Text Analytics & Sentiment Analysis

Alex Link, Application Engineer
Big Picture

Unstructured Data
- Reports
- Social Media

Structured Data
- Database
- In-house Spreadsheets

Text Analytics

Traditional Data Analytics

Decision
### Where is this data?

#### Data Sources for Big Data Analysis

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Currently Analyzing</th>
<th>Planning to Analyze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transactions</td>
<td>67%</td>
<td>22%</td>
</tr>
<tr>
<td>Log Data (Including Web)</td>
<td>56%</td>
<td>26%</td>
</tr>
<tr>
<td>Geospatial and Location</td>
<td>38%</td>
<td>34%</td>
</tr>
<tr>
<td>Social Media Profile</td>
<td>27%</td>
<td>38%</td>
</tr>
<tr>
<td>Social Media Chat and Interaction</td>
<td>23%</td>
<td>42%</td>
</tr>
<tr>
<td>E-Mails and Documents</td>
<td>24%</td>
<td>38%</td>
</tr>
<tr>
<td>Open Data</td>
<td>32%</td>
<td>30%</td>
</tr>
<tr>
<td>Data from Data Brokers</td>
<td>34%</td>
<td>24%</td>
</tr>
<tr>
<td>Internet of Things</td>
<td>20%</td>
<td>38%</td>
</tr>
<tr>
<td>Free-Form Text</td>
<td>23%</td>
<td>33%</td>
</tr>
<tr>
<td>Protected Data</td>
<td>31%</td>
<td>21%</td>
</tr>
<tr>
<td>Images</td>
<td>13%</td>
<td>32%</td>
</tr>
<tr>
<td>Video</td>
<td>8%</td>
<td>27%</td>
</tr>
<tr>
<td>Audio</td>
<td>5%</td>
<td>24%</td>
</tr>
<tr>
<td>Other*</td>
<td>5%</td>
<td>11%</td>
</tr>
<tr>
<td>None of the Above</td>
<td>8%</td>
<td>7%</td>
</tr>
</tbody>
</table>

*Source: Nick Heekelaar, "Best Practice for Designing Your Data Lake," Gartner, 16 October 2015.*

*Non-inclusive of processing, monitoring and basic security and customer data.

Gartner is a registered trademark of Gartner, Inc. or its affiliates.
What is Text Analytics/NLP?

Text Analytics = Natural Language Processing + Machine Learning
Why now?

“About 80% of data of an organization processes daily is **unstructured data**” – Gartner 2018

60 to 80% growth per year
What is text analytics being used for today?

Key Applications

- Sentiment Analysis
- Topic Modeling
- Text Classification
- Text Generation

Applications

- Voice of Customers
- Market/Competitive Intelligence
- Fraud/Legal Risk Detection
- Compliance Check
- Triage/Routing
- Conversational Agents
Sentiment Analysis

**Goal:** Determining real-time sentiment scores for use in financial trading strategies

Positive (+)  
(growth, advances, up, strong)  
Buy!

Neutral  
Hold

Negative (-)  
(bust, difficulty, lack, struggle)  
Sell!

**Other Applications:**

- Automating the classification of reviews, whether positive or negative
- Analyzing surveys to understand why customers are satisfied or dissatisfied
- Assessing counterparty credit risk
Text Analytics Workflow

1. **Access Text**
   - Text, PDF, Word, Excel, HTML
   - Languages
   - Visualize

2. **Preprocess**
   - Raw Data
   - Clean Data

3. **Develop Predictive Models**
   - Deep Learning
   - Machine Learning
   - Statistics

4. **Share and Integrate**
   - Integrate
     - MATLAB, Excel, Java, C/C++, .exe, .NET, dll, Python
   - Scale
   - Deployment (Desktop, Web, Enterprise)