Ask the Expert

SAS Text Miner: Getting Started

§sas
THE POWER TO KNOW.

Ask the Expert
SAS Text Miner: Getting Started

Presenter: Twanda Baker
Senior Associate Systems Engineer
SAS Customer Loyalty Team

Q&A: Melodie Rush
Senior Analytical Engineer
SAS Customer Loyalty Team

§sas

§sas

Goals

- Increase awareness of and comfort with capabilities in SAS® Text Miner
- Share resources for learning more

CUSTOMER LOYALTY TEAM · Support You Can Count On
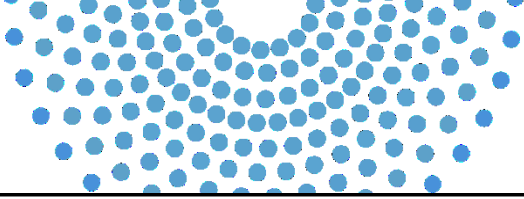

Today's Agenda

This presentation will demonstrate the basic steps for getting started using SAS® Text Miner, such as how to:

- Parse test data
- Filter text data
- Analyze text data including topic discovery and cluster analysis
- Use text mining results as input to predictive modeling

SAS® Text Miner
Prerequisites for this session

This example assumes basic knowledge of using the interface in SAS® Enterprise Miner™ and how to create Data Sources and Diagrams.

For an introduction or review, visit these resources.

- Six part video tutorial on YouTube
- "Getting Started with SAS® Enterprise Miner™" in the Ask the Expert series
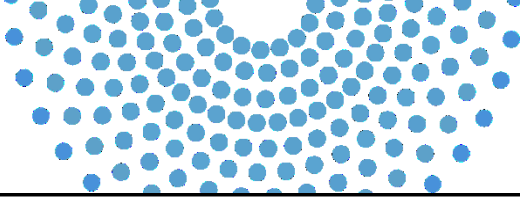- Review the free downloadable step-by-step tutorial.
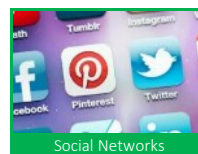
§sas

Is there valuable information "locked away" in your unstructured data?

§sas

---

§sas

Unstructured Text: Where is it?

Call Center Notes	Associate Comments	Survey Feedback	Research & Publications	Claims & Case Notes
Live Chat	Factory/Tech'n Notes	HR data	Medical/Health Records	Contracts & Applications
Online Forums	Blogs	Consumer Reviews	Online News	Social Networks
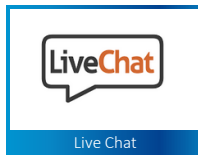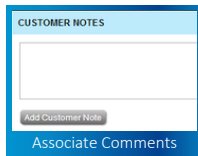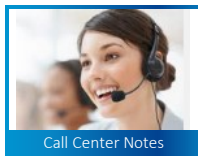
What if you could…

| Discover new insights from large text data sources	Extract key patterns from text data to predict the future
Customers	
Discover current topics about your products from customer opinions	Find patterns within customer feedback, that predicts good interest in upsell opportunities
Fraud	
Detect anomalies from usual topics described in text reports, text applications or feedback	Find patterns in reports that may seem to predict/ relate to suspicious behavior
Public Opinion	
Understand previously unknown issues/ concerns, from citizen discussions on twitter/ forums	Extract key opinions from citizen feedback to forecast citizen sentiments in the near future
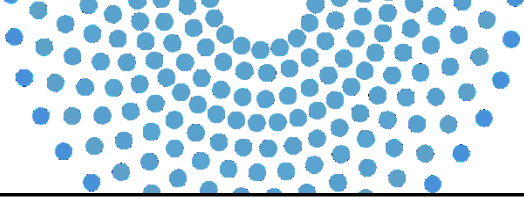
Where is Text Mining used?

Text Mining has numerous applications in any industry	Government	Finance	Insurance
	Detect fraudulent activity. Spot emerging trends and public concerns.	Retention of current customer base using call center transcriptions or transcribed audio. Identification of potentially fraudulent activities.	Identify fraudulent claims. Track competitive intelligence. Brand management
Retail	**Manufacturing**	**Telecommunications**	**Life Sciences**
Identify the most profitable customers and the underlying reasons for their loyalty. Brand management	Reduce time to detect root cause of product issues. Identify trends in market segments.	Help prevent churn and suggest up-sell/cross-sell opportunities for individual customers.	Identify adverse events. Recommend appropriate research materials.

§sas

SAS TEXT ANALYTICS

TEXT MINING

Import and explore textual data to uncover valuable patterns & themes, and incorporate text into predictive models

CONTEXTUAL ANALYSIS

- Information Retrieval
- Automatic Topic Detection
- Content Categorization
- Entity/Fact Extraction
- Sentiment Analysis

VISUALIZATION

REPORTING

INTEGRATED ANALYTICS

Integrate structured and unstructured data for enhanced:
- Forecasting
- Optimization
- Predictive Modeling
- Network Analysis

§sas

SAS® Text Miner

- SAS® Text Miner is an add-on product to SAS® Enterprise Miner™
- This adds the capabilities of analyzing unstructured data to the broad set of techniques in Enterprise Miner™.
- The interface is Enterprise Miner™ – when Text Miner is licensed, there is an additional tab of tools.

| Sample | Explore | Modify | Model | Assess | Utility | Credit Scoring | HPDM | Applications | Text Mining | Time Series |
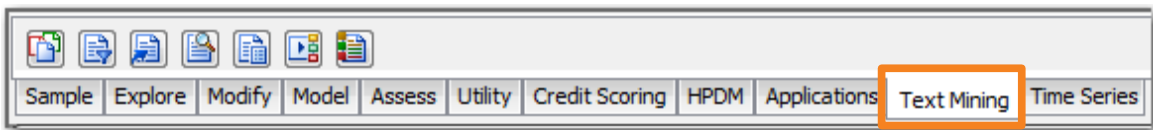
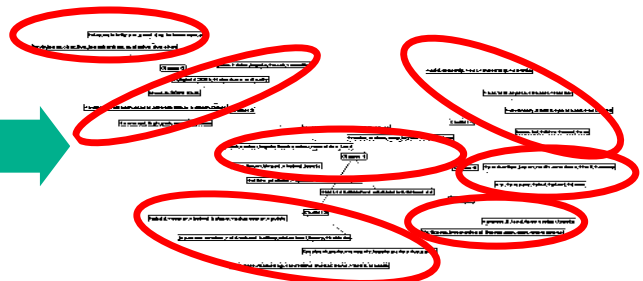# How does Text Mining work?
Exploring & Discovering Insights

1. Input text messages – e.g. twitter data, reports, email, news, forum messages

2. Parse & explore Text Data –break down text and explore relationships of key concepts such as persons, places, organizations…

3. Discover Topics – cluster documents of similar content and describe them with important key words

How does Text Mining work?
Discover patterns for predictive modeling

1. Input text messages with relevant structured data –
e.g. email, call center notes, applications

2. Parse Text Data and Discover Topics – Break down text into structured data, group messages of similar content

3. Predictive Modeling with text data – text data input into models may provide reliable info to predict outcome & behavior
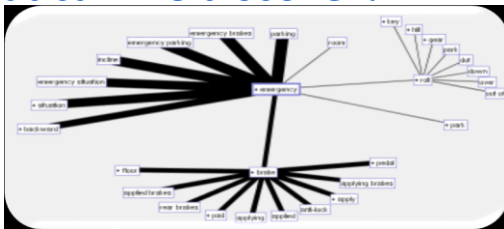
Customer data

this text

*Predict activity that is likely fraudulent…*

§sas

What can we discover?

Discover relationships between concepts described in large corpus of text data –
how are persons, places, organizations related?

Discover topics mentioned in text data–
what are main topics mentioned?
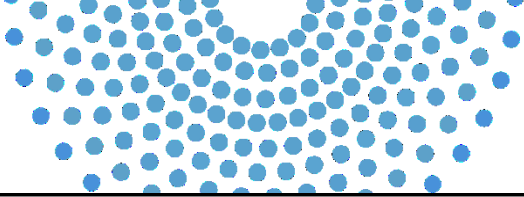What are the rare topics?

Variables – Tree

(none) ▼ ☐ not Equal to ▼ ...

Columns: ☑ Label ☐ Mining

Name /	Label	Use
TextTopic_1	Airbag deployment	Default
TextTopic_10	+dealer,+problem,+truck,+fix,+find	Default
TextTopic_11	+replace,+dealer,+problem,+car,+mechanic	Default
TextTopic_12	+dealership,+vehicle,+problem,+replace,+light	Default
TextTopic_13	+mile,+rotor,+replace,+dealer,+vehicle	Default
TextTopic_14	+time,+problem,+vehicle,+dealer,+car	Default
TextTopic_15	+brake,+vehicle,+break,+rotor,+apply	Default
TextTopic_16	+noise,+vehicle,+problem,+dealer,+service	Default
TextTopic_17	+engine,+vehicle,+light,+fuel,+stall	Default
TextTopic_18	+seat,+belt,+passenger,+consumer,+vehicle	Default
TextTopic_19	+door,+vehicle,+problem,+passenger,+driver	Default

+dealer,+problem,+truc...

Airbag deployment

+vehicle,+problem,+de...
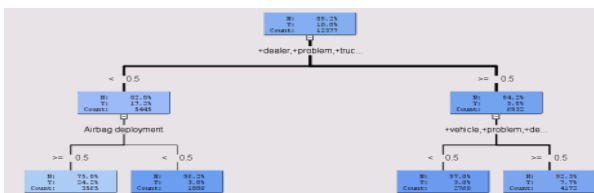
Discover patterns related to structured data –
e.g. how is feedback related to customer purchase behavior?

§sas

---

§sas
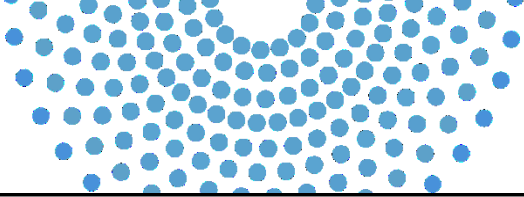
SAS® Text Miner

SAS® TEXT MINER – ANALYTICAL WORKFLOW
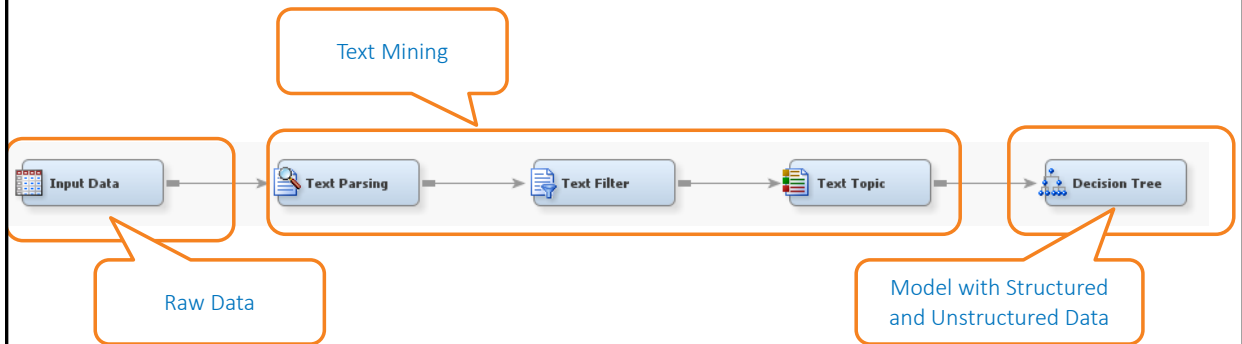
Text Mining

| Input Data | Text Parsing | Text Filter | Text Topic | Decision Tree |

Raw Data

Model with Structured and Unstructured Data

Text Mining Process Flows

| Input Data | → | Text Parsing | → | Text Filter | → | Text Topic |

① ② ③

Ssas

Text Mining Process Flows

| Input Data | → | Text Parsing | → | Text Filter | → | Text Topic |

⬆

Start with a table that contains either:
- Documents saved as a variable (column)
- A column that points to physical text files

Ssas

Example Input Data

CAGE_YR	SEX	SYMPTOM_TEXT	SYM_CNT	NUMDAYS
64	F	Information has been received from an RN concerning a 64 year old white, obese female who on 11/14/01, at 11:...	10	0
29	F	Information has been received from an NP concerning a 29 year old female pt who on an unspecified date was va...	2	66
0.3	F	Memory loss, family loss, Mother and father are dead. Permanent coma. Delusional thoughts if my brain died. Ado...	2	.
1.6	F	Sabin tri vaccines were not good ones. They make you taller and handicapped looking.	1	.
4	M	Cellulitus at administration site.	3	1
69	F	Approximately 5 minutes after receiving flu and pneumonia vaccine pt began hollering, "Oh, Oh my neck is hurting...	8	0
77	M	Demyelinating disease; dizziness, blurred vision; difficulty hearing and walking.	5	12
0	F	Pt presented to ER complaining of swelling in hands and lips; Bilateral hand paresthesia, tightness in throat and c...	4	0
74	F	Immediate pain, redness. In about 1 hour shot raised about 1 inch round lump and pain continued. In a few hours...	4	0
0.2	F	Infant had a sudden event with cyanosis, pallor, bradycardia and shallow breathing. This lasted approximately 20...	4	0
1	M	Autistic mannerisms, system "shutdown". Blank stares, catatonic state.	4	7
1.3	M	Loss of speech and coordination.	2	4
39	F	Reportedly called in after first dose to report had a rash that sounded like hives 2 days after immunization. Very itc...	1	2
0.1	M	Rash immediately following 03/21/1997 hep vaccine with arrested head growth and neurological complications. Ar...	7	.
40	F	Large firm, red region at site of injection. Skin is itchy but not painful. Estimate 3 inch in diameter. Will treat with w...	3	0
0.2	M	The child attorney alleged that his mother received two injections of Rho { D } immune globulin, which contained t...	2	.
0.4	F	This report is concerning a 4 month old female who on 29-NOV-2001 was vaccinated with one dose of Comvax. C...	3	2
0.2	M	Five to ten minute seizure, generalized. No meds given for seizure. On ampicillin and Garamycin. Rule out sepsis...	2	0
0.5	M	Fever for 3 days. Emesis (recurrent) . Empiric antibiotics.	2	1
0	U	Fever, fussiness, questionable hematuria. 07/14: Right otitis media. 07/15: Persistent fever despite antibiotics. Fu...	4	2
0.2	M	Temp 104. 5 within 24 hours of immunizations. Pt had septic URI admitted for 48 hours to local hospital then sent...	2	1
60	F	Five day post vaccination, the patient began to experience swelling of all of her joints. She also loss strength. She...	2	6
0	M	The patient c/o of pain, erythema, edema and cellulitis at injection site.	4	2
33	M	Patient experienced a baseball sized edema. The area was erythematic and hot to the touch. As of 12/03/2001, th...	4	1
0.7	M	Immediately after receiving the DTaP vaccine in his left thigh, the patient cried for three seconds and then stop br...	1	0
58	F	Patient c/o very sore left arm, she states that she ca not lift it without pain.	2	0

Example Input Data
Variable contains pointer to text file

	TEXT	url	
154	VEHICLE WAS PARKED, PULLED OUT OF GEAR, ROLLED, IMPACT 06:00 POSITION...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks237.txt.htm	
155	KEY CAN BE REMOVED FROM IGNITION IN ANY GEAR. *LDG	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks238.txt.htm	
156	KEY WAS REMOVED FROM IGNITION WHILE VEHICLE WAS STILL IN GEAR, VEHIC...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks239.txt.htm	
157	WHILE DRIVING BETWEEN 35-40MPH CONSUMER HIT ANOTHER VEHICLE HEAD ...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks24.txt.htm	
158	SUDDEN ACCELERATION, WHILE BRAKE WAS APPLIED. *AJ	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks240.txt.htm	
159	SUDDEN ACCELERATION IN REVERSE CAUSED ACCIDENT. *AJ	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks241.txt.htm	
160	SUDDEN ACCELERATION, CAUSING ACCIDENT. *LDG	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks242.txt.htm	
161	WHILE IN PARK, VEHICLE ROLLED DOWN DRIVEWAY, CAUSING ACCIDENT PEB	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks243.txt.htm	
162	VEHICLE MOVED WHILE IN PARK AFTER RECALL REPAIRS, CAUSED ACCIDENT (9...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks244.txt.htm	
163	VEHICLE ROLLED BACKWARDS FROM PARKED POSITION, CAUSING ACCIDENT. *...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks245.txt.htm	
164	VEHICLE JUMPED OUT OF GEAR PRIOR TO RECALL (91V189), RESULTING IN AN A...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks246.txt.htm	
165	AUTOMATIC TRANSMISSION SLIPPED OUT OF PARK INTO NEAUTRAL AND THEN ...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks247.txt.htm	
166	VEHICLE PARKED IN DRIVEWAY AND ROLLED BACKWARDS INTO ANOTHER PAR...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks248.txt.htm	
167	WHEN AVOIDING ACCIDENT, VEHICLE BEGAN TO ACCELERATE OUT OF CONTRO...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks249.txt.htm	
168	VEHICLE HIT A BARRIER AT 45 MPH, THEN ROLLED, AND AIRBAG DID NOT DEPL...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks25.txt.htm	
169	AXEL BROKE, CAUSING REAR TIRE TO COME OFF, RESULTING IN ROLLOVER/INJ...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks250.txt.htm	
170	VEHICLE WAS PARKED W/ HATCH BACK UP & KEY OUT OF IGNITION. CHILD WAS...	file://\\L72307\Public\TM\Examples\NHTSA\datafiles\trucks251.txt.htm	
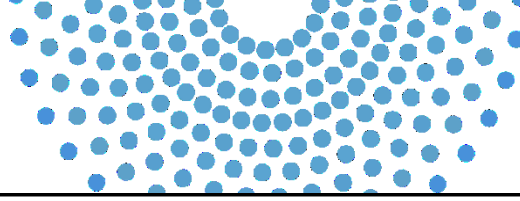
Importing Text
If your data isn't already in a SAS Table
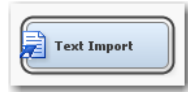
- The **Text Import** node enables you to create data sets dynamically from files contained in a directory or from the Web.

- The **Text Import** node takes an import directory containing text files in potentially proprietary formats such as MS Word and PDF files as input.

- The tool traverses this directory and filters or extracts the text from the files, places a copy of the text in a plain text file, and a snippet (or possibly even all) of the text in a SAS data set.

Text Mining Process Flows

Apply natural language processing algorithms to **parse the documents** and **quantify information** about the terms in the corpus.

Text Parsing Node

- Tokenization - break sentences or documents into terms
- Part of speech identification (noun, verb, etc.)

Text Mining Process Flows

Perform **spell-checking** and refine **synonym lists**. Discover related concepts using **Concept Linking**. Perform **full text search**. **Subset documents and/or terms** for further analysis.

sas.com

Text Mining Process Flows

Input Data → Text Parsing → Text Filter → Text Topic

③

Analyze the documents to **create topics** and assign each document to one or more topics. In addition to derived topics, users can add their own topic definitions.

§sas

Text Mining Process Flows

Input Data → Text Parsing → Text Filter → Text Cluster → Segment Profile

③
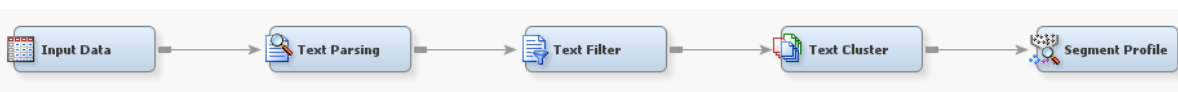
Analyze the documents to **create clusters** and assign each document to a single cluster.

§sas

sas.com

§sas

Text Mining Process Flows

| Input Data | → | Text Parsing | → | Text Filter | → | Text Cluster | → | Segment Profile |
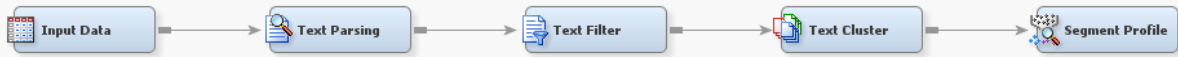
Clusters can be further explored using the Segment Profile node to **identify factors that differentiate data segments from the population**.

§sas

Text Mining Process Flows: Prediction

Text Topic → Decision Tree

Text Cluster → Regression

Text Rule Builder

→ Model Comparison
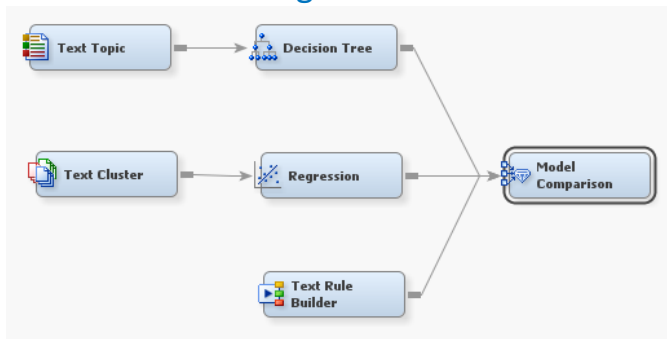
Several methods are available to use the unstructured data to **create predictions**.

§sas

sas.com

§.sas
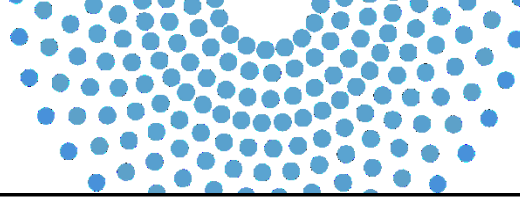
SAS® Text Miner
Example

- Vaccine Adverse Events (VAERS)

- Publicly available dataset from the U.S. Department of Health and Human Services (HHS)

- ~ 25,000 patient records, including physician/patient comments

- Not representative of all recipients of vaccines

- Predict hospitalizations

- This example is covered in detail in "Getting Started with SAS Text Miner"
  http://support.sas.com/documentation/onlinedoc/txtminer/index.html

§sas

SAS® Text Miner
VAERS Example

- **Symptom Text** – text variable containing the reported symptoms

- **Serious** – binary flag that is used as a target to predict serious side effects that resulted in hospitalizations, disability or life-threatening event

- 16 vaccine flag variables for 16 most common vaccines

- Age and gender of patient

- Nominal variables to indicate by whom vaccine was administered and how it was funded

- Pediatric flag

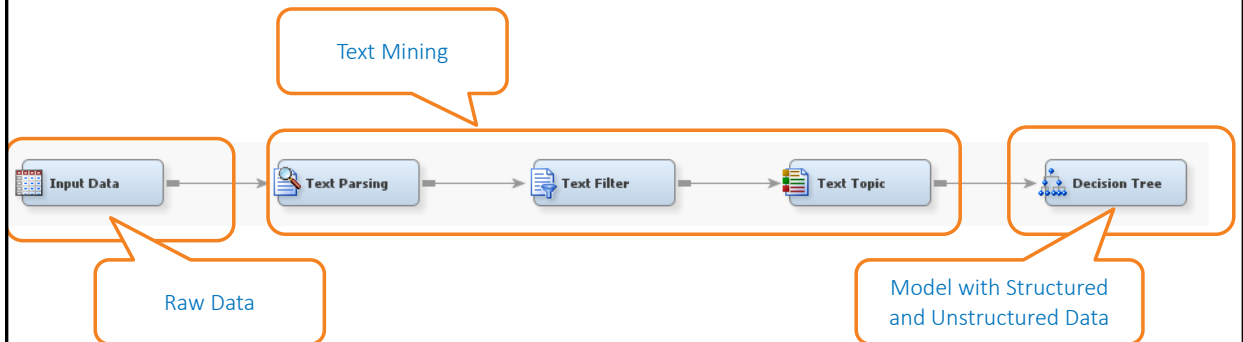- Vaccine count – number of vaccines administered in visit

§sas

§sas

Summary

Text Mining

Input Data → Text Parsing → Text Filter → Text Topic → Decision Tree

Raw Data

Model with Structured and Unstructured Data

Review

Text Mining Discovery
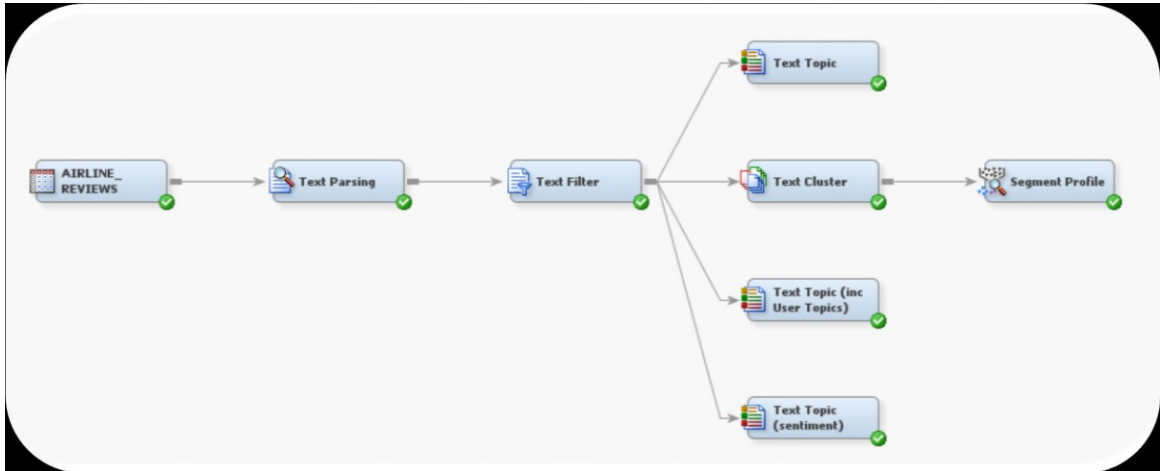
Review

Text Mining Prediction
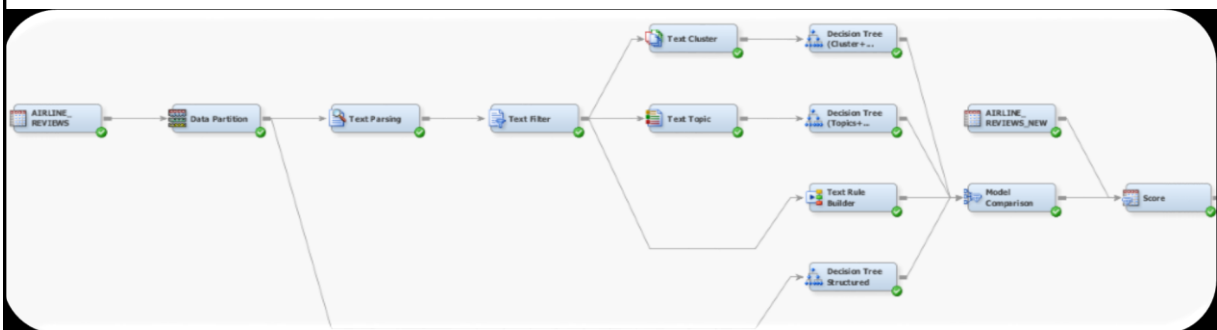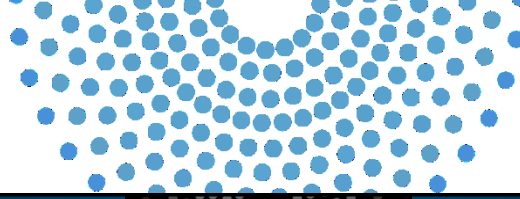
Learning More

Potential Next Steps

- Work through the example in "Getting Started with SAS® Text Miner" - Both the data and the documentation are available online

  http://support.sas.com/documentation/onlinedoc/txtminer/index.html

- Contact SAS Technical Support if you get stuck
  - There is no charge for this – it is included in your SAS software license.
    - http://support.sas.com/techsup/

SAS® Text Miner

http://support.sas.com/documentation/onlinedoc/txtminer/index.html

RESOURCES
- Products & Solutions
- System Requirements
- Samples
- Install Center
- Third-Party Software Reference
- Documentation
  - What's New in SAS
  - Product Index A-Z
  - SAS Viya
  - SAS 9.4
  - SAS Analytical Products 14.2
  - SAS Analytical Products 14.1
  - SAS Analytical Products 13.2
  - SAS Analytical Products 13.1
  - SAS 9.3
  - SAS Analytical Products 12.1
  - SAS 9.2
  - Earlier SAS Releases
- Papers

SAS® Text Miner

| 14.2 and 14.1 | 13.2 and 13.1 | 12.3 and 12.1 |

Click tabs for other versions

SAS Text Miner 13.2
- Getting Started with SAS Text Miner 13.2 PDF | HTML | Japanese PDF
- Example Data for Getting Started with SAS Text Miner 13.2 ZIP
- SAS Text Miner 13.2: Reference Help PDF

SAS Text Miner 13.2 High-Performance Text Mining
- SAS Text Miner 13.2: High-Performance Procedures PDF

SAS Text Miner 13.1
- What's New in SAS Text Miner 13.1 PDF
- SAS Text Miner 13.1: Reference Help PDF

SAS Text Miner 13.1 High-Performance Text Mining
- SAS Text Miner 13.1: High-Performance Procedures PDF

Links

- SAS Text Miner Documentation and Tutorial
  - http://support.sas.com/documentation/onlinedoc/txtminer/index.html
- SAS Technical Support
  - http://support.sas.com/techsup/
- SAS Text Miner Training
  - https://support.sas.com/edu/prodcourses.html?code=TM&ctry=
- Support Website
  - http://support.sas.com/
- SAS Text Miner Technical Forum (Join Today!)
  - https://communities.sas.com/community/support-communities/sas_data_mining_and_text_mining
- SAS Customer Loyalty
  - http://support.sas.com/contact/customerloyalty/
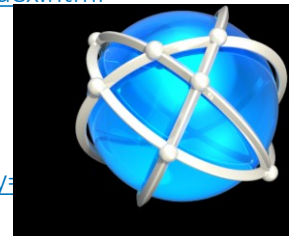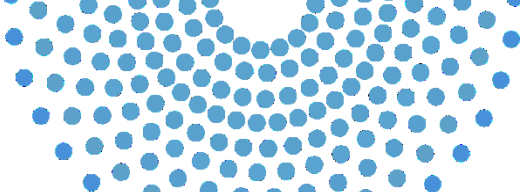
Learn SAS with SAS Education

SAS Education will support you in continual learning to grow your career.

- SAS Training Courses
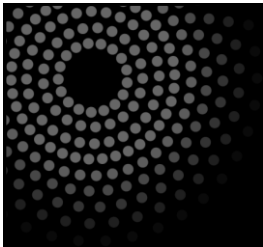 support.sas.com/training

- Get SAS Certified
  support.sas.com/certify

- SAS Books
  support.sas.com/books

Contact SAS Training Customer Service
(800) 727-0025 or training@sas.com

§sas

Questions?
Thank you for your time and attention!

sas.com

CUSTOMER LOYALTY TEAM · Support You Can Count On

§sas
THE POWER TO KNOW.

§sas