2017,.pv,219batches,logitwithconfM

Nikhil Muthukrishnan

4 February 2019

The data loaded is of all batches of 2017 (219 batches) with a binary outcome variable "inspec". The batches are represented by mean value of each .pv pi tag which act as predictors for Inspec/Offspec

```
##   SC3_FIC20301.pv InSpec
## 1        81.54993      0
## 2        66.19363      1
## 3        48.98670      1
## 4        42.99971      1
## 5        36.46492      1
## 6        31.42811      1
```

table of dependent variables describing class imbalance

```
##
##   0   1
##  45 174
```

The logistic regression on this data set shows actual offspec equals 45 and actual inspec equal 174. Of which 9 offspec were incorrectly predicted and 3 onspec were incorrectly predicted totalling 12 incorrect predictions or 95% accuracy. However the dim of the data set is 219 rows and 116 coloumns which is obviously overfit.

```
##          Predicted 0 Predicted 1 Total
## Actual 0          36           9    45
## Actual 1           3         171   174
## Total             39         180   219
```