

Text Analytics for Executives

What Can Text Analytics Do for Your Organization?



Contents

Introduction.....	1
Getting Started with Text Analytics.....	2
Text Analytics in Government.....	2
Text Analytics in Manufacturing	3
Text Analytics in Financial Services.....	4
Text Analytics in Health Care	5
Conclusion.....	6
Learn More	6

Introduction

Text is the largest human-generated data source. It grows exponentially every day as we post on social media, interact with chatbots and digital assistants, send emails, conduct business online, generate reports and essentially document our daily thoughts and activities using computers and mobile devices.

As we continually transmit and store this information through machines, our human experiences, perspectives and knowledge are increasingly captured as unstructured text data. And organizations want to know how all of that data can be used to drive improvements. For many, unstructured text represents a massive untapped data source with great potential for producing valuable insights that could result in significant business transformations or spur incredible social innovation.

Text analysis is designed to derive value from text data when it is no longer humanly feasible to manually review and categorize that content. It allows organizations to augment and scale the human act of reading, organizing and quantifying text - with the added benefit of uncovering patterns and information buried within.

Text analytics provides structure to unstructured data so it can be easily analyzed. Natural language processing (NLP) is foundational to the text analytics process. It performs linguistic analysis to help a machine read text - essentially helping the machine recognize and understand the construct of a language. Machine learning runs outputs from NLP through data mining and machine learning algorithms to automatically extract key features and relational concepts. Human input from linguistic rules adds to the process, enabling contextual comprehension, or natural language understanding (NLU), of content such as slang, sarcasm, intent and sentiment.

The natural language capabilities that take unstructured data and provide entity and fact extraction, categorization, search and summarization are critical to developing artificial intelligence applications. The powerful insights that can be generated by AI applications are dependent on having large amounts of quality data to analyze. With unstructured text being the largest data source produced by people, text analytics functionality should be core to enterprise deployments of artificial intelligence.

Results from text analytics can be applied to a wide variety of business use cases, including detecting and tracking service or quality issues, quantifying customer feedback, assessing risk, improving operational processes, enhancing predictive models and many more. Text analytics solves a variety of business problems and applications across all industries are expanding. Let's take a look at a few use cases with some advice on where to begin with text analytics in your organization.

Text analytics allows organizations to scale the human act of reading, organizing and quantifying textual data with the added benefit of automated categorization and analysis to produce new insights. Its use across all industries is expanding.

Getting Started with Text Analytics

Developing a text analytics strategy should occur as part of a larger business strategy. What problem are you trying to solve? This requires a visionary who can create direction and articulate a plan. When it comes to text analytics and AI implementations, there are several important factors to consider.

Enterprise applications work best with a well-defined strategy that includes:

- A detailed data management plan.
- Identified roles and expectations for personnel.
- A road map for integrating text analytics into business processes.
- Identified use cases and performance expectations of the technology itself.

Other considerations are:

- Where is the data coming from?
- What data is being collected?
- What information would you like to glean from the data?
- Is the type of data being collected going to provide the type of information you need?
- As you begin to apply text analytics, what trends are being surfaced that may drive new requirements and how will those requirements be addressed or documented?
- What regulatory compliance measures must be adhered to when accessing and analyzing data?

Each organization will have unique considerations and answers to these questions. Here are a few examples of how SAS is helping organizations find innovative ways to discover value with text analytics.

Text Analytics in Government

The US Department of State (DOS) produces the Trafficking in Persons (TIP) report annually. It assesses the state of human trafficking in approximately 200 countries and is used worldwide as a tool to engage in dialogs to advance anti-trafficking reforms and examine where resources are most needed. Freeing victims, preventing trafficking and bringing traffickers to justice are the ultimate goals of the publically available report and of the US government's anti-trafficking policy.

But the information in these reports is scattered across hundreds of free-form text documents. How can the value of that data be transformed into consumable insights that help drive actions for change?

SAS text analytics and visualization capabilities were used to analyze the reports and answer three questions.

- What are the general trends present through the reports?
- Are there any key indicators of trafficking activity?
- Can we visualize these results in a meaningful way?

First, general trends were automatically extracted from the TIP reports. These trends highlighted types of trafficking across geographies, patterns of victim exploitation and identification of vulnerable populations.

Term maps were used to discover key behavior indicators of trafficking activity such as connections between violence, exploitation and intimidation.

Finally, we automatically classified which countries were not meeting the minimal standards of compliance for human trafficking laws and extracted a geospatial pattern across all trafficking reports between source and target countries. This was depicted visually in a network analysis diagram. The network analysis diagram included controls for filtering on trafficking victims, trafficking type and the year of the report.

Annually, after reports are compiled from the DOS and made available to the public, this established text process will quickly assess the most recent report, automatically generating consumable analytics and data visualizations for both the individual report and the aggregate collection.

Analysts and decision makers within humanitarian organizations or offices where the solution is made available will use both the results and insights to further engage in dialogs with governments or stakeholders. Situational awareness generated from text analytics will drive strategic and tactical efforts, providing additional opportunities for collaboration and data aggregation.

Text Analytics in Manufacturing

When someone purchases a car, there are many opportunities to provide feedback via customer surveys. These surveys can be used to collect consumer opinions of the vehicle at the time of purchase, at 30/60/90 day intervals and sometimes even later. How can this disparate data be used to improve vehicle design and engineering?

By consolidating and analyzing the data from customer surveys, an automotive manufacturer used SAS text analytics to create models that automatically linked complaints from those surveys to specific vehicle features. Engineers and product managers used the outputs of this process to guide the modification of those features on a given vehicle.

Government

Issue: Text data widely scattered across free-form documents.

Goal: Compile text data and analytically assess the state of human trafficking around the world.

Result: Visualizations depicting patterns and other information used to spur anti-trafficking reforms.

One specific example was a trend in customer complaints around the glare off of a vehicle's dashboard. The manufacturer had added a stylish carbon fiber dashboard in a particular model, but this feature received many complaints from customers stating it was very reflective. A typical customer complaint was: "Carbon fiber dash looks great in the garage but creates such an intense glare from sunlight that it makes driving the vehicle dangerous."

Engineers switched the material to a less reflective carbon fiber specification. Subsequently, this complaint disappeared from consumer surveys. Customer satisfaction and safety increased as a direct result of incorporating the output from text analytics into the engineering and design process. Most importantly, the manufacturer was able to do this proactively instead of in response to a class action lawsuit.

Text Analytics in Financial Services

Structured account and demographic information has been used to combat money laundering for years. Now financial institutions are beginning to analyze unstructured data sources to augment existing anti-money laundering (AML) efforts.

There is a lot of interest in the financial crime and compliance industry around the application of artificial intelligence to improve the effectiveness of detection programs and automate the manual tasks being performed by investigators.

Text analytics is one area that has enormous potential for financial crimes applications, given that compliance departments have vast amounts of untapped, unstructured data sources. These sources contain rich information including who, where, what, when and how that can be used as an input to many financial crimes use cases such as negative news monitoring, trade finance monitoring and suspicious activity/suspicious transaction reporting (SAR/STR) quality assurance.

How can artificial intelligence help banks extract and derive meaning from text and organize it in a way that helps them perform complex tasks that were previously accessible only through manual human review?

Processing techniques such as theme detection, categorization and entity or fact extraction are all ways to provide structure to free-form text. SAS text analytics can provide structure to free-form text and then use this data to ensure compliance. Here are a few examples.

- **Negative news monitoring.** As an industry standard, financial institutions typically look for negative news related to high-risk customers and customers who have an open AML case. With the wide array of digital news made available daily, the identification of credible news can be challenging. Negative news not relevant to compliance can bias an investigator's decision process, while missed news can leave an institution open to reputational risk. Coupled with bank policy and risk tolerance, an automated process to identify negative news and successfully link this information to customers provides both cost and time savings through automation.

Manufacturing

Issue: Difficulty integrating and using data from various operational systems and customer surveys.

Goal: Proactively understand customer complaints.

Result: Improved automobile design.

- Network analytics. Perhaps one of the best pieces of information for investigating AML is to understand relationships among your customers, as well as non-customers. Most institutions have structured data for known relationships among their customers, but often there are gaps with unknown relationships and those relationships with non-customers. Relationships and networks often surface through normal investigative procedures and are documented in case notes and SAR data. Storing this valuable information and displaying it for future use along with geographic tagging provides deeper insights to the investigations process.
- SAR attribution detection. The detection of money laundering is an exercise in correctly identifying rare events in vast amounts of data. As the AML compliance industry starts to explore the application of artificial intelligence and machine learning to replace Boolean rules, the need for reliably labeled data (target variables) for training becomes even more important. Often, SARs are filed based on external information, but are attributed to the success of one or more rule-based scenarios. Text mining can help determine the correlation. This is critical to not only tune existing models, but also to allow banks to predict newly identified patterns in the future.
- Trade finance document categorization. Deciphering trade documents is a tedious, manual process. Character recognition and natural language processing are being used to train models that read trade finance documents. This automation holds great promise for improving accuracy and reducing the time it takes to process the documents.

Text Analytics in Health Care

Improved Customer Satisfaction

A U.S. based health care payer was faced with declining customer satisfaction as measured by net promoter score (NPS). Despite having multiple data sources, they were unable to clearly identify drivers of declining NPS.

Can text analytics be used to identify and improve customer satisfaction?

The health care company provided SAS with data sets that included customer surveys and survey respondents who had a recent interaction with the customer call center. SAS performed data cleansing on this data and used text analytics to develop both rules-based and machine learning-based models to identify potential drivers of NPS.

Employees are now able to further research and test those results and achieve a deeper understanding of how that information can be used to drive positive change. Data scientists are tasked with building and maintaining the text models, while business users are responsible for improving the score using results shared by the data science team.

The success of such an analytics investment lies on the commitment, dedication and innovation of the employees across the enterprise who collectively contribute to the organization's performance and customer engagement.

Financial Services

Issue: Time-consuming and tedious to categorize and extract facts and entities in documents.

Goal: Automate manual processes and enrich structured data.

Result: Reduced compliance risk and improved productivity.

Health Care

Issue: Decreasing customer satisfaction.

Goal: Combine and analyze surveys and call center notes to identify problems.

Result: Rules-based and machine learning models that pinpoint drivers of declining provider scores.

Automated Case Handling

A clinical research organization runs and administrates clinical trials for pharmaceutical companies. Case reports are manually read to identify severe adverse events during clinical trials for new medications. These manual processes add time, expense and variability to the clinical trial evaluation process.

Using SAS text analytics, they were able to train a model to score new case reports for severe adverse events. And as a result, they are automating the first-pass triage of case reports. Low and high scores can be classified without manual review while those with inclusive results are sent for manual review. This can add accuracy and timeliness to the clinical trials process.

Conclusion

SAS text analytics provides a flexible framework capable of tackling a variety of use cases in support of a single initiative. This framework supports the end-to-end text analytics life cycle, including preparing data, visually exploring topics, extracting entities and facts, analyzing sentiment, building a variety of text models and deploying them within existing systems or processes. Machine learning powers topic generation, categorization, entity/fact and sentiment extraction to automatically identify relationships and patterns that exist within text.

When looking to implement a text analytics and artificial intelligence application within your organization, it's important to consider a technology platform that can integrate diverse and disparate data sources to support an enterprise deployment of analytics, inclusive of appropriate security and governance models.

The SAS Platform fosters collaboration by providing a toolbox where best practice pipelines and methods can be shared. SAS also seamlessly integrates with existing systems and open source technology. Open APIs enable users to call SAS text analytics using a programming environment and language of familiarity. APIs are available for Java, Python, Lua, R and RESTful web services.

Your unstructured text data is growing daily, and data without analytics is opportunity yet to be realized. Discover the value in your data with text analytics capabilities from SAS.

Learn More

To learn more about SAS text analytics, visit sas.com/vta.

Health Care

Issue: Too many case reports to manually read.

Goal: Automate first-pass of case studies to quickly identify severe adverse events (SAEs).

Result: Faster, more accurate clinical trials process.

To contact your local SAS office, please visit: sas.com/offices

