How Artificial Intelligence (AI) Has Transformed the Insurance Industry

Summer National Meetings // NAIC - CIPR
New York City, NY // 5th August, 2019

https://www.halosinsurance.com/
https://www.datarobot.com/
HOW do we build AI
WHY do we trust AI
Satadru Sengupta
Founder & CEO at Halos

- built multi-million dollar insurance business from pre-revenue at AI pioneer DataRobot: 2015-2018
- built & operationalized multiple AI applications at AIG
- formerly, actuarial data scientist at Liberty Mutual & Deloitte
- CSPA designee Casualty Actuarial Society
HOW do we build an AI application

let's use an example to understand the key concepts and workflow
What is AI?

computer systems able to do tasks that require human intelligence
Problem statement
predict the "likelihood of fraud" of an incoming claim based on policy data and claims data at FNOL

Scope of use:
to be used to triage claims and help SIU in targeted investigation

Key Objectives:
- accurate predictions
- prediction explanations
WHY do we need a computer to do a human task

- processing large, unstructured data
- an objective way of making decision
- fast and automated
<table>
<thead>
<tr>
<th>ID</th>
<th>FRAUD</th>
<th>DISTINCT_PARTIES_ONCLAIM</th>
<th>CLM_AFTER_RNWL</th>
<th>CLAIM_DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0 this via others themselves inc become within ours slow parking lot f</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0 would less bottom de what then find cry motorbike breaks van sudd</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>0 indeed none you to somehow call whereas anyhow driving left sch</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0 am not fire same now over whence therein right left not indicating c</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0 formerly by fifteen again are please four bottom caravan motorbike</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0 nor put see not seems serious is herself motorbike caravan parking</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0 not others into who its these else during car sun right school driving</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0 describe except yourself what whom every because within slow ma</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0 more being third us part but found neither not indicating windscreer</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0 would couldnt etc or wherever her may this carpark van sun parking</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0 have co further three cant found whereafter nevertheless mall round</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0 cant still front among whom wherein serious part not indicating rour</td>
</tr>
</tbody>
</table>
HOW do we teach the computer?

1. A historical dataset

2. Machine learning algorithms
The ability to learn from the past to predict the future without being explicitly programmed.

(a) Clear, two-feature condition (CLEAR-2).
Same dataset but a different algorithm, more complex this time

(c) Clear, eight-feature condition (CLEAR-8).
Algorithms matter

If the model is inaccurate, there could be terrible outcome: bad customer experience to insolvencies
GREAT NEWS!!!
We have explanation tools.
These tools are algorithm agnostic.

- Most impactful features
- Directionality of the feature
- Explain every prediction
Feature importance: which predictors drive the model performance

A few top predictors: # prior claims, claim type, # people involved
EXPLAINABLE AI

Directionality of the feature:
How a predictor influences overall outcome

# prior claims:
3 or more prior claims ⇒ higher chances of fraud
EXPLAINABLE AI

Prediction explanation (AI storytelling):

Going forward, once implemented, we can tell what are the factors behind a prediction

A male claimant reported a theft involving 8 people within a month of buying the policy. He had 4 claims in the last 5 years.
We generated these explanation from a fairly complex model: XGBoost: a very complex and powerful algorithm
Few things that you need to build an AI application

Team:
- data engineers
- data scientists
- domain experts
- users (in this case, fraud/ SIU analyst)

A dataset:
We are using a dataset with 10,000 claims from past with fraud indicator and 40 possible predictors (aka: features)

An AI platform:
we are using DataRobot Automated Machine Learning platform
**DATASET**

**What happened in the past**

Many historical examples...

**Features or variables accompanied the historical outcome**

<table>
<thead>
<tr>
<th>ID</th>
<th>FRAUD</th>
<th>DISTINCT_PARTIES_ON_CLAIM</th>
<th>CLM_AFTER_RNWL</th>
<th>CLAIM_DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>this via others themselves inc become within ours slow parking lot fast vehicle</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>would less bottom de what then find cry motorbike brakes van suddenly not inc</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>21</td>
<td>indeed none you to somehow call whereas anyhow driving left school motorbike</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>am not fire same now over whence therein right left not indicating car carpark s</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>formerly by fifteen again are please four bottom caravan motorbike not indicating</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>nor put see not seems serious is herself motorbike caravan parking lot car right</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>not others into who its these else during car sun right school driving not indicating</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>describe except yourself what whom every because within slow mall vehicle with</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>more being third us part but found neither not indicating windscren vehicle brake</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>would couldnt etc or wherever her may this carpark van sun parking lot slow left</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>have can further three cant found whereafter nevertheless mall roundabout stop</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>cant still front among whom wherein serious part not indicating roundabout can</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>formerly rather it but might former neither done mall roundabout brakes fast inc</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>become no being throughout someone twelve part whole motorbike slow round</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>without among each none system who many well vehicle right slow left school</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>two its was already in this somehow fifty school carpark parking lot indicating round</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>though too full no take together a seem parking lot vehicle caravan windscren</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>due describe hundred therefore became bottom others so vehicle fast brakes on</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>under whence co only therefore eg no around sun parking lot motorbike school</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>six whereupon please nothing interest noone often several slow stopped caravan</td>
</tr>
</tbody>
</table>
Data Science Iron Man
The AI Life Cycle

1. Define Project Objectives
2. Acquire & Explore Data
3. Model Data
4. Interpret & Communicate
5. Implement, Document & Maintain
The AI Life Cycle

1. **Define Project Objectives**
   - Specify business problem
   - Acquire subject matter expertise
   - Define unit of analysis and prediction target
   - Prioritize modeling criteria
   - Consider risks and success criteria
   - Decide whether to continue

2. **Acquire & Explore Data**
   - Find appropriate data
   - Merge data into single table
   - Conduct exploratory data analysis
   - Find and remove any target leakage
   - Feature engineering

3. **Model Data**
   - Variable selection
   - Build candidate models
   - Model validation and selection

4. **Interpret & Communicate**
   - Interpret model
   - Communicate model insights

5. **Implement, Document & Maintain**
   - Set up batch or API prediction system
   - Document modeling process for reproducibility
   - Create model monitoring and maintenance plan
Why Should We TRUST Artificial Intelligence (AI)
“Sometimes attaining the deepest familiarity with a question is our best substitute for actually having the answer.”

BRIAN GREENE
THEORETICAL PHYSICIST & MATHEMATICIAN, COLUMBIA UNIVERSITY
learned the right lessons
was built correctly
made the right decision
has an ethical purpose
remains healthy
is not unfairly biased
“Man is fallible, but maybe men are less so.”

Atul Gawande
Satadru Sengupta

email: satadru@halosinsurance.com
text/ call: +1.617.301.2471

https://www.datarobot.com/
https://www.halosinsurance.com/