

SAS® Visual Text Analytics

Combine the power of natural language processing, machine learning and linguistic rules to reveal insights in data



Text analytics helps solve a variety of everyday business problems – things like managing

and interpreting notes, assessing risk or fraud, and incorporating customer feedback for earlier detection of problems. With more unstructured data than ever, the use of text analytics is expanding across all industries.

SAS Visual Text Analytics analyzes large volumes of unstructured data using predefined templates, machine learning methods and natural language processing (NLP) to produce deeper insights using more data, faster than ever before. This software combines text mining, contextual extraction, categorization, sentiment analysis and search within a modern and flexible framework. An end-to-end visual pipeline makes it easy to prepare data, visually explore topics, extract entities and facts, analyze sentiment, build text models and deploy them within existing systems or processes.

What does SAS® Visual Text Analytics do?

SAS Visual Text Analytics offers a wide variety of modeling approaches for getting the most value from unstructured data, including supervised and unsupervised machine learning, linguistic rules, categorization, entity extraction, sentiment analysis and topic detection.

Why is SAS® Visual Text Analytics important?

Text data is pervasive across all industries with volumes growing each day. SAS Visual Text Analytics provides a comprehensive solution that overcomes the challenges of identifying and categorizing text data. It enables organizations to scale the human act of reading, organizing and extracting useful information from huge volumes of textual data. You build models that analyze and categorize a set of documents. Unstructured data is automatically converted into meaningful insights that feed machine learning models.

For whom is SAS® Visual Text Analytics designed?

It's designed for business analysts, domain experts, research analysts, linguists, knowledge workers and data scientists who need to analyze large amounts of unstructured data to glean new insights.

Benefits

- **Uncover emerging trends and spot new opportunities for action with artificial intelligence.** Automatically convert unstructured data into meaningful insights. Increase the accuracy of text models by combining machine learning methods with a rules-based approach that can be enhanced with subject-matter expertise.
- **Automate comprehensive analytics.** SAS Visual Text Analytics offers great breadth and depth of analytical capabilities through a rich mix of rules and machine learning, as well as integrated deep learning. It provides quick start pipelines and can use industry taxonomies to readily support both predictive and prescriptive analytics.
- **Go from data to decisions more quickly.** Empower decision making at the source of the data. If someone leaves a comment or clicks through an app on a mobile

device, that data is analyzed immediately and accelerates the data-to-decision timeline. Organizations can reduce the gap between when information is received and when it's acted upon with in-memory, in-database and streaming technologies.

- **Fuel collaboration and information sharing in an open ecosystem.** SAS Visual Text Analytics provides a flexible environment that supports the entire analytical life cycle – from data preparation and visual exploration to analysis and deployment. Tackle and experiment with a variety of analytical use cases to support a single initiative. Whether you're a data scientist preparing data, a domain expert applying linguistic rules or an IT person deploying models, collaboration is possible at all levels. This unified solution integrates seamlessly with existing systems and open source technologies.

Overview

When the volume of text-based data becomes too large to manually review and analyze, it's time to add text analytics to your insights arsenal. SAS Visual Text Analytics supports the entire analytics life cycle with data management, comprehensive analytics and flexible deployment options.

Embedded data preparation and visualization

SAS Visual Text Analytics comes with embedded data integration and preparation capabilities to help access, integrate, profile, cleanse and transform data. You can import text directly from more than 35 out-of-the-box data connectors, including multiple document formats, relational databases, remote file system data sources, local data file types, social media connectors and Esri. The software also includes self-service data visualization capabilities for exploring and understanding your text data.

Combined machine learning, deep learning and rules-based methods

SAS Visual Text Analytics combines a rich mix of artificial intelligence (AI) techniques (supervised and unsupervised machine learning and deep learning) and rules-based approaches to automatically surface themes in text data, generate rules for categories of interest, produce visual representations of related terms for exploratory analysis and create best practice templates for specific business use cases.

SAS Visual Text Analytics derives topics from groups of important terms in your documents. You can explore trending topics in text and see how they change over time. These insights can kick off immediate actions. Or, you can fine-tune them with subject-matter expertise.

If discovered themes require additional tuning or custom definitions are preferred, you can create precise, rules-based categories and concepts. Despite its advances, machine learning still can't capture specific nuances and complexities of language ambiguity. With advanced rule tuning and a wide

range of Boolean operators and linguistic qualifiers, users can dig deep into their unstructured data.

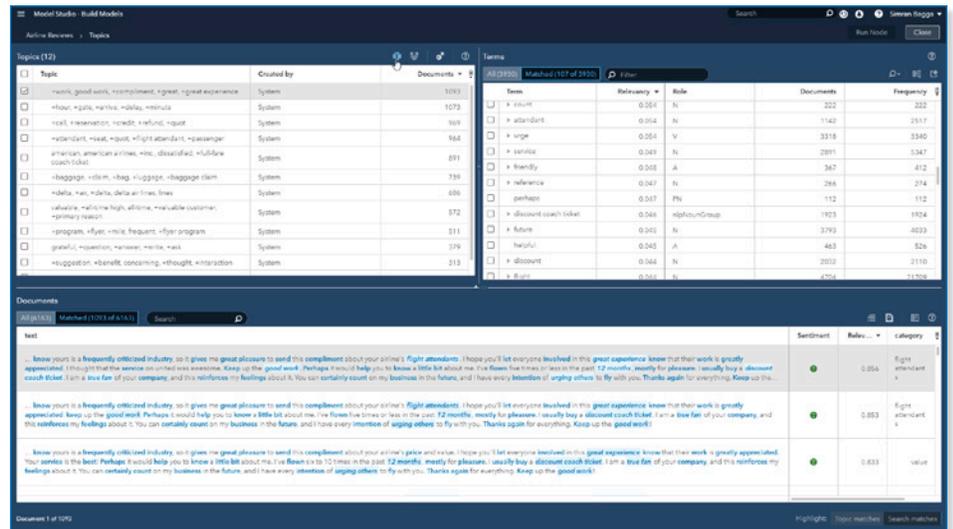
Contextual extraction

The contextual extraction technique is often used to isolate and pull out important pieces of information where the value of matches to specific context is of utmost importance. By extracting snippets buried in free-form text without the need for manual markup, you can derive new variables to incorporate into reports, predictive models, or enhance search and filtering applications. You can use predefined concepts to detect and extract data elements and relationships from unstructured text, or you can create custom concepts and definitions.

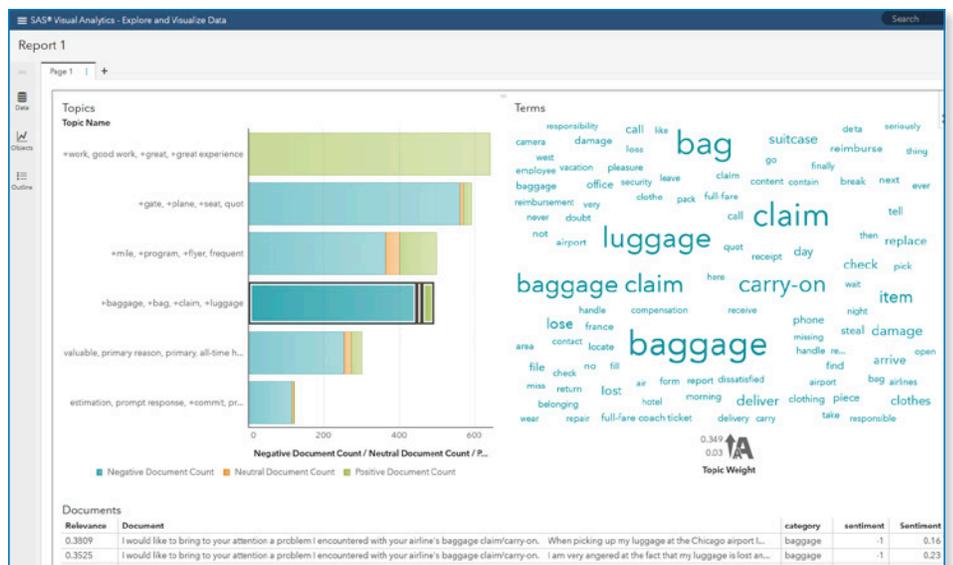
SAS Visual Text Analytics provides the most comprehensive set of tools for achieving precision and contextual specificity to aid in contextual extraction with a wide range of rule types, Boolean and linguistic operators, term qualifiers, regular expressions, part of speech tags and more.

Flexible deployment

With SAS, you can deploy models in batch, in Hadoop, in stream and via APIs. Data does not have to go through the user interface to be enriched by a text model. Models can be run closer to where data is located. This reduces data movement and produces faster results for scoring new data.



Automatic topic discovery groups documents based on common themes.



Powerful NLP capabilities allow you to easily identify and extract important patterns in text.

Multiuser environment fosters teamwork and collaboration

As part of the SAS Platform, SAS Visual Text Analytics fosters teamwork by providing a workspace to share best-practice pipelines and methods. Any extensive analysis, such as identifying relevant terms, modifying or creating user topics, creating linguistic rules, etc., can be shared as a Best Practice node in a pipeline across multiple SAS projects.

SAS also integrates with existing systems and open source languages, including Python, R, Java and Lua. And with RESTful APIs, you can easily add the power of SAS to other applications.

The software provides an intuitive user interface that accounts for important factors such as localization/internationalization and accessibility. Additionally, by offering open APIs and a microservices architecture, SAS offers users the ability to use their own user interface or build a custom search application.

Natural language processing

Manually reviewing documents is time-consuming and prone to errors. Natural language processing (NLP) reduces the need for tedious manual analysis and allows you to more easily identify and extract important patterns in the text.

NLP is often the first step in the text analytics process. It performs linguistic analysis to help a machine “read” text. SAS Visual Text Analytics uses NLP to analyze and transform text into formal representations for text processing and understanding.

There’s no need to cobble together disparate NLP libraries or custom code to perform word and sentence tokenization, segmentation, stemming, compound decomposition, part-of-speech tagging, named entity recognition and semantic parsing. You spend less time programming the computer how to interpret text and more time deriving business value from textual data.

Key Features

Machine learning and rules-based approaches in a single project

- Unsupervised machine learning automates topic generation.
- Supervised/probabilistic machine learning models include BoolRule (enables automatic rule generation for document categorization), and conditional random fields and probabilistic semantics (used to label and sequence data and can automate entity and relationship extraction by learning the contextual rules of a given entity).
- Automatic rule builders promote topics to categories with supervised machine learning.
- Rules-based linguistic methods are used for extracting key concepts.
- Rules can be tested on an input data set prior to deployment.
- Automatic parsing can be used along with deep learning algorithms (recurrent neural networks) to classify documents and sentiment more accurately.

Contextual extraction

- Use predefined concepts to extract common entities such as names, dates, currency values, measurements, people, places and more.
- Create custom concepts using keywords, Boolean operators, regular expressions, predicate logic and a wide array of linguistic operators.
- Programmatically populate fields that contain the desired matched information. No need to manually mark up document for entities or facts of interest.
- Reference a predefined or custom concept in a categorization rule, for extra contextual specificity or reach.

Flexible deployment

- Concepts, Sentiment, Topics and Categories nodes provide score code needed to deploy models on an external data set.
- Score code is natively threaded for distributed processing, taking maximum advantage of computing resources to reduce latency to results, even on very large data sets.
- Register text models and monitor performance and model decay with SAS Model Manager.

Multiuser environment

- Graphical user interface with visual programming flow.
- Share projects with other users.
- Five nodes available through text analytics pipeline (Concepts, Text Parsing, Sentiment, Topics and Categories).
- Prepopulated default pipeline to represent typical workflow of a text analytics project.
- Flexible pipelines enable users to create additional nodes or modify default pipeline.
- Register models in SAS Model Manager for easier management of text models.

Natural language processing

- Automated parsing, tokenization, part-of-speech tagging and lemmatization.
- Ability to apply start and stop lists.
- Ability to use special tags, qualifiers and operators in linguistic rules that take advantage of part-of-speech tagging, tokenization and lemmatization (allows for more precision or better recall/abstraction capabilities).
- Detect misspellings.

Automated feature extraction with machine-generated topics

- Automatic topic discovery groups documents based on common themes; each document may contain zero, one or more themes.

Automated, machine-generated topic detection

SAS Visual Text Analytics provides two machine learning methods for automatic topic discovery within your documents: singular value decomposition and latent Dirichlet allocation, which is popular within the open source community. With these methods, you can quickly discover text topics and inspect the terms and documents that make up these natural groupings. Discover themes you may not have thought to look for. Then, deploy the model against new data or convert topics to logic that can be manually modified or extended by subject-matter experts.

Native linguistic support for multiple languages

SAS Visual Text Analytics supports a wide variety of languages through dictionaries and linguistic assets created by native language experts. This helps support the global challenges organizations face. Standardization of part-of-speech tags and named entities across languages is key for organizations looking to implement text analytics in a variety of languages. SAS Visual Text Analytics includes out-of-the-box analysis functionality for 32 languages. These language packs enable native language analysis as opposed to language translation prior to analysis.

Sentiment analysis

Sentiment analysis identifies an author's tone or attitude (positive, negative or neutral) that is expressed through text. SAS Visual Text Analytics identifies and analyzes terms, phrases and character strings that imply sentiment. Continuously track sentiment to see how dimensions of interest are changing. Better understand and categorize feedback and adjust decisions based on perspectives.

Key Features (continued)

- Relevance scores are produced that characterize how well each document belongs to each topic, as well as a binary flag showing topic membership above a given threshold.
- Merge or split topics automatically generated by the machine (unsupervised machine learning) to create user-defined topics (subject-matter expertise to refine automated AI output).

Native linguistic support for multiple languages

- Out-of-the-box text analysis for 32 languages: Arabic, Chinese, Croatian, Czech, Danish, Dutch, English, Farsi, Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Indonesian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovene, Spanish, Swedish, Tagalog, Turkish, Thai and Vietnamese.
- Default stop list provided for each language the application supports.
- Built-in lexicons that contain part-of-speech information and dictionary-based expansion to detect and resolve surface forms to root form (verb conjugations, plurals, etc.).

Sentiment analysis

- Visual depiction of document-level sentiment through sentiment indicator display at a document and topic level.
- Default domain-independent sentiment analysis taxonomy for 14 languages: Arabic, Chinese (simplified), Chinese (traditional), Dutch, English, Farsi, French, German, Italian, Japanese, Korean, Portuguese, Spanish and Turkish.
- Ability to import and execute custom sentiment models built in SAS Sentiment Analysis.
- Ability to use recurrent neural networks for more accurate sentiment classification.

Open APIs

- Seamlessly integrate with existing systems and open source technology.
- Add the power of SAS Analytics to other applications using SAS® Viya® REST APIs.
- Out-of-the-box analytical programming interfaces for text summarization, text data segmentation, text parsing and mining, topic modeling, text rule development and scoring, text rule discovery, term mapping and topic term mapping, conditional random field and search.

TO LEARN MORE »

To learn more about SAS Visual Text Analytics, view screenshots and see other related materials, please visit sas.com/vta.

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

