

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)
```

Data Setup

There are a couple of irrelevant attributes from the original set. The **account_name** is not important in this case, although if the transaction had been more sequential (i.e., from new business through renewal), the name would have matter very much. This is an important consideration for using this with other data. Similarly, the **branch_name** was just the name of the agency and every observation had the same value. If there had been multiple agencies (or branch locations), this feature would have been very important.

With the irrelevant attributes identified, the following gets rid of those and then show the data structure and separate the **train** and **test** data sets.

```
# Trim the phat - i.e., data that's irrelevant
agency_data_used <- agency_data_orig[c(-1, -4)]
```

```
# Show the data structure
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 ...
## $ status : Factor w/ 2 levels "Active","Inactive": 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 ...

# Setup train and test data, 70% / 30%
train_ad <- sample_frac(agency_data_used, 0.7)
sid <- as.numeric(rownames(train_ad)) # because rownames() returns character
test_ad <- agency_data_used[-sid,]
```

Build the Model

This step is set aside because with a big data set, it would be slow or put into a distributed system. This model is using `transaction_type` as the target variable. In simple terms, the model is making predictions for the various transaction types given other conditions, such as the line of business.

The transactions, which are the basic policy lifecycle transactions across lines of business, are:

1. New Business - when a client is first getting a policy (this assumes one or more quotes have been provided)
2. Policy Change - any type of change to a policy, including fixing spellings through endorsements for additional coverage
3. Cancel - there are three versions of this
 - Conf - a confirmed cancellation by the insured
 - Req - a requested cancellation by the insured
 - DNR - a cancellation caused by a non-renewal
4. Renew - when a policy term ends, it goes into a renewal state, which has multiple conditions:
 - Just renew, meaning the policy renewed, which can be automatic in many situations
 - Renewal Quote - this trigger premium calculations for the renewal
 - Renewal Re-Quote - often time, the insured will ask for different options, which results in a re-quote (which can occur multiple times)
5. Non-Renewal - this means the policy term end is the end of the policy altogether and happen when an insurance company no longer wants to cover the insured
6. Re-issue - this happens when a policy is cancelled but then “uncancelled,” a situation most often due to payment problems (that can be the fault of the insurance company) - there is normally a lapse in coverage for a re-issue
7. Re-instate - like re-issue, a re-instatement happens when a cancelled policy is renewed, but occurs with no lapse in coverage
8. Re-write - an insurance policy is a contract and requires endorsements to make even simple changes; if a policy has anything major that requires significant changes, the policy is cancelled and then a re-write is issued with the correction(s)

The situations where predicting which transaction will occur is important. For example, can the model predict the likelihood of a renewal for a personal auto policy (it can go deeper than that, but it’s a good illustration of the principle).

As previously stated, a Naive Bayes model is used. For a quick description of Naive Bayes, please refer to this nice Wikipedia description.

```
nb_model <- naiveBayes(transaction_type~., data=train_ad)
```

Conditional Probabilities

The model creates the conditional probabilities based on the training data. This output shows those, which is interesting unto themselves

Intuitively, it may appear that this is simple math - e.g., if there were 100 transactions and 40 of them were renewals, then the probability should be 40%. However, this model is multi-factored, which means a more accurate way of considering it is, given the line of business was personal auto, there's some non-obvious probability that a renewal will occur. That's the kind of probabilities being reported here.

For the sake of readability, the bulk of the conditional probabilities can be found in the corresponding .rmd and .pdf files within the data directory.

```
nb_model$stables$policy_term
```

##		policy_term		
## Y		12 Months	6 Months	7 Months
##	Cancel Conf	0.55696203	0.44303797	0.00000000
##	Cancel Req	0.42857143	0.57142857	0.00000000
##	Cancel/DNR	0.20000000	0.80000000	0.00000000
##	DB Synch	0.66666667	0.33333333	0.00000000
##	New Business	0.96273292	0.03105590	0.00621118
##	Non-Renewal	1.00000000	0.00000000	0.00000000
##	Policy Change	0.84196185	0.15803815	0.00000000
##	Re-Issue	1.00000000	0.00000000	0.00000000
##	Reinstate	0.78048780	0.21951220	0.00000000
##	Renew Quote	0.00000000	1.00000000	0.00000000
##	Renewal	0.77777778	0.22222222	0.00000000
##	Renewal Re-Quote	1.00000000	0.00000000	0.00000000
##	Rewrite	0.95454545	0.04545455	0.00000000

```
nb_model$stables$policy_type
```

##		policy_type	
## Y		Commercial	Personal
##	Cancel Conf	0.03234880	0.96765120
##	Cancel Req	0.28571429	0.71428571
##	Cancel/DNR	0.00000000	1.00000000
##	DB Synch	0.00000000	1.00000000
##	New Business	0.08695652	0.91304348
##	Non-Renewal	0.00000000	1.00000000
##	Policy Change	0.03269755	0.96730245
##	Re-Issue	0.00000000	1.00000000
##	Reinstate	0.07317073	0.92682927
##	Renew Quote	0.00000000	1.00000000
##	Renewal	0.08771930	0.91228070
##	Renewal Re-Quote	0.00000000	1.00000000
##	Rewrite	0.00000000	1.00000000

```
# create one for binned premium
```

```
nb_model$stables$rating_state
```

##		rating_state			
## Y		FL	ND	NV	OK
##	Cancel Conf	0.995780591	0.001406470	0.000000000	0.000000000
##	Cancel Req	1.000000000	0.000000000	0.000000000	0.000000000
##	Cancel/DNR	1.000000000	0.000000000	0.000000000	0.000000000
##	DB Synch	1.000000000	0.000000000	0.000000000	0.000000000

```
## New Business      0.993788820 0.000000000 0.000000000 0.000000000
## Non-Renewal       1.000000000 0.000000000 0.000000000 0.000000000
## Policy Change     0.989100817 0.000000000 0.000000000 0.010899183
## Re-Issue          1.000000000 0.000000000 0.000000000 0.000000000
## Reinstate         1.000000000 0.000000000 0.000000000 0.000000000
## Renew Quote       1.000000000 0.000000000 0.000000000 0.000000000
## Renewal           0.997076023 0.000000000 0.000000000 0.002923977
## Renewal Re-Quote  1.000000000 0.000000000 0.000000000 0.000000000
## Rewrite           1.000000000 0.000000000 0.000000000 0.000000000
## rating_state
## Y                OR                UT                WA                WY
## Cancel Conf      0.001406470 0.001406470 0.000000000 0.000000000
## Cancel Req       0.000000000 0.000000000 0.000000000 0.000000000
## Cancel/DNR       0.000000000 0.000000000 0.000000000 0.000000000
## DB Synch         0.000000000 0.000000000 0.000000000 0.000000000
## New Business     0.000000000 0.000000000 0.006211180 0.000000000
## Non-Renewal      0.000000000 0.000000000 0.000000000 0.000000000
## Policy Change     0.000000000 0.000000000 0.000000000 0.000000000
## Re-Issue         0.000000000 0.000000000 0.000000000 0.000000000
## Reinstate        0.000000000 0.000000000 0.000000000 0.000000000
## Renew Quote      0.000000000 0.000000000 0.000000000 0.000000000
## Renewal          0.000000000 0.000000000 0.000000000 0.000000000
## Renewal Re-Quote 0.000000000 0.000000000 0.000000000 0.000000000
## Rewrite          0.000000000 0.000000000 0.000000000 0.000000000
```

Some important points about the conditional probabilities:

1. The conditional probabilities by `assigned_agent` and transaction can help in workforce capacity planning. For example, if some agents are handling renewals more frequently, perhaps those agents need help.
2. `master_company` data can be used similarly to how `assigned_agent` is used in order to ensure companies with the most valuable business, or most frequently number of transactions, have the greatest amount of help.
3. For `policy_term` there is something unexpected: virtually all `New Business` transactions are 12 month, whereas renewals are 76%/24% 12 to 6 month. Why?

Make Predictions

As with the model, this step can take some real time with a large data set, so it's segregated from processing / output.

```
nb_predict <- predict(nb_model, test_ad)
```

Show the Predictions

This is similar to the conditional probability, but applies it to show the actual predicted values. The points about conditional probabilities all apply for the predictions, but this is where number of transactions are shown. As such, this is a more valuable workforce capacity planning metric.

However, the numbers shown here are based on the size of the test set. If you want to actually simulate workforce capacity, you'll need a sample that is similar to what you'll expect in a year (a model like this one can be used to make that prediction).

```
table(nb_predict, test_ad$account_type)
```

```
##
```

```
## nb_predict      Commercial Personal
## Cancel Conf      11      667
## Cancel Req        2        1
## Cancel/DNR        0       12
## DB Synch          0        2
## New Business       9        3
## Non-Renewal        0        0
## Policy Change      0        0
## Re-Issue           0        0
## Reinstate          1        4
## Renew Quote        0        0
## Renewal            1        0
## Renewal Re-Quote   0        0
## Rewrite            0        0
```

```
table(nb_predict, test_ad$assigned_agent)
```

```
##
## nb_predict      Boehm, Sebastian Doyle, Walter Insurance, Nichersen
## Cancel Conf      85      289      38
## Cancel Req        0        2        1
## Cancel/DNR        0        0       12
## DB Synch          0        1        1
## New Business       0       11        0
## Non-Renewal        0        0        0
## Policy Change      0        0        0
## Re-Issue           0        0        0
## Reinstate          3        0        0
## Renew Quote        0        0        0
## Renewal            0        0        0
## Renewal Re-Quote   0        0        0
## Rewrite            0        0        0
```

```
##
## nb_predict      Padberg, Tracy Rohan, Elwood Tremblay, Deena
## Cancel Conf        0        0      266
## Cancel Req          0        0        0
## Cancel/DNR          0        0        0
## DB Synch            0        0        0
## New Business         0        1        0
## Non-Renewal          0        0        0
## Policy Change        0        0        0
## Re-Issue             0        0        0
## Reinstate            0        0        2
## Renew Quote          0        0        0
## Renewal              0        0        1
## Renewal Re-Quote     0        0        0
## Rewrite              0        0        0
```

```
table(nb_predict, test_ad$lob)
```

```
##
## nb_predict      Auto (Commercial) Auto (Personal) BOP Liability
## Cancel Conf      3      479      0
## Cancel Req        1        1      1
## Cancel/DNR        0      12      0
```

##	DB Synch	0	2	0
##	New Business	0	0	0
##	Non-Renewal	0	0	0
##	Policy Change	0	0	0
##	Re-Issue	0	0	0
##	Reinstate	1	4	0
##	Renew Quote	0	0	0
##	Renewal	0	0	0
##	Renewal Re-Quote	0	0	0
##	Rewrite	0	0	0
##				
##	nb_predict	BOP Property	Business Owners	Policy Commercial Prpty
##	Cancel Conf	1		2 0
##	Cancel Req	0		0 0
##	Cancel/DNR	0		0 0
##	DB Synch	0		0 0
##	New Business	0		1 0
##	Non-Renewal	0		0 0
##	Policy Change	0		0 0
##	Re-Issue	0		0 0
##	Reinstate	0		0 0
##	Renew Quote	0		0 0
##	Renewal	0		1 0
##	Renewal Re-Quote	0		0 0
##	Rewrite	0		0 0
##				
##	nb_predict	Dwelling fire	Flood Genl Liability	Homeowners
##	Cancel Conf	6	0	5 174
##	Cancel Req	0	0	0 0
##	Cancel/DNR	0	0	0 0
##	DB Synch	0	0	0 0
##	New Business	0	0	8 3
##	Non-Renewal	0	0	0 0
##	Policy Change	0	0	0 0
##	Re-Issue	0	0	0 0
##	Reinstate	0	0	0 0
##	Renew Quote	0	0	0 0
##	Renewal	0	0	0 0
##	Renewal Re-Quote	0	0	0 0
##	Rewrite	0	0	0 0
##				
##	nb_predict	Inland marine (comm)	Inland Marine (pers)	
##	Cancel Conf	0		1
##	Cancel Req	0		0
##	Cancel/DNR	0		0
##	DB Synch	0		0
##	New Business	0		0
##	Non-Renewal	0		0
##	Policy Change	0		0
##	Re-Issue	0		0
##	Reinstate	0		0
##	Renew Quote	0		0
##	Renewal	0		0
##	Renewal Re-Quote	0		0

```
## Rewrite 0 0
##
## nb_predict Personal Liability Professional Liab Renters Signs
## Cancel Conf 0 0 0 0
## Cancel Req 0 0 0 0
## Cancel/DNR 0 0 0 0
## DB Synch 0 0 0 0
## New Business 0 0 0 0
## Non-Renewal 0 0 0 0
## Policy Change 0 0 0 0
## Re-Issue 0 0 0 0
## Reinstate 0 0 0 0
## Renew Quote 0 0 0 0
## Renewal 0 0 0 0
## Renewal Re-Quote 0 0 0 0
## Rewrite 0 0 0 0
```

```
##
## nb_predict Umbrella - Comm Umbrella - Personal
## Cancel Conf 2 5
## Cancel Req 0 0
## Cancel/DNR 0 0
## DB Synch 0 0
## New Business 0 0
## Non-Renewal 0 0
## Policy Change 0 0
## Re-Issue 0 0
## Reinstate 0 0
## Renew Quote 0 0
## Renewal 0 0
## Renewal Re-Quote 0 0
## Rewrite 0 0
```

```
##
## nb_predict Watercraft (small boat) Workers comp
## Cancel Conf 0 0
## Cancel Req 0 0
## Cancel/DNR 0 0
## DB Synch 0 0
## New Business 0 0
## Non-Renewal 0 0
## Policy Change 0 0
## Re-Issue 0 0
## Reinstate 0 0
## Renew Quote 0 0
## Renewal 0 0
## Renewal Re-Quote 0 0
## Rewrite 0 0
```

```
table(nb_predict, test_ad$master_company)
```

```
##
## nb_predict Aimonetti Insurance Beulah General Agency
## Cancel Conf 209 0
## Cancel Req 2 0
## Cancel/DNR 0 0
## DB Synch 0 0
```

##	New Business	0	0
##	Non-Renewal	0	0
##	Policy Change	0	0
##	Re-Issue	0	0
##	Reinstate	3	0
##	Renew Quote	0	0
##	Renewal	0	0
##	Renewal Re-Quote	0	0
##	Rewrite	0	0
##			
##	nb_predict	Beaulah Insurance	Beyer Strategic Insurance
##	Cancel Conf	63	37
##	Cancel Req	0	0
##	Cancel/DNR	0	0
##	DB Synch	0	0
##	New Business	0	0
##	Non-Renewal	0	0
##	Policy Change	0	0
##	Re-Issue	0	0
##	Reinstate	0	0
##	Renew Quote	0	0
##	Renewal	0	0
##	Renewal Re-Quote	0	0
##	Rewrite	0	0
##			
##	nb_predict	Borer Group, LLC	Brakus Fire & Casualty
##	Cancel Conf	0	2
##	Cancel Req	0	0
##	Cancel/DNR	0	0
##	DB Synch	0	0
##	New Business	0	0
##	Non-Renewal	0	0
##	Policy Change	0	0
##	Re-Issue	0	0
##	Reinstate	0	0
##	Renew Quote	0	0
##	Renewal	0	0
##	Renewal Re-Quote	0	0
##	Rewrite	0	0
##			
##	nb_predict	Breitenberg Insurance	DuBuque Insurance
##	Cancel Conf	133	63
##	Cancel Req	0	0
##	Cancel/DNR	12	0
##	DB Synch	1	0
##	New Business	0	1
##	Non-Renewal	0	0
##	Policy Change	0	0
##	Re-Issue	0	0
##	Reinstate	0	2
##	Renew Quote	0	0
##	Renewal	0	0
##	Renewal Re-Quote	0	0
##	Rewrite	0	0


```

##
## nb_predict      Friesen Insurance Helen Herman Group Hui Insurance
## Cancel Conf      33              2              1
## Cancel Req        0              1              0
## Cancel/DNR        0              0              0
## DB Synch          0              0              0
## New Business      0              0              3
## Non-Renewal       0              0              0
## Policy Change     0              0              0
## Re-Issue          0              0              0
## Reinstate         0              0              0
## Renew Quote       0              0              0
## Renewal           0              0              0
## Renewal Re-Quote  0              0              0
## Rewrite           0              0              0
##
## nb_predict      Kilback Excess & Surplus Roderick Insurance
## Cancel Conf      0              4
## Cancel Req        0              0
## Cancel/DNR        0              0
## DB Synch          0              0
## New Business      0              0
## Non-Renewal       0              0
## Policy Change     0              0
## Re-Issue          0              0
## Reinstate         0              0
## Renew Quote       0              0
## Renewal           0              0
## Renewal Re-Quote  0              0
## Rewrite           0              0
##
## nb_predict      Ryan Hyatt Insurance Schimel Mutual Insurance
## Cancel Conf      8              8
## Cancel Req        0              0
## Cancel/DNR        0              0
## DB Synch          0              0
## New Business      0              8
## Non-Renewal       0              0
## Policy Change     0              0
## Re-Issue          0              0
## Reinstate         0              0
## Renew Quote       0              0
## Renewal           0              1
## Renewal Re-Quote  0              0
## Rewrite           0              0
##
## nb_predict      Schuppe Group Simonis Insurance Spinka West Insurance
## Cancel Conf      53              42              15
## Cancel Req        0              0              0
## Cancel/DNR        0              0              0
## DB Synch          1              0              0
## New Business      0              0              0
## Non-Renewal       0              0              0
## Policy Change     0              0              0

```

```
## Re-Issue 0 0 0
## Reinstate 0 0 0
## Renew Quote 0 0 0
## Renewal 0 0 0
## Renewal Re-Quote 0 0 0
## Rewrite 0 0 0
```

```
##
## nb_predict Streich Group Insurance Sultan Insurance
## Cancel Conf 5 0
## Cancel Req 0 0
## Cancel/DNR 0 0
## DB Synch 0 0
## New Business 0 0
## Non-Renewal 0 0
## Policy Change 0 0
## Re-Issue 0 0
## Reinstate 0 0
## Renew Quote 0 0
## Renewal 0 0
## Renewal Re-Quote 0 0
## Rewrite 0 0
```

```
#table(nb_predict, test_ad$effective_date) # bin this
table(nb_predict, test_ad$policy_term)
```

```
##
## nb_predict 12 Months 6 Months 7 Months
## Cancel Conf 387 291 0
## Cancel Req 2 1 0
## Cancel/DNR 0 12 0
## DB Synch 2 0 0
## New Business 12 0 0
## Non-Renewal 0 0 0
## Policy Change 0 0 0
## Re-Issue 0 0 0
## Reinstate 3 2 0
## Renew Quote 0 0 0
## Renewal 1 0 0
## Renewal Re-Quote 0 0 0
## Rewrite 0 0 0
```

```
table(nb_predict, test_ad$policy_type)
```

```
##
## nb_predict Commercial Personal
## Cancel Conf 13 665
## Cancel Req 2 1
## Cancel/DNR 0 12
## DB Synch 0 2
## New Business 9 3
## Non-Renewal 0 0
## Policy Change 0 0
## Re-Issue 0 0
## Reinstate 1 4
## Renew Quote 0 0
```

```
## Renewal 1 0
## Renewal Re-Quote 0 0
## Rewrite 0 0
```

```
table(nb_predict, test_ad$rating_state)
```

```
##
## nb_predict FL ND NV OK OR UT WA WY
## Cancel Conf 675 1 0 0 0 1 1 0 0
## Cancel Req 3 0 0 0 0 0 0 0 0
## Cancel/DNR 12 0 0 0 0 0 0 0 0
## DB Synch 2 0 0 0 0 0 0 0 0
## New Business 12 0 0 0 0 0 0 0 0
## Non-Renewal 0 0 0 0 0 0 0 0 0
## Policy Change 0 0 0 0 0 0 0 0 0
## Re-Issue 0 0 0 0 0 0 0 0 0
## Reinstate 5 0 0 0 0 0 0 0 0
## Renew Quote 0 0 0 0 0 0 0 0 0
## Renewal 1 0 0 0 0 0 0 0 0
## Renewal Re-Quote 0 0 0 0 0 0 0 0 0
## Rewrite 0 0 0 0 0 0 0 0 0
```

```
table(nb_predict, test_ad$status)
```

```
##
## nb_predict Active Inactive
## Cancel Conf 0 678
## Cancel Req 0 3
## Cancel/DNR 0 12
## DB Synch 0 2
## New Business 0 12
## Non-Renewal 0 0
## Policy Change 0 0
## Re-Issue 0 0
## Reinstate 0 5
## Renew Quote 0 0
## Renewal 0 1
## Renewal Re-Quote 0 0
## Rewrite 0 0
```

Confusion Matrix

Build a simple confusion matrix and then plot it. Despite its name, there's no confusion within a confusion matrix, which is really more about understanding the error rate for statistical classification (which is what this model does). In fact, the confusion matrix is also called an error matrix.

For more, you can read this [Wikipedia description](#) of confusion matrix.

```
confusion_matrix <- table(nb_predict, test_ad$transaction_type)
```

Finally, Show Stats on the Confusion Matrix

```
## Confusion Matrix and Statistics
##
##
## nb_predict Cancel Conf Cancel Req Cancel/DNR DB Synch New Business
```

```

## Cancel Conf          678          0          0          0          0
## Cancel Req           3          0          0          0          0
## Cancel/DNR          12          0          0          0          0
## DB Synch             2          0          0          0          0
## New Business         12          0          0          0          0
## Non-Renewal          0          0          0          0          0
## Policy Change        0          0          0          0          0
## Re-Issue             0          0          0          0          0
## Reinstate            5          0          0          0          0
## Renew Quote          0          0          0          0          0
## Renewal              1          0          0          0          0
## Renewal Re-Quote     0          0          0          0          0
## Rewrite              0          0          0          0          0
##
## nb_predict           Non-Renewal Policy Change Re-Issue Reinstate
## Cancel Conf          0          0          0          0
## Cancel Req           0          0          0          0
## Cancel/DNR           0          0          0          0
## DB Synch             0          0          0          0
## New Business         0          0          0          0
## Non-Renewal          0          0          0          0
## Policy Change        0          0          0          0
## Re-Issue             0          0          0          0
## Reinstate            0          0          0          0
## Renew Quote          0          0          0          0
## Renewal              0          0          0          0
## Renewal Re-Quote     0          0          0          0
## Rewrite              0          0          0          0
##
## nb_predict           Renew Quote Renewal Renewal Re-Quote Rewrite
## Cancel Conf          0          0          0          0          0
## Cancel Req           0          0          0          0          0
## Cancel/DNR           0          0          0          0          0
## DB Synch             0          0          0          0          0
## New Business         0          0          0          0          0
## Non-Renewal          0          0          0          0          0
## Policy Change        0          0          0          0          0
## Re-Issue             0          0          0          0          0
## Reinstate            0          0          0          0          0
## Renew Quote          0          0          0          0          0
## Renewal              0          0          0          0          0
## Renewal Re-Quote     0          0          0          0          0
## Rewrite              0          0          0          0          0
##
## Overall Statistics
##
## Accuracy : 0.9509
## 95% CI : (0.9324, 0.9656)
## No Information Rate : 1
## P-Value [Acc > NIR] : 1
##
## Kappa : 0
##
## McNemar's Test P-Value : NA

```

```

##
## Statistics by Class:
##
##          Class: Cancel Conf Class: Cancel Req
## Sensitivity          0.9509          NA
## Specificity          NA          0.995792
## Pos Pred Value      NA          NA
## Neg Pred Value      NA          NA
## Prevalence          1.0000          0.000000
## Detection Rate      0.9509          0.000000
## Detection Prevalence 0.9509          0.004208
## Balanced Accuracy    NA          NA
##
##          Class: Cancel/DNR Class: DB Synch Class: New Business
## Sensitivity          NA          NA          NA
## Specificity          0.98317      0.997195      0.98317
## Pos Pred Value      NA          NA          NA
## Neg Pred Value      NA          NA          NA
## Prevalence          0.00000      0.000000      0.00000
## Detection Rate      0.00000      0.000000      0.00000
## Detection Prevalence 0.01683      0.002805      0.01683
## Balanced Accuracy    NA          NA          NA
##
##          Class: Non-Renewal Class: Policy Change
## Sensitivity          NA          NA
## Specificity          1          1
## Pos Pred Value      NA          NA
## Neg Pred Value      NA          NA
## Prevalence          0          0
## Detection Rate      0          0
## Detection Prevalence 0          0
## Balanced Accuracy    NA          NA
##
##          Class: Re-Issue Class: Reinstate Class: Renew Quote
## Sensitivity          NA          NA          NA
## Specificity          1          0.992987      1
## Pos Pred Value      NA          NA          NA
## Neg Pred Value      NA          NA          NA
## Prevalence          0          0.000000      0
## Detection Rate      0          0.000000      0
## Detection Prevalence 0          0.007013      0
## Balanced Accuracy    NA          NA          NA
##
##          Class: Renewal Class: Renewal Re-Quote Class: Rewrite
## Sensitivity          NA          NA          NA
## Specificity          0.998597      1          1
## Pos Pred Value      NA          NA          NA
## Neg Pred Value      NA          NA          NA
## Prevalence          0.000000      0          0
## Detection Rate      0.000000      0          0
## Detection Prevalence 0.001403      0          0
## Balanced Accuracy    NA          NA          NA

```

Results Interpretation

The predictive quality for agency data shows promise. Some very specific points about the results:

1. A 95.09% accuracy for `transaction_type` is very good. While it doesn't guarantee anything, this value is baseline for the usefulness of the model. This model should be checked over time against reality to

see if it remains valid.

2. There's a 95% chance that the accuracy is between 93.24% and 96.56% - also very good. The name - confidence interval - is well done and indicates how likely the accuracy is going to be correct. Do ****not*** take this to mean there are guarantees - save one: the model will make the wrong predictions sometimes, but the right predictions far more often.
3. The accuracy is greater than the No Information Rate, which indicates significance. (Not always, but the upper CI is.)

It's also very much worth noting that if you run this model over and over, you'll get slightly different (albeit similar) results. This is because the predictions vary since the underlying logic re-runs the entire set of logic (including separating the test and train data). This is normal and expected, but don't just pick the best results as **the** results - the best results are just **a** result.

Using the Model - Out of Sample

When you know all the data and are making predictions, that's useful only to understand the viability of the model. The application of the model is for data that's new so you can understand (more) about what's going to happen. That's a true prediction. To do that, a dummy record is created (which could be a real record or some guesses at what you'll have over the next week, month, etc.) and then fed into the model.

```
new_data = data.frame(  
  'account_type' = 'Personal',  
  'assigned_agent' = 'Walter Doyle',  
  'lob' = 'Auto (Personal)',  
  'master_company' = 'Beaulah Insurance',  
  'effective_date' = as.Date('2014-07-29'),  
  'policy_term' = '12 Months',  
  'policy_type' = 'Personal',  
  'annual_premium' = 1234.33,  
  'written_premium' = 993.23,  
  'rating_state' = 'FL',  
  'status' = 'Active'  
)  
  
single_model <- predict(nb_model, newdata=new_data)  
table(single_model)
```

```
## single_model  
##      Cancel Conf      Cancel Req      Cancel/DNR      DB Synch  
##           0           0           0           0  
##      New Business      Non-Renewal      Policy Change      Re-Issue  
##           0           0           0           0  
##      Reinstate      Renew Quote      Renewal Renewal Re-Quote  
##           0           0           1           0  
##      Rewrite  
##           0
```

The transaction with the 1 is what the model predicted the transaction would be. You can do this with large sets of data to understand (for example) how many underwriters are required in Florida.

Areas for Improvement

While the model is quite accurate, it can still be improved. For example, the premium values are skewed and could be transformed using \log_{10} (I did this just to see what would happen and the distribution becomes almost normal). The target variable can also be changed, which will likely yield a different view of the data.

References

- [1] O’Neil, Cathy and Schutt, Rachel. (2014). Doing Data Science. Sebastopol, CA. O’Reilly Media.
- [2] Lander, Jared P. (2017). R for Everyone: Advanced Analytics and Graphics. Hoboken, NJ. Pearson Education as Addison-Wesley.
- [3] James, Gareth, Witten, Daniela, Hastie, Trevor, and Tibshirani, Robert. (2015). New York, NY. Springer.
- [4] Abbott, Dean. (2014). Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst. Indianapolis, IN. John Wiley & Sons.