
native.countryPeru                          -1.472e+00  1.019e+00  -1.445 0.148509
native.countryPhilippines                   -3.004e-01  6.059e-01  -0.496 0.620012
native.countryPoland                        -3.515e-01  6.996e-01  -0.502 0.615337
native.countryPortugal                       4.337e-01  7.335e-01   0.591 0.554316
native.countryPuerto-Rico                   -7.600e-01  6.862e-01  -1.108 0.268046
native.countryScotland                      -8.406e-01  1.072e+00  -0.784 0.433142
native.countrySouth                         -2.177e+00  6.948e-01  -3.134 0.001727
native.countryTaiwan                        -1.376e+00  7.158e-01  -1.923 0.054521
native.countryThailand                      -2.206e+00  1.041e+00  -2.120 0.033965
native.countryTrinadad&Tobago               -1.158e+00  1.014e+00  -1.142 0.253296
native.countryUnited-States                 -2.594e-01  5.825e-01  -0.445 0.656117
native.countryVietnam                       -2.597e+00  8.741e-01  -2.971 0.002969
native.countryYugoslavia                     1.151e-01  8.576e-01   0.134 0.893272

(Intercept)                                 ***
age                                         ***
workclassself-emp                           ***
workclassSL-gov
workclassState-gov                          ***
workclassWithout-pay
fnlwgt                                      ***
education11th
education12th
education1st-4th                            .
```

education5th-6th	
education7th-8th	**
education9th	
educationAssoc-acdm	***
educationAssoc-voc	***
educationBachelors	***
educationDoctorate	***
educationHS-grad	***
educationMasters	***
educationPreschool	
educationProf-school	***
educationSome-college	***
educational.num	
marital.statusNever-married	***
marital.statusNot-Married	*
occupationArmed-Forces	
occupationCraft-repair	
occupationExec-managerial	***
occupationFarming-fishing	***
occupationHandlers-cleaners	***
occupationMachine-op-inspct	***
occupationOther-service	***
occupationPriv-house-serv	.
occupationProf-specialty	***
occupationProtective-serv	***
occupationSales	**
occupationTech-support	***
occupationTransport-moving	.
relationshipNot-in-family	***
relationshipOther-relative	***
relationshipOwn-child	***
relationshipUnmarried	***
relationshipWife	***
raceAsian-Pac-Islander	***
raceBlack	
raceOther	
raceWhite	*
genderMale	***
capital.gain	***
capital.loss	***
hours.per.week	***
native.countryCanada	
native.countryChina	*
native.countryColumbia	**
native.countryCuba	
native.countryDominican-Republic	*
native.countryEcuador	
native.countryEl-Salvador	

```
native.countryEngland
native.countryFrance
native.countryGermany
native.countryGreece
native.countryGuatemala
native.countryHaiti
native.countryHoland-Netherlands
native.countryHonduras
native.countryHong
native.countryHungary
native.countryIndia                      .
native.countryIran
native.countryIreland
native.countryItaly
native.countryJamaica
native.countryJapan
native.countryLaos
native.countryMexico                     .
native.countryNicaragua
native.countryOutlying-US(Guam-USVI-etc)
native.countryPeru
native.countryPhilippines
native.countryPoland
native.countryPortugal
native.countryPuerto-Rico
native.countryScotland
native.countrySouth                      **
native.countryTaiwan                     .
native.countryThailand                   *
native.countryTrinadad&Tobago
native.countryUnited-States
native.countryVietnam                    **
native.countryYugoslavia
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35452  on 31655  degrees of freedom
Residual deviance: 20526  on 31566  degrees of freedom
AIC: 20706

Number of Fisher Scoring iterations: 12
```

0.8 Zad 8

0.8.1 Refine the model using the step function

[42]:
```
step_model = step(model)
```

```
Start:  AIC=20705.59
income ~ age + workclass + fnlwgt + education + educational.num +
    marital.status + occupation + relationship + race + gender +
    capital.gain + capital.loss + hours.per.week + native.country


Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"


Step:  AIC=20705.59
income ~ age + workclass + fnlwgt + education + marital.status +
    occupation + relationship + race + gender + capital.gain +
    capital.loss + hours.per.week + native.country


Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
```

```
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"
Warning message:
"glm.fit: fitted probabilities numerically 0 or 1 occurred"

                 Df Deviance   AIC
<none>                 20526 20706
- race            4    20551 20723
- fnlwgt          1    20547 20725
- native.country 40    20638 20738
- marital.status  2    20568 20744
- workclass       4    20574 20746
- gender          1    20636 20814
- age             1    20758 20936
- capital.loss    1    20845 21023
- hours.per.week  1    20890 21068
- occupation     13    21161 21315
- relationship    5    21198 21368
- education      15    21531 21681
- capital.gain    1    22363 22541
```

0.9 Zad 9

0.9.1 Pre-date the model on data (confusion table)

[43]:
```
predictions = predict(step_model, newdata = test_set)
y_hat = ifelse(predictions > 0.5, ">50K", "<=50K")
pred = table(predicted = y_hat, actual = test_set$income)
pred
```

```
         actual
predicted    0    1
    <=50K 9762 1736
```

```
    >50K   442 1626
```

[44]:
```
install.packages("pROC")
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

[45]:
```
library(pROC)
roc_obj = roc(test_set$income, predictions)
roc_obj
```

```
Type 'citation("pROC")' for a citation.


Attaching package: 'pROC'


The following objects are masked from 'package:stats':

    cov, smooth, var


Setting levels: control = 0, case = 1

Setting direction: controls < cases



Call:
roc.default(response = test_set$income, predictor = predictions)

Data: predictions in 10204 controls (test_set$income 0) < 3362 cases
 ↪(test_set$income 1).
Area under the curve: 0.901
```

0.10 Zad 10

0.10.1 Calculate the F1-score for the model developed

[46]:
```
TP <- pred[1,1]
FP <- pred[1,2]
FN <- pred[2,1]
TN <- pred[2,2]

TPR <- TP/(TP+FN)
TNR <- TN/(TN+FP)
PPV <- TP/(TP+FP)
```

```
(F_wynik = 1/(0.5*(1/TPR + 1/PPV)))
# 0.899640586121095
```

0.899640586121095

0.11 Results and summary

0.11.1 pROC

Data: predictions in 10204 controls (test_set$income0) < 3362cases(test_set$income 1).

Area under the curve: 0.901

The high value of area under the curve (0.901) indicates the high quality of the classifier.

The ROC curve shows the relationship between sensitivity (recall) and specificity (1 - percentage of false positives) according to the different thresholds used for classification.

0.11.2 F-score

The F-score is high (0.899640586121095) - it shows that the model's performance is very good and we can predict 89 % of variability.

A higher F-value indicates better model performance in recognising both positive and negative cases. The F-value can range from 0 to 1, where 1 indicates excellent performance and 0 indicates the worst performance.

0.11.3 Summary

Function step helped me to choose the variables that has the best impact on model and leaving only those helped to tune model so that it had great performance.