



Models Cross Validation

Sanatan Das

May 24, 2018

Contents

Initial Data Preparation	2
Load the Data	2
Split the Data (training set and test set)	2

Initial Data Preparation

Load the Data

Split the Data (training set and test set)

```
## # A tibble: 18,372 x 20
##   RecordId OPEID  Name    Address    City  State StateDesc ZipCode ZipExt
##   <dbl> <chr>  <fct>  <chr>      <chr> <fct> <chr>      <fct> <chr>
## 1      335 001170 CLAREM~ 500 EAST ~ CLAR~ CA    CALIFORN~ 91711 6400
## 2     10023 022704 SOUTHE~ 2545 VALL~ BIRM~ AL    ALABAMA 35244 2083
## 3      1899 001969 KENTUC~ 3000 FRED~ OWEN~ KY    KENTUCKY 42301 6057
## 4      9062 020788 COLLEC~ 7353 SOUT~ MIDV~ UT    UTAH     84047 3022
## 5     22438 037063 AMERIC~ 5000C COC~ MARG~ FL    FLORIDA 33063 3901
## 6     18447 001785 ANDERS~ 1100 EAST~ ANDE~ IN    INDIANA 46012 3495
## 7     19902 003509 UNIVER~ SOUTHERN ~ MEM~ TN    TENNESSEE 38152 4611
## 8     20830 009192 SIERRA~ 999 TAHOE~ INCL~ NV    NEVADA 89451 0000
## 9      1822 001936 NEOSHO~ 800 WEST ~ CHAN~ KS    KANSAS 66720 2699
## 10    17798 041559 AVEDA ~ 6020 EAST~ INDI~ IN    INDIANA 46250 4746
## # ... with 18,362 more rows, and 11 more variables: ProgLength <fct>,
## #   SchoolType <fct>, Year <chr>, Num <dbl>, Denom <dbl>, Drate <dbl>,
## #   Prate <fct>, EthnicCode <fct>, CongDis <fct>, Region <chr>, Avg <chr>

## # A tibble: 4,593 x 20
##   RecordId OPEID  Name    Address    City  State StateDesc ZipCode ZipExt
##   <dbl> <chr>  <fct>  <chr>      <chr> <fct> <chr>      <fct> <chr>
## 1      4.00 001003 FAULKNE~ 5345 ATL~ MONT~ AL    ALABAMA 36109 3398
## 2      9.00 001004 UNIVERS~ PALMER C~ MONT~ AL    ALABAMA 35115 6000
## 3     11.0 001005 ALABAMA~ 915 SOUT~ MONT~ AL    ALABAMA 36104 5714
## 4     14.0 001007 CENTRAL~ 1675 CHE~ ALEX~ AL    ALABAMA 35010 0000
## 5     15.0 001007 CENTRAL~ 1675 CHE~ ALEX~ AL    ALABAMA 35010 0000
## 6     22.0 001012 BIRMING~ 900 ARKA~ BIRM~ AL    ALABAMA 35254 0002
## 7     23.0 001012 BIRMING~ 900 ARKA~ BIRM~ AL    ALABAMA 35254 0002
## 8     36.0 001019 HUNTING~ 1500 EAS~ MONT~ AL    ALABAMA 36106 2148
## 9     41.0 001022 JEFFERS~ 2601 CAR~ BIRM~ AL    ALABAMA 35215 3098
## 10    44.0 001023 JUDSON ~ 302 BIBB~ MARI~ AL    ALABAMA 36756 2504
## # ... with 4,583 more rows, and 11 more variables: ProgLength <fct>,
## #   SchoolType <fct>, Year <chr>, Num <dbl>, Denom <dbl>, Drate <dbl>,
## #   Prate <fct>, EthnicCode <fct>, CongDis <fct>, Region <chr>, Avg <chr>
```