project375

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Warning: package 'tibble' was built under R version 4.0.3
```

```
## Warning: package 'tidyr' was built under R version 4.0.3
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
## Warning: package 'purrr' was built under R version 4.0.3
```

```
## Warning: package 'dplyr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.3
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(rpart)
```

```
## Warning: package 'rpart' was built under R version 4.0.3
```

```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.0.3
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.0.3
```

```
## randomForest 4.6-14
```

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

```
##
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':
##
##     combine

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

```
library(glmnet) # backward criterion
```

```
## Warning: package 'glmnet' was built under R version 4.0.3

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loaded glmnet 4.0-2
```

```
library(jpeg)   # high-quality plotting
```

```
## Warning: package 'jpeg' was built under R version 4.0.3
```

```
credit <- read_csv("https://docs.google.com/spreadsheets/d/1jFkOKgD5NGeD8mDj_42oBNJfFVK42-1cMKk0JxVFxeA/
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   duration = col_double(),
##   credit_amount = col_double(),
##   installment_commitment = col_double(),
##   residence_since = col_double(),
##   age = col_double(),
##   existing_credits = col_double(),
##   num_dependents = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
glimpse(credit)
```

```
## Rows: 1,000
## Columns: 21
## $ checking_status        <chr> "<0'", "0<=X<200'", "no checking'", "<0'", "...
## $ duration               <dbl> 6, 48, 12, 42, 24, 36, 24, 36, 12, 30, 12, 4...
## $ credit_history         <chr> "critical/other existing credit'", "existing...
## $ purpose                <chr> "radio/tv", "radio/tv", "education", "furnit...
## $ credit_amount          <dbl> 1169, 5951, 2096, 7882, 4870, 9055, 2835, 69...
## $ savings_status         <chr> "no known savings'", "<100'", "<100'", "<100...
## $ employment             <chr> ">=7'", "1<=X<4'", "4<=X<7'", "4<=X<7'", "1<...
## $ installment_commitment <dbl> 4, 2, 2, 2, 3, 2, 3, 2, 2, 4, 3, 3, 1, 4, 2,...
## $ personal_status        <chr> "male single'", "female div/dep/mar'", "male...
## $ other_parties          <chr> "none", "none", "none", "guarantor", "none",...
```

```
## $ residence_since        <dbl> 4, 2, 3, 4, 4, 4, 4, 2, 4, 2, 1, 4, 1, 4, 4,...
## $ property_magnitude     <chr> "real estate'", "real estate'", "real estate...
## $ age                    <dbl> 67, 22, 49, 45, 53, 35, 53, 35, 61, 28, 25, ...
## $ other_payment_plans    <chr> "none", "none", "none", "none", "none", "non...
## $ housing                <chr> "own", "own", "own", "for free'", "for free'...
## $ existing_credits       <dbl> 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1,...
## $ job                    <chr> "skilled", "skilled", "unskilled resident'",...
## $ num_dependents         <dbl> 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ own_telephone          <chr> "yes", "none", "none", "none", "none", "yes"...
## $ foreign_worker         <chr> "yes", "yes", "yes", "yes", "yes", "yes", "y...
## $ class                  <chr> "good", "bad", "good", "good", "bad", "good"...
```

```
credit <-  credit %>% mutate_if(is.character, as.factor)
glimpse(credit)
```

```
## Rows: 1,000
## Columns: 21
## $ checking_status        <fct> <0', 0<=X<200', no checking', <0', <0', no c...
## $ duration               <dbl> 6, 48, 12, 42, 24, 36, 24, 36, 12, 30, 12, 4...
## $ credit_history         <fct> critical/other existing credit', existing pa...
## $ purpose                <fct> radio/tv, radio/tv, education, furniture/equ...
## $ credit_amount          <dbl> 1169, 5951, 2096, 7882, 4870, 9055, 2835, 69...
## $ savings_status         <fct> no known savings', <100', <100', <100', <100...
## $ employment             <fct> >=7', 1<=X<4', 4<=X<7', 4<=X<7', 1<=X<4', 1<...
## $ installment_commitment <dbl> 4, 2, 2, 2, 3, 2, 3, 2, 2, 4, 3, 3, 1, 4, 2,...
## $ personal_status        <fct> male single', female div/dep/mar', male sing...
## $ other_parties          <fct> none, none, none, guarantor, none, none, non...
## $ residence_since        <dbl> 4, 2, 3, 4, 4, 4, 4, 2, 4, 2, 1, 4, 1, 4, 4,...
## $ property_magnitude     <fct> real estate', real estate', real estate', li...
## $ age                    <dbl> 67, 22, 49, 45, 53, 35, 53, 35, 61, 28, 25, ...
## $ other_payment_plans    <fct> none, none, none, none, none, none, none, no...
## $ housing                <fct> own, own, own, for free', for free', for fre...
## $ existing_credits       <dbl> 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1,...
## $ job                    <fct> skilled, skilled, unskilled resident', skill...
## $ num_dependents         <dbl> 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ own_telephone          <fct> yes, none, none, none, none, yes, none, yes,...
## $ foreign_worker         <fct> yes, yes, yes, yes, yes, yes, yes, yes, yes,...
## $ class                  <fct> good, bad, good, good, bad, good, good, good...
```

```
# splitting into training and testing dataset
credit_split_70 = createDataPartition(credit$class, p = 0.7, list = FALSE)
credit_split_80 = createDataPartition(credit$class, p = 0.8, list = F)
training_70 = credit[credit_split_70,]
```

```
## Warning: The `i` argument of ``[`()` can't be a matrix as of tibble 3.0.0.
## Convert to a vector.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
training_80 = credit[credit_split_80,]
test_70_features = credit[-credit_split_70, !(colnames(credit) %in% c('class'))]
test_70_target = credit[-credit_split_70, 'class']
test_80_features = credit[-credit_split_80, !(colnames(credit) %in% c('class'))]
test_80_target = credit[-credit_split_80, 'class']
```

```
# create decision trees
credit_tree_70 = rpart(class ~ . , data = training_70)
credit_tree_80 = rpart(class ~ . , data = training_80)
```

70% training:

```
credit_tree_70$variable.importance
```

```
##        checking_status          credit_amount               duration
##             35.0364974             18.3708206             16.6689073
##                purpose         savings_status         credit_history
##             12.6168167             12.1374316             11.2432345
##             employment        personal_status          other_parties
##              6.3984054              4.9897206              4.2250509
##     property_magnitude       existing_credits                    job
##              3.9038504              2.5251330              1.8349415
##                housing installment_commitment    other_payment_plans
##              1.7911154              1.1409589              0.9109744
##                    age        residence_since        foreign_worker
##              0.6852500              0.3070420              0.1165217
```

```
jpeg(filename="credit_70.jpeg", width=1920, height = 1080)
rpart.plot(credit_tree_70)
dev.off()
```

```
## pdf
##   2
```

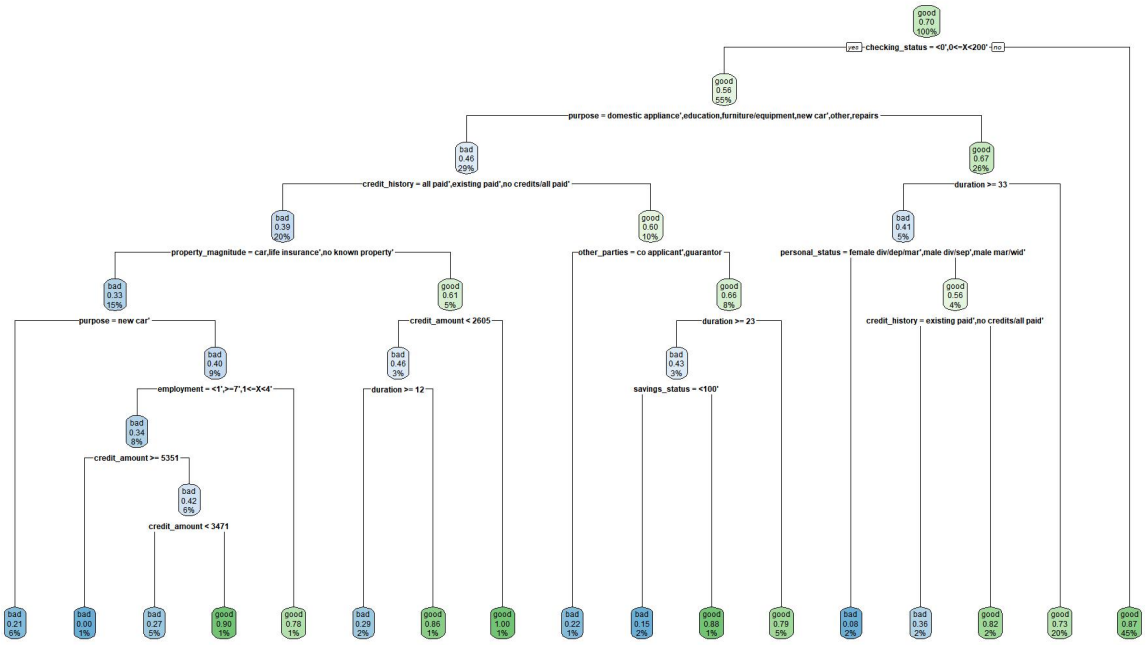

Figure 1: credit_70_tree