```r
library(readr)
data <- read_csv("creditcard.csv")
table(data$Class)
```

```
##
##      0      1
## 284315    492
```

```r
#imbalance


library(caret)

set.seed(1)
data$Class <- factor(data$Class)
newdata <- upSample(data[,-ncol(data)],
         data$Class
         )

table(newdata$Class)
```

```
##
##      0      1
## 284315 284315
```

```r
#now balanced

#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 = newdata,
               method = "rpart",
               trControl = ctrl)

model
```

```
## CART
##
## 568630 samples
##     30 predictor
##      2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 454904, 454904, 454904, 454904, 454904
## Resampling results across tuning parameters:
##
##   cp           Accuracy   Kappa
##   0.006939486  0.9375165  0.8750330
##   0.013791042  0.9325871  0.8651742
##   0.843339254  0.6682852  0.3365704
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.006939486.
```

```r
#find best cp for decision model

#the best model is about cp = 0.007

#evaluate the best model

pred <- predict(model)

#performances
confusionMatrix(pred, newdata$Class, positive = "1")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 268450  20834
##          1  15865 263481
##
##                Accuracy : 0.9355
##                  95% CI : (0.9348, 0.9361)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8709
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9267
##             Specificity : 0.9442
##          Pos Pred Value : 0.9432
##          Neg Pred Value : 0.9280
##              Prevalence : 0.5000
##          Detection Rate : 0.4634
##    Detection Prevalence : 0.4913
##       Balanced Accuracy : 0.9355
##
##        'Positive' Class : 1
##
```