Naive Bayes Agency Data Predictive Model

Frank Neugebauer

9/17/2019

Agency Data Simple Naive Bayes Predictive Model

Agency data intuitively has value to an agency and potentially to an agent's carrier partner(s). One way to gain value from that data - which exists within agency management systems - is to create one or more predictive models. Gaining insight from predictive modeling using agency data is the fundamental nature of this effort.

This project's objectives are to:

- Understand **if** a predictive model can be created
- Determine the relative value of such a predictive model to an agency
- Explore other methods aside from Naive Bayes for further investigation

The data itself is contained within a codebook, the output of which can be found within the `data` folder of this project. There is also a small data preparation file in the same location.

Epilogue It's clear that predictive modelling with agency has value. The contents of this analysis are a single perspective of such an effort, although subsequently, at least one other perspective is presented to give the reader a glimpse into how varied this analysis could potentially be.

This uses a largely unaltered simple Naive Bayes model to set a baseline for subsequent predictive models, which can be created using other algorithms.

```
# Read the original data set
#agency_data_orig <- read.xlsx('./data/AgencyData_clean.xlsx', sheetIndex=1, stringsAsFactors=T)
source('load_trim_data.R')
```

```
## 'data.frame':    2376 obs. of  12 variables:
##  $ account_type    : Factor w/ 2 levels "Commercial","Personal": 2 2 2 2 2 2 2 2 2 2 ...
##  $ assigned_agent  : Factor w/ 6 levels "Boehm, Sebastian",..: 6 2 2 6 6 6 6 6 6 6 ...
##  $ lob             : Factor w/ 20 levels "Auto (Commercial)",..: 10 10 2 10 2 2 10 2 2 2 ...
##  $ master_company  : Factor w/ 20 levels "Aimonetti Insurance",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ effective_date  : Date, format: "2014-10-07" "2014-10-06" ...
##  $ policy_term     : Factor w/ 3 levels "12 Months","6 Months",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ policy_type     : Factor w/ 2 levels "Commercial","Personal": 2 2 2 2 2 2 2 2 2 2 ...
##  $ annual_premium  : num  291 868 1239 789 1648 ...
##  $ written_premium : num  291 868 1239 789 1648 ...
##  $ rating_state    : Factor w/ 8 levels "FL","ND","NV",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ status          : Factor w/ 2 levels "Active","Inactive": 2 2 2 2 2 2 2 2 2 2 ...
##  $ transaction_type: Factor w/ 13 levels "Cancel Conf",..: 13 13 13 13 13 13 13 13 13 13 ...
```

```
str(agency_data_used)
```

```
## 'data.frame':    2376 obs. of  12 variables:
##  $ account_type    : Factor w/ 2 levels "Commercial","Personal": 2 2 2 2 2 2 2 2 2 2 ...
##  $ assigned_agent  : Factor w/ 6 levels "Boehm, Sebastian",..: 6 2 2 6 6 6 6 6 6 6 ...
##  $ lob             : Factor w/ 20 levels "Auto (Commercial)",..: 10 10 2 10 2 2 10 2 2 2 ...
##  $ master_company  : Factor w/ 20 levels "Aimonetti Insurance",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ effective_date  : Date, format: "2014-10-07" "2014-10-06" ...
##  $ policy_term     : Factor w/ 3 levels "12 Months","6 Months",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ policy_type     : Factor w/ 2 levels "Commercial","Personal": 2 2 2 2 2 2 2 2 2 2 ...
```

```
##  $ annual_premium  : num  291 868 1239 789 1648 ...
##  $ written_premium : num  291 868 1239 789 1648 ...
##  $ rating_state    : Factor w/ 8 levels "FL","ND","NV",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ status          : Factor w/ 2 levels "Active","Inactive": 2 2 2 2 2 2 2 2 2 2 ...
##  $ transaction_type: Factor w/ 13 levels "Cancel Conf",..: 13 13 13 13 13 13 13 13 13 13 ...
```