Classificação de Veículos

Luan Roger Santos Santana

```
## Loading required package: ggplot2
```

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

Preparing data

Loading Data

```
data_raw <- read.csv("../data_sets/Material 03 - 6 - C - Previsao do Tempo - Dados.csv")
data_raw_new_cases <- read.csv("../data_sets/Material 03 - 6 - C - Previsao do Tempo - Dados - Novos Cas
```

Cleaning data

```
data <- data_raw
data_new_cases <- data_raw_new_cases
print(head(data))
```

```
##          Ceu Temperatura Umidade Vento Chovera
## 1 Ensolarado     Elevada    Alta   Sim     SIM
## 2    Chuvoso     Elevada  Normal   Sim     NAO
## 3    Coberto     Elevada    Alta   Nao     NAO
## 4 Ensolarado     Elevada    Alta   Nao     SIM
## 5 Ensolarado     Elevada    Alta   Sim     SIM
## 6    Coberto     Elevada    Alta   Nao     NAO
```

```
print(head(data_new_cases))
```

```
##          Ceu Temperatura Umidade Vento Chovera
## 1 Ensolarado     Elevada    Alta   Sim       ?
## 2    Chuvoso     Elevada  Normal   Sim       ?
## 3    Coberto     Elevada    Alta   Nao       ?
```

Creating data partitioning

```
set.seed(1988)
ran <- sample(1:nrow(data), 0.8 * nrow(data))
training_data <- data[ran,]
test_data <- data[-ran,]
```

Training

Using KNN

Creating the model

```
tuneGrid <- expand.grid(k = c(1,3,5,7,9))
set.seed(1988)
knn <- train(Chovera ~ ., data = training_data, method = "knn", tuneGrid=tuneGrid)
print(knn)
```

```
## k-Nearest Neighbors
##
## 31 samples
##  4 predictor
##  2 classes: 'NAO', 'SIM'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 31, 31, 31, 31, 31, 31, ...
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   1  0.9050390  0.7926744
##   3  0.7773243  0.5096363
##   5  0.6856154  0.3512346
##   7  0.6842218  0.3559695
##   9  0.6930789  0.3756379
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 1.
```

```
prediction.knn <- predict(knn, test_data)
cf_matrix <- confusionMatrix(prediction.knn, as.factor(test_data$Chovera))
print(cf_matrix)
```

Checking the model with training data

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction NAO SIM
##        NAO   5   0
##        SIM   0   3
##
##                Accuracy : 1
##                  95% CI : (0.6306, 1)
##     No Information Rate : 0.625
##     P-Value [Acc > NIR] : 0.02328
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
```

```
##             Sensitivity : 1.000
##             Specificity : 1.000
##          Pos Pred Value : 1.000
##          Neg Pred Value : 1.000
##              Prevalence : 0.625
##          Detection Rate : 0.625
##    Detection Prevalence : 0.625
##       Balanced Accuracy : 1.000
##
##        'Positive' Class : NAO
##
```

Checking for new cases

```
prediction.knn_new_data <- predict(knn, data_new_cases)
data_new_cases$Chovera <- NULL
result <- cbind(data_new_cases, Chovera=prediction.knn_new_data)
print(result)
```

```
##          Ceu Temperatura Umidade Vento Chovera
## 1 Ensolarado     Elevada    Alta   Sim     SIM
## 2    Chuvoso     Elevada  Normal   Sim     NAO
## 3    Coberto     Elevada    Alta   Nao     NAO
```