

```

library(data.table)
library(caret)
credit_card_raw <- fread("creditcard.csv")

# Create train and test credit_card_rawset
credit_card_raw[, test:=0]
credit_card_raw[, "Time" := NULL]
credit_card_raw[sample(nrow(credit_card_raw), 284807*0.2), test:=1]
test <- credit_card_raw[test==1]
train <- credit_card_raw[test==0]
train[, "test" := NULL]
test[, "test" := NULL]
credit_card_raw[, "test" := NULL]

# Convert credit_card_rawtables to credit_card_rawframes for downsampling
setDF(train)
setDF(test)

# Downsample
set.seed(1)
train$Class <- factor(train$Class)
downsample.train <- downSample(train[, -ncol(train)], train$Class)

test$Class <- factor(test$Class)
downsample.test <- downSample(test[, -ncol(test)], test$Class)

#build decision tree
#apply 5-folds cross validation to find the best parameter cp for decision tree
ctrl <- trainControl(method = "cv", number = 5)

model <- train(Class ~ ., data = downsample.train,
               method = "rpart",
               trControl = ctrl)

model

## CART
##
## 788 samples
## 29 predictor
## 2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 630, 631, 631, 630, 630
## Resampling results across tuning parameters:
##
##   cp          Accuracy   Kappa
## 0.007614213 0.9149722 0.8299509
## 0.019035533 0.9136983 0.8274268
## 0.827411168 0.6561961 0.3136974

```

```
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.007614213.
```

```
#find best cp for decision model
```

```
#the best model is about cp = 0.015
```

```
#evaluate the best model using test data
```

```
pred <- predict(model, downsample.test)
```

```
#performances
```

```
confusionMatrix(pred, downsample.test$Class, positive = "1")
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction  0  1
```

```
##           0 94 12
```

```
##           1  4 86
```

```
##
```

```
##           Accuracy : 0.9184
```

```
##           95% CI : (0.8708, 0.9526)
```

```
## No Information Rate : 0.5
```

```
## P-Value [Acc > NIR] : < 2e-16
```

```
##
```

```
##           Kappa : 0.8367
```

```
##
```

```
## McNemar's Test P-Value : 0.08012
```

```
##
```

```
##           Sensitivity : 0.8776
```

```
##           Specificity : 0.9592
```

```
## Pos Pred Value : 0.9556
```

```
## Neg Pred Value : 0.8868
```

```
## Prevalence : 0.5000
```

```
## Detection Rate : 0.4388
```

```
## Detection Prevalence : 0.4592
```

```
## Balanced Accuracy : 0.9184
```

```
##
```

```
## 'Positive' Class : 1
```

```
##
```