

# STOR556 Project

Austin Dunn

4/17/2018

```
#Import Tidy data
crimes = read.csv("crimes_tidy.csv")

#Add variable 'night' for crimes that occur between 8pm - 6am
crimes = mutate(crimes, night = 0)
crimes$night <- ifelse( crimes$hour_occur >= 20 | crimes$hour_occur <= 5, 1, 0)

#Randomly divide the data into a training and test set with ratio 3:1
n <- nrow(crimes)
set.seed(123)
shuffle <- sample.int(n)
shuffle <- cut(shuffle, breaks = c(0, quantile(1:n, 0.75), n), labels = c("train", "test"))
crimes.train <- split(crimes, shuffle)$train
crimes.test <- split(crimes, shuffle)$test
```

## Logistic Regression

```
## [1] Confusion Matrix of Logistic Regression
```

```
##      Predict
## Truth FALSE  TRUE  Sum
##  0   13865  1359 15224
##  1    2000   374  2374
##  Sum 15865  1733 17598
```

```
## [1] Proportional Table of Confusion Matrix
```

```
##      Predict
## Truth FALSE  TRUE  Sum
##  0   0.7879 0.0772 0.8651
##  1   0.1136 0.0213 0.1349
##  Sum 0.9015 0.0985 1.0000
```

## LDA

```
## [1] Confusion Matrix of LDA
```

```
##      Predict
## Truth    0    1  Sum
##  0   10332   844 11176
##  1    5471   951  6422
##  Sum 15803  1795 17598
```

```
## [1] Proportional Table of Confusion Matrix
```

```
##      Predict
## Truth    0    1  Sum
##  0   0.587 0.048 0.635
##  1   0.311 0.054 0.365
##  Sum 0.898 0.102 1.000
```

```
## [1] Row-Proportional Table of Confusion Matrix
##      Predict
## Truth    0    1    Sum
##      0 0.9245 0.0755 1.0000
##      1 0.8519 0.1481 1.0000
```

## QDA

```
## [1] Confusion Matrix of QDA
```

```
##      Predict
## Truth    0    1    Sum
##      0  5626  5550 11176
##      1  2268  4154  6422
##      Sum  7894  9704 17598
```

```
## [1] Proportional Table of Confusion Matrix
```

```
##      Predict
## Truth    0    1    Sum
##      0 0.320 0.315 0.635
##      1 0.129 0.236 0.365
##      Sum 0.449 0.551 1.000
```

```
## [1] Row-Proportional Table of Confusion Matrix
```

```
##      Predict
## Truth    0    1    Sum
##      0 0.503 0.497 1.000
##      1 0.353 0.647 1.000
```

## Comparison

Test errors, FPRs and FNRs in all methods above are reported as follows.

	Logistic Regression	LDA	QDA
<b>Test Error</b>	0.191	0.359	0.444
<b>FPR</b>	0.0893	0.0755	0.497
<b>FNR</b>	0.842	0.852	0.353