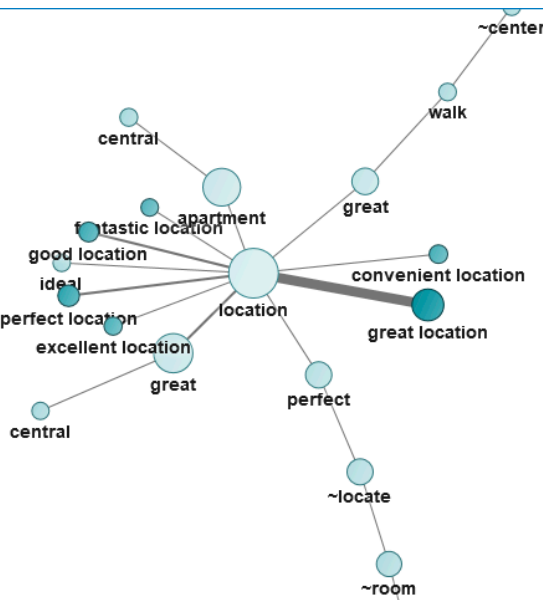


SAS® Visual Text Analytics

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



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 automatically analyze and categorize a set of documents. Unstructured data is automatically converted into meaningful insights that feed machine-learning and predictive 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.



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.

Benefits

- Uncover emerging trends and spot new opportunities for action.** 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.
- Use automation to drive faster results.** Intelligent algorithms and NLP techniques automatically identify relationships and patterns in text data, eliminating time-consuming manual analysis. Identifying and extracting important topics in freeform text produces new variables that enhance predictive models, reports and search or filtering applications.
- 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 in-stream 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 combines:

- A visual programming flow, called a pipeline, with nodes you can control.
- Machine learning and rules-based linguistic methods for categorization and entity extraction.
- Document-level scoring for each component.

The output of text analysis can be applied to a variety of business uses, such as detecting and tracking service or quality issues, quantifying customer feedback, assessing risk, improving operational processes, enhancing forecasting and predictive models and many more.

Combined machine-learning and rules-based methods

Manually pinpointing central themes and topics across multiple documents is a difficult, if not impossible, task. SAS Visual Text Analytics uses machine-learning techniques to automatically derive topics from groups of important terms in your documents. (Topics are collections of terms that describe and characterize a main theme or idea.) 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.

When 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. You need linguistic rules to understand slang, detect sarcasm and infer intention. With advanced rule tuning and a wide range of Boolean operators and linguistic qualifiers, users can dig as deep into their unstructured data as they want to.

Contextual extraction

When key data elements such as names, dates, measurements or cause-and-effect relationships are buried within a freeform text field, they are less accessible for activities such as filtering, search or reporting. SAS Visual Text Analytics provides the most comprehensive set of tools for achieving precision and contextual specificity, including part-of-speech tags, regular expressions and many different rule types, linguistic qualifiers and Boolean operators.

You can use predefined concepts to detect and extract data elements and relationships

from unstructured text, or you can create custom concepts and definitions. By checking for correct context and extracting only the critical pieces of these data elements, new structured attributes are produced to enhance downstream applications, as well as feed natural language understanding (NLU) and artificial intelligence (AI) applications.

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.

Multiuser environment fosters teamwork and collaboration

The SAS Platform offers more depth and breadth of integrated analytical capabilities than any other software. This integration is key to collaboration and information sharing. As part of the platform, SAS Visual Text Analytics fosters teamwork by providing a workspace to share best-practice pipelines and methods.

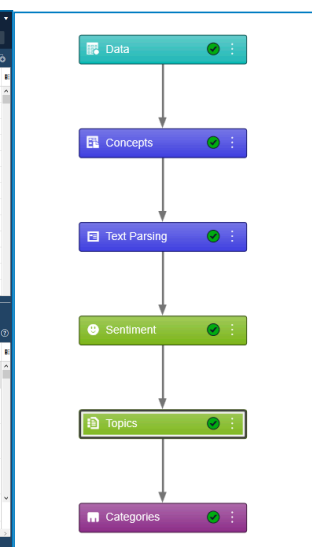
Any extensive analysis, such as identifying relevant terms, modifying or creating user topics, creating linguistic rules, etc., can be shared as a best practice node in a pipeline across multiple SAS projects.

The screenshot displays the SAS Visual Text Analytics interface. It features a 'Kept Terms' table with columns for Term, Role, Documents, and Frequency. Below this is a 'Dropped Terms' table with similar columns. At the bottom, a 'Documents' table shows a list of documents with columns for Document ID, Review Score, and Room Type. The interface also includes a search bar and various navigation icons.

Term	Role	Documents	Freq
room	N	1825	2330
comfortable	A	2997	2247
recommend	V	2132	2191
make	V	1662	1977
good	A	1702	1954
again	ADV	1789	1863
need	V	1575	1812
time	N	1555	1802
really	ADV	1407	1773
easy	A	1582	1750
neighborhood	N	1458	1571
also	ADV	1302	1570
boston	noun/Place	1401	1541

Term	Role	Documents	Frequency
be	V	9099	38995
the	DET	8333	33152
and	CONJ	8705	28489
to	PPOS	6801	17387
a	DET	6832	16964
in	PPOS	5733	10793
i	PRO	4252	9108
we	PRO	3502	7505
for	PPOS	4438	7258
it	PRO	3994	6865
of	PPOS	3960	6821
have	V	4043	6648
with	PPOS	3749	5521

Document	Review Score	Room Type
Cristina is a great host. Her apartment is clean, neat, and functional, just as described. She arranged everything efficiently and effectively. We had a great time and would recommend her to everyone who is considering renting her place.	91-100	Entire home/apt
Cristina and the apartment were terrific. Everything well organized and easy, and the apartment is stylish and very clean, thoughtfully set up for guests, and even nicer in person than in the photos. Right around the corner from a great coffee/lunch place, a T stop, drugstore, and groceries, lots of restaurants in walking distance, but apartment very quiet. We hope to stay here again when visiting friends in Brooklyn and highly recommend.	91-100	Entire home/apt
Cristina was a great host! It was such a great space to stay at and Cristina's straightforward description and instructions made the whole process easy and stress free. I loved everything about the space, not only was it clean and organized, but the decor was great, and truly made me feel like I was staying at a friend's house. I have nothing but gratitude for Cristina!	91-100	Entire home/apt
Cristina's apartment was in a great location. The place is clean and spacious for three people, and looks just like the photos if not even better. Loved staying in this beautiful apartment. Could not recommend it enough. She left us clear instructions and was easy to communicate with.	91-100	Entire home/apt
Cristina is very patient and friendly. She sent her husband to meet us, tell us everything and gave us the parking space. The apartment is very good, neat and comfortable. I enjoyed my stay there.	91-100	Entire home/apt



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

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.

Natural language processing

Manually reviewing documents is time-consuming and can leave you open 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.

Natural language understanding (NLU) is a subset of NLP that enables contextual understanding of content. NLP capabilities, such as entity and fact extraction, categorization, search and summarization, are critical to developing AI applications.

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.

Using 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.

Key Features

Machine-learning and rules-based approaches in a single project

- Unsupervised machine learning automates topic generation.
- 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.

Contextual extraction

- Use predefined concepts to extract common entities such as names, dates, currency values, measurements 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.

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.

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).

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.
- 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.

Native linguistic support for multiple languages

- Out-of-the-box text analysis for 30 languages.
- 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.).

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 30 languages: Arabic, Chinese, Croatian, Czech, Danish, Dutch, English, Farsi, Finnish, French, German, Greek, Hebrew, Hindi, Indonesian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Slovak, Slovene, Spanish, Swedish, Tagalog, Turkish, Thai and Vietnamese. 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. You gain a better understanding of topics when emotional context is tied to someone's experience.

With SAS Visual Text Analytics, you can move beyond simply returning results for specific keywords to extracting related trends with increased relevance to your unique business applications. Continuously track sentiment to see how dimensions of interest are changing. Better understand and categorize feedback from customers, employees, citizens, etc. and adjust decisions based on perspectives.

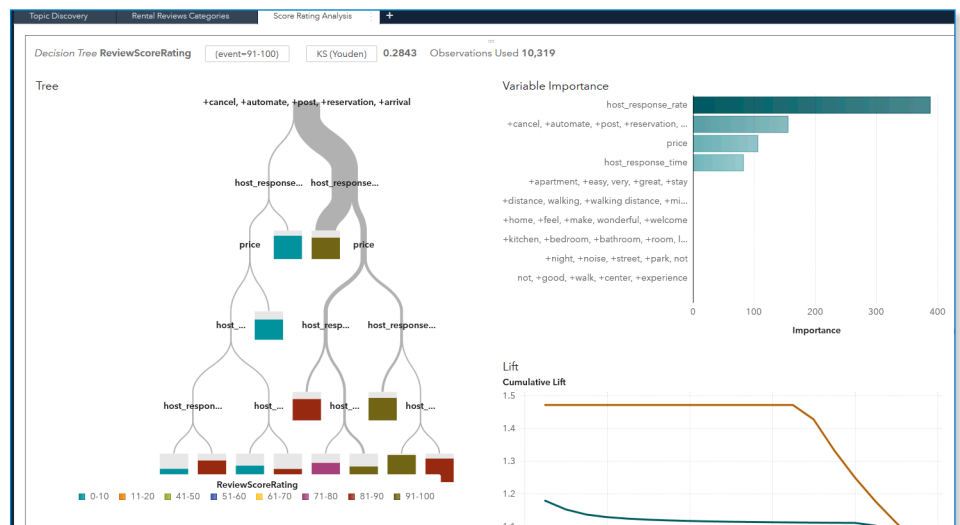
Key Features (continued)

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.

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.



Explore results easily with automatic visualization of your text outputs.

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

