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

- **Transactions**: 67% Currently Analyzing, 22% Planning to Analyze, 89% Total
- **Log Data (Including Web)**: 56% Currently Analyzing, 26% Planning to Analyze, 82% Total
- **Geospatial and Location**: 36% Currently Analyzing, 34% Planning to Analyze, 70% Total
- **Social Media Profile**: 27% Currently Analyzing, 38% Planning to Analyze, 65% Total
- **Social Media Chat and Interaction**: 23% Currently Analyzing, 42% Planning to Analyze, 65% Total
- **E-Mails and Documents**: 24% Currently Analyzing, 34% Planning to Analyze, 62% Total
- **Open Data**: 32% Currently Analyzing, 30% Planning to Analyze, 62% Total
- **Data from Data Brokers**: 34% Currently Analyzing, 24% Planning to Analyze, 58% Total
- **Internet of Things**: 20% Currently Analyzing, 38% Planning to Analyze, 58% Total
- **Free-Form Text**: 23% Currently Analyzing, 33% Planning to Analyze, 56% Total
- **Protected Data**: 31% Currently Analyzing, 21% Planning to Analyze, 52% Total
- **Images**: 13% Currently Analyzing, 32% Planning to Analyze, 45% Total
- **Video**: 8% Currently Analyzing, 27% Planning to Analyze, 33% Total
- **Audio**: 5% Currently Analyzing, 24% Planning to Analyze, 29% Total
- **Other**: 5% Currently Analyzing, 11% Planning to Analyze, 16% Total
- **None of the Above**: 8% Currently Analyzing, 7% Planning to Analyze, 15% Total


*Other includes growth, management, sales, security, and customer data.
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
What is text analytics being used for today?

Key Applications

Sentiment Analysis

Text Generation

Text Classification

Topic Modeling

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

Access Text → Preprocess → Develop Predictive Models → Share and Integrate

- **Text, PDF, Word, Excel, HTML**
  - **Languages**
  - **Visualize**
- **Raw Data**
  - **Clean Data**
- **Deep Learning**
  - **Machine Learning**
  - **Statistics**
- **Integrate**
  - MATLAB, Excel, Java, C/C++, .exe, .NET, dll, Python
- **Scale**
  - **Deployment** (Desktop, Web, Enterprise)