Text Analytics in the Insurance World:
Discussion on the Benefits of Going Beyond Traditional Tools

Frédérick Guillot, M.Sc.
Text Analytics World – Chicago – June 22nd, 2016
Agenda

Current State

Challenges

Our Approach
Current State

Prior the project, we were heavily relying on **structured** data

- **Asked to clients**
  - Gender
  - Date of birth
  - City

- **Policy / Claims Transactions**
  - Coverages
  - Premiums
  - Losses

- **External Data**
  - Census
  - Credit score
  - Gov’t
Current State

No exploitation of our unstructured data

CRM

“Amy just married and will be moving out shortly... follow up in 2 months to confirm status.”

Claims

“Due to the severe flooding this time the water almost reached the furnace - approx 3 feet of water.”
CRM

“Amy just married and will be moving out shortly... follow up in 2 months to confirm status.”

Claims

“Due to the severe flooding this time the water almost reached the furnace - approx 3 feet of water.”
Current State

Missed opportunities in not mining our unstructured data!

Closing predictors | X-sell opportunities | Churn proxies
Challenges

Traditional tools & techniques not completely identify actionable topics in our text fields :/
Challenges

A lot of un-actionable stuff

“TRIED TO CALL NO ANSWERE AND GOES STRAIGHT TO VOICEMAIL. LM AND I HAVE TRIED TO EMAIL BUT THAT CAME BACK UNDELIVERABLE.”

“SPKE TO JIM, NEED TO F/U WITH RICKI TO SCH. APPOINT. AM BETWEEN 9-12NOON.”
Challenge
Insurance-specific abbreviations / terminology

- Ex-Dates
- Leads
- Med Rehab
- QT
- TPL
- liab
- comp
- NF
- BI
- Pol #
- xsell
- coll
- AB
- Veh sub
### Challenges

Many numbers with various meanings

<table>
<thead>
<tr>
<th>Dates</th>
<th>Aug 11</th>
<th>September 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Money</td>
<td>$173.94</td>
<td>173,94$</td>
</tr>
<tr>
<td>Coverages</td>
<td>500 coll / 300 comp</td>
<td>2 mill liab</td>
</tr>
</tbody>
</table>
Our Approach

Intuitive topics → Initial Rules → Read Identified Notes → Final Rules

Improve Rules ↔ New Topics
Our Approach

- Business discussions
- Traditional text mining techniques
  - N-Grams
  - Latent Dirichlet allocation (LDA)
Our Approach

N-Grams: Basic actionable topics

- Collect
- Premium
- Policy
- New
- Baby
- Happy
- Sold
- Home
- Auto
- Change
- NSF
- Date
- Coverage
- Insurance
- Renewal
- Non payment
- Cancellation
- Effective
- Client
- Frustred
- Error
Our Approach

LDA : Found lot of customers upset about our oil tank rules

“I explained that the co-operators does not insure 14 Gauge oil tanks older than 10 years.”

“Mrs Orton upset about request, stated he has petrol Canada technician service the oil tank every year.”

“Client called re letter, he is very upset that co-operators will not be renewing his home policy.”

“I explained our concerns with older oil tanks and the cost of clean up. William did not want to listen to anything I was trying to tell him and hung up.”
Our Approach

- Start with intuitive keywords
  Include / exclude type
- Augment with algos such as Word2Vec
Our Approach

- **Read Identified Notes**
  - Read & classify 10,000+ notes
  - Found new topics
  - Refined rules

- **New Topics**
- **Improve Rules**
  - Intuitive topics
  - Initial Rules
  - Final Rules

Summer students
- Read & classify 10,000+ notes
- Found new topics
- Refined rules
Our Approach

CLIENT HAPPY - 2,505 Identified

IN = ['happy', 'satisfied', 'glad', 'appreciate', 'thoroughly', ...]

OUT = ['upset', 'not happy', 'isnt happy', 'no happy', 'complain', ...]

“HE SAID HE REALLY APPRECIATED IT. I ASKED IF HE WANTED ME TO CALL HIM BACK IN ABOUT SIX MONTHS TO BOOKED HE SAID THAT WOULD NICE.”
Our Approach

CLIENT UNSATISFIED – 1,764 Identified

IN = ['upset', 'frustrated', 'unhappy', 'complain', 'angry', 'rude', ...]

OUT = []

“very upset that we keep calling to her heras her and doesn't feel the house needs to be inspected.”
Our Approach

PRICE TOO HIGH – 13,003 Identified

IN = ['reduce premium', 'cost down', 'save money', 'expensive', ...]

OUT = ['documents saved', 'full coverage', 'sold', 'satisfied', ...]

“Client advised she could not afford that amount as she was on a pension. Gave reports to take home and review and consider lower coverage amount.”
Our Approach

IN = ['child', 'daughter', 'son', 'birth', 'pregnancy', 'born', 'baby', ...]

OUT = ['daughter in law', 'son in law', 'date of birth', ...]

NEW BABY – 10,083 Identified

“Jean is expecting a baby - end of Feb. E-mailed him re: life insurance.”
Our Approach

NEW MORTGAGE – 17,139 Identified

IN = ['mortgage', 'lender', 'creditor', 'credit union', 'remortgage', ... ]

OUT = ['car', 'auto', 'vehicle', 'motorcycle', 'jewellery', 'boat', ... ]

“clients have remortgaged house and are looking to increase coverages from $150,000 to $250,000.”
Conclusion

• Traditional tools was not enough

• “Hands on” work needed

• Great results with low tech (Python, R, ...)