



# ASK THE EXPERT

## Getting Started with SAS® Visual Text Analytics

George Fernandez  
George.Fernandez@sas.com





# George Fernandez

Senior Analytical Training Consultant, SAS

George has more than 30 years of experience in teaching advanced analytics courses. He leads training in predictive modeling, forecasting, text analytics, model management and survey analysis.

# AGENDA

0 *Visual Text Analytics – A gentle introduction*

0 *Visual Text Analytics – Steps*

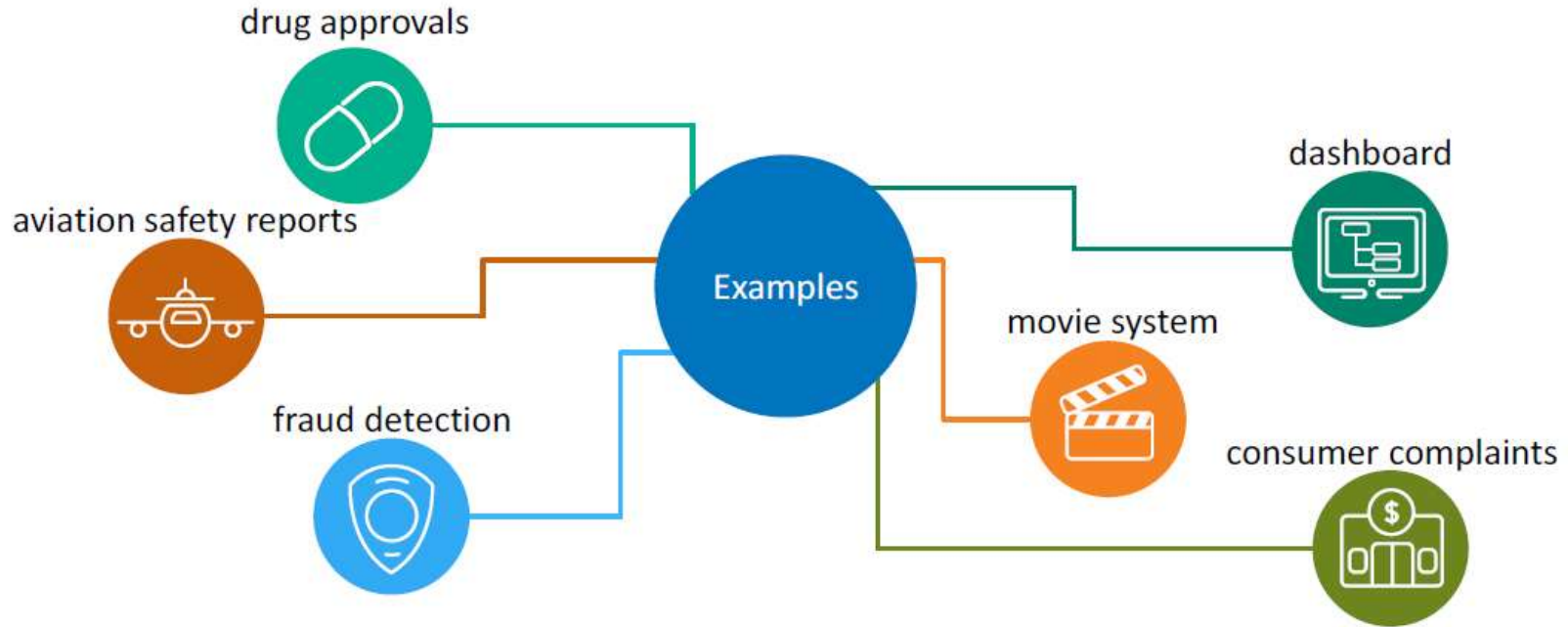
0 *Visual Text Analytics – Features*



# SAS Visual Text Analytics

SAS Visual Text Analytics combines the power of **natural language processing, machine learning,** and **linguistic rules** in an end-to-end analytics framework with the goal of creating and refining models for extracting insights from textual data.

# Text Analytics: Examples



## Why Is SAS Visual Text Analytics Important?

SAS Visual Text Analytics automatically converts unstructured data into meaningful insights. Considering the staggering amount of unstructured data that's generated every day, automation is critical to fully analyze text and speech data efficiently.



- Sepsis Indicators
- Painkiller Addict
- Environmental Influences



- Is product meeting quality expectations?



- Terrorism
- Cyber Bullying
- At-Risk Persons



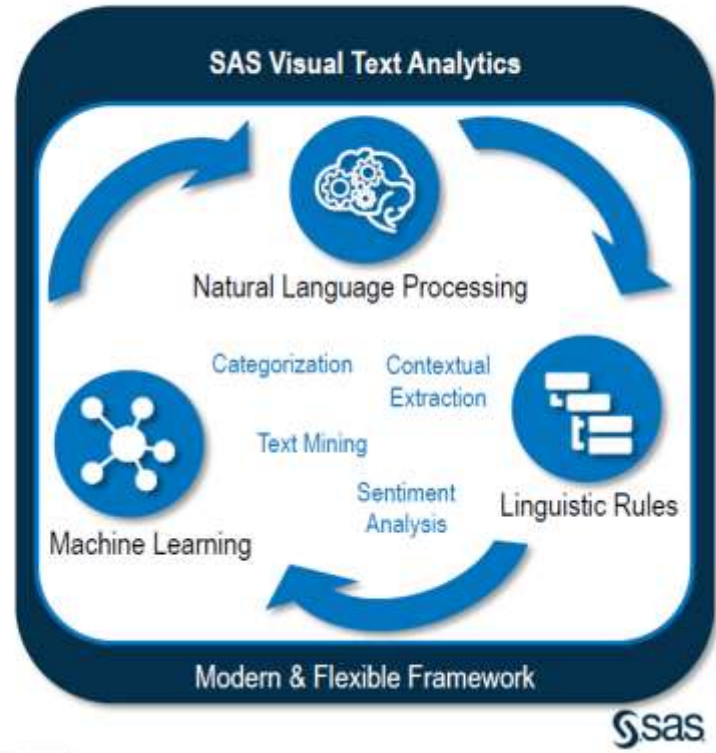
- What amenities influence bookings?

## What Is SAS Visual Text Analytics?

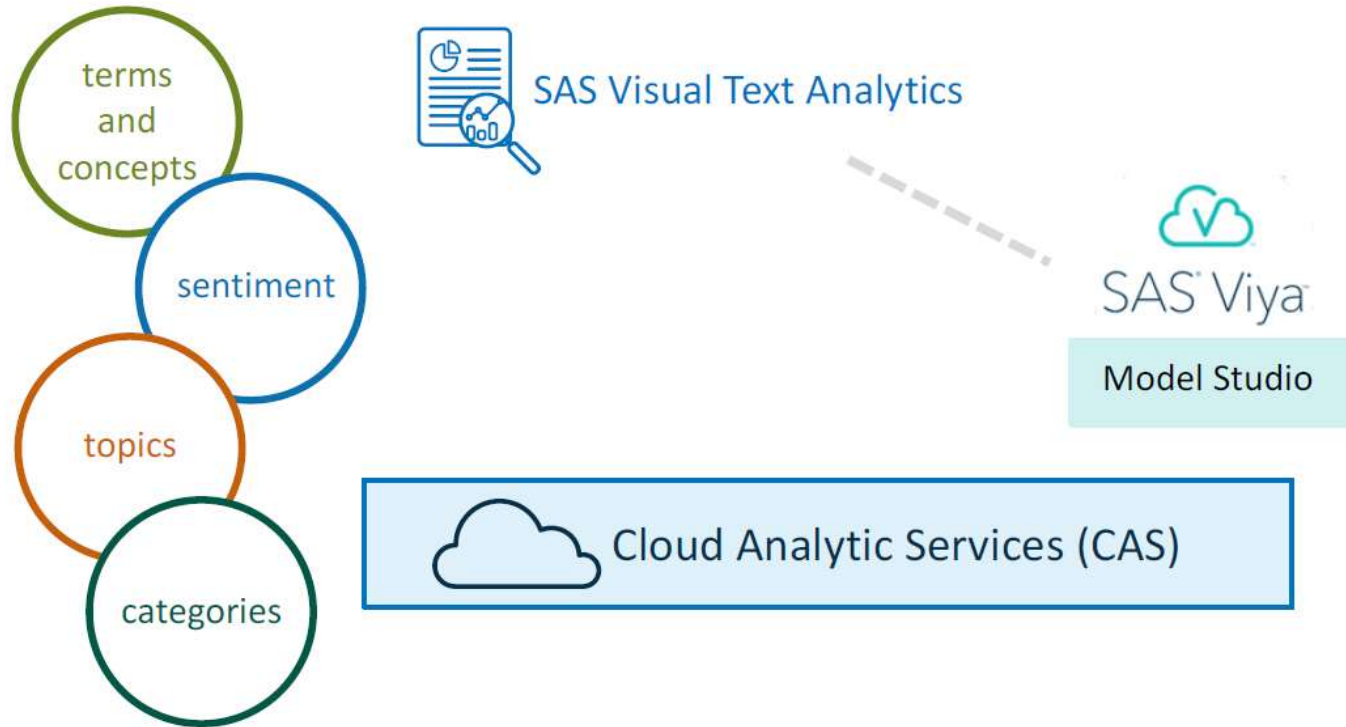
SAS Visual Text Analytics is the SAS offering designed to effectively extract insights from unstructured data in large scale.

Offered on the SAS Viya architecture, Visual Text Analytics combines the power of

- natural language processing (NLP)
- machine learning (ML)
- linguistic rules (algorithmically derived or human supplied).



# SAS Visual Text Analytics: Big Picture





# Global Language Support



33 Languages

|          |            |            |
|----------|------------|------------|
| Arabic   | Greek      | Portuguese |
| Chinese  | Hebrew     | Romanian   |
| Croatian | Hindi      | Russian    |
| Czech    | Hungarian  | Slovak     |
| Danish   | Indonesian | Slovene    |
| Dutch    | Italian    | Spanish    |
| English  | Japanese   | Swedish    |
| Farsi    | Kazakh     | Tagalog    |
| Finnish  | Korean     | Thai       |
| French   | Norwegian  | Turkish    |
| German   | Polish     | Vietnamese |

# Why Use SAS Visual Text Analytics?



eliminates manual reviews

removes manual bias errors

regulates consistent classification

removes requirement of training corpus

# Unstructured Text to Structured Data

unstructured

text data



- information overload
- cognitive bias
- sampling bias
- automatic knowledge or intelligence discovery
- multiple languages and cultures
- limited information lifespan

- document collections are massive
- text parsing is computationally intensive
- decisions about strategic versus tactical time frames
- real-time knowledge capture; real-time deployment problematic

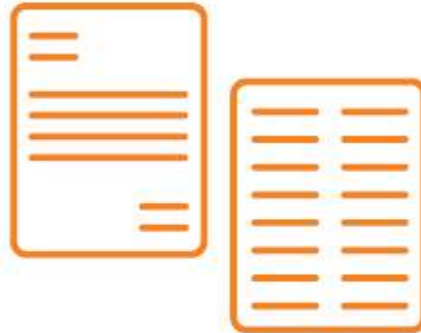


structured

# Text Analytics Projects Requirements



well-defined  
goal

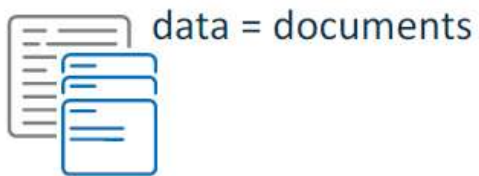


document  
collection



metrics to  
identify success

# Text Analytics



Text Analytics Project



natural language processing  
(NLP)

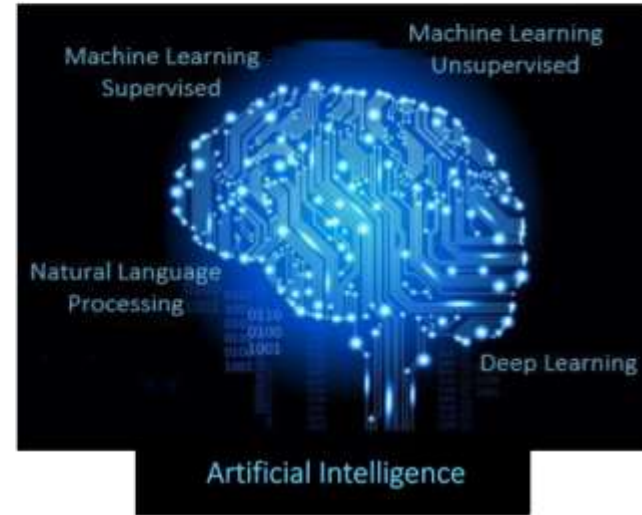


collection of dictionaries

# What Is Natural Language Processing (NLP)

Branch of artificial intelligence that helps computers

- understand
- interpret, and
- manipulate human language.

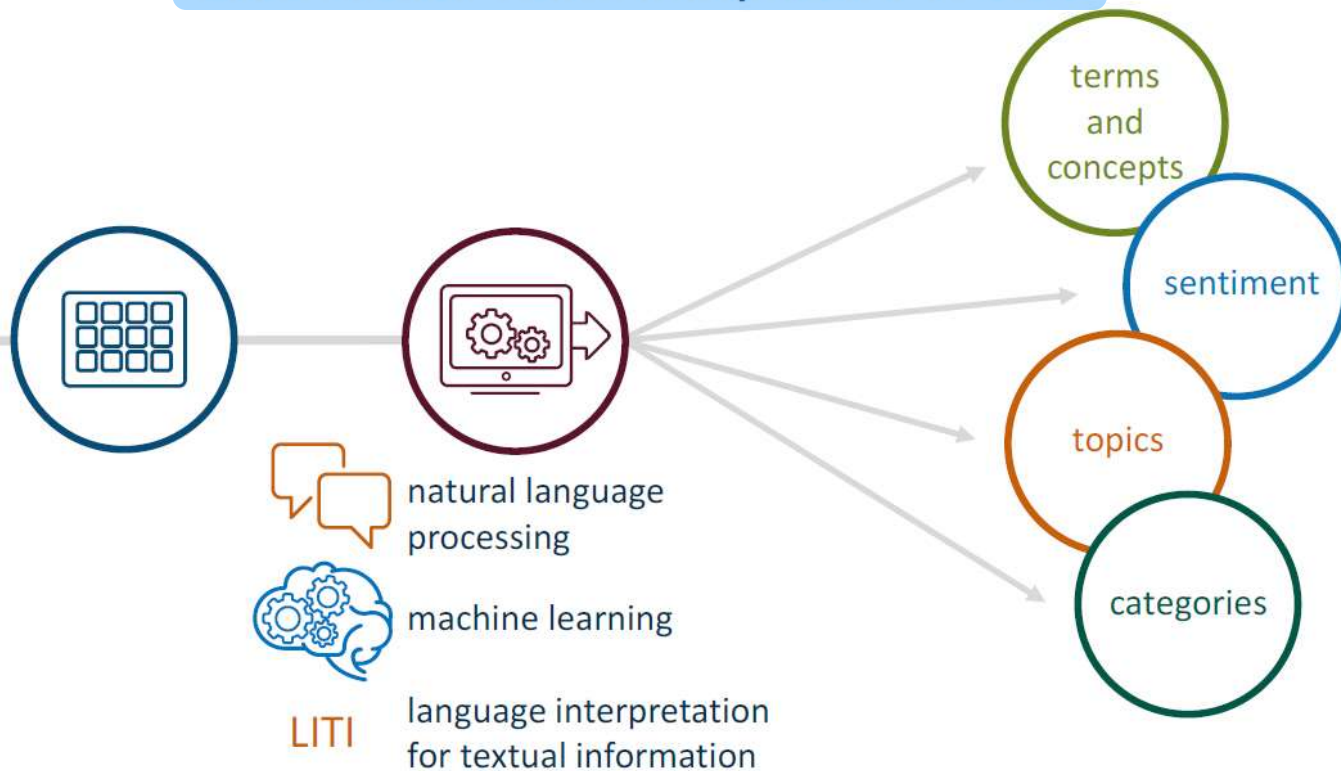


NLP helps **resolve ambiguity** in language and **adds useful numeric structure** to the data for many downstream applications, such as speech recognition or text analytics.



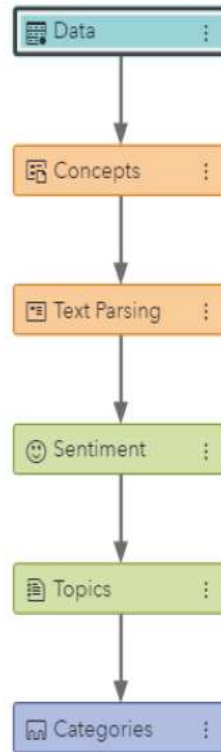


# SAS Visual Text Analytics: Process

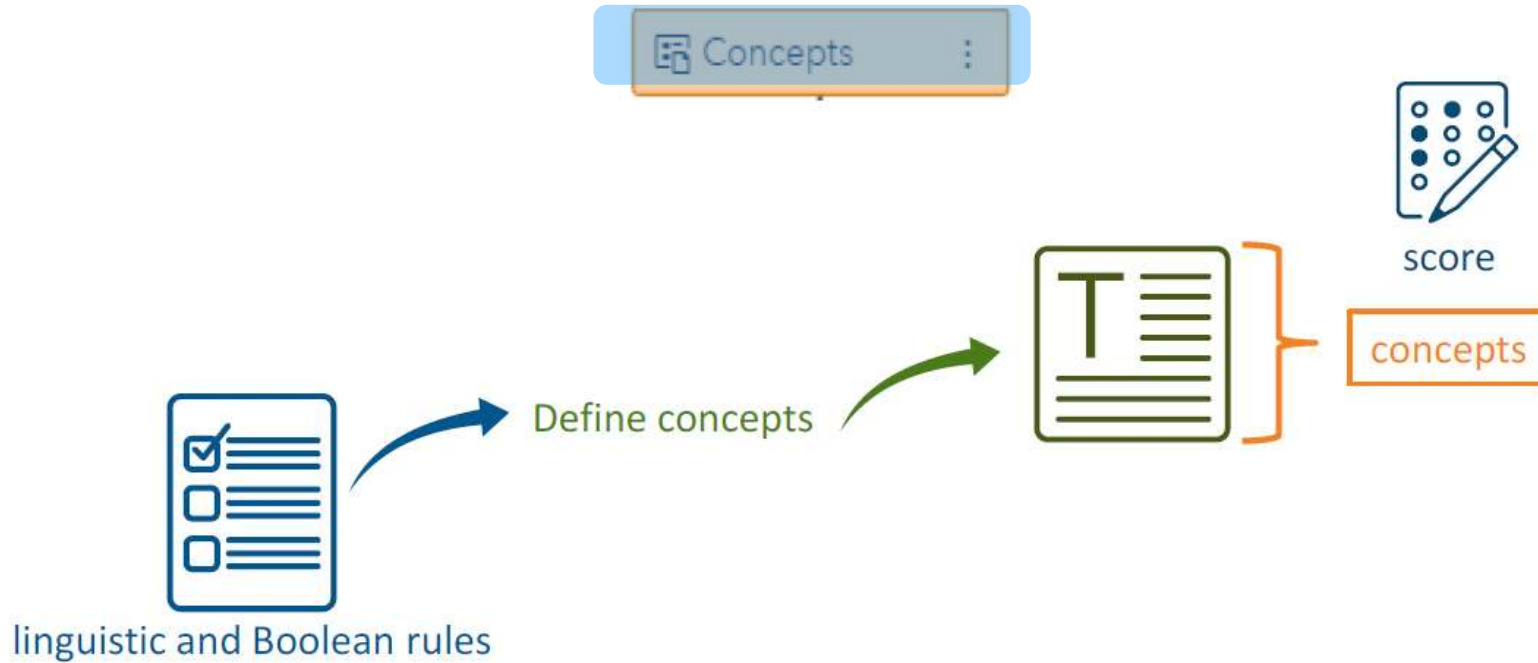




# SAS Visual Text Analytics: Default Pipeline



# Text Analytics Components



# Concept Extraction



Input text data  
(Twitter data, reports, email,  
news, or forum messages)



Parse defined  
concepts  
(linguistic and  
Boolean rules)



Output results  
(scored using score  
code; satisfy concept  
definitions)

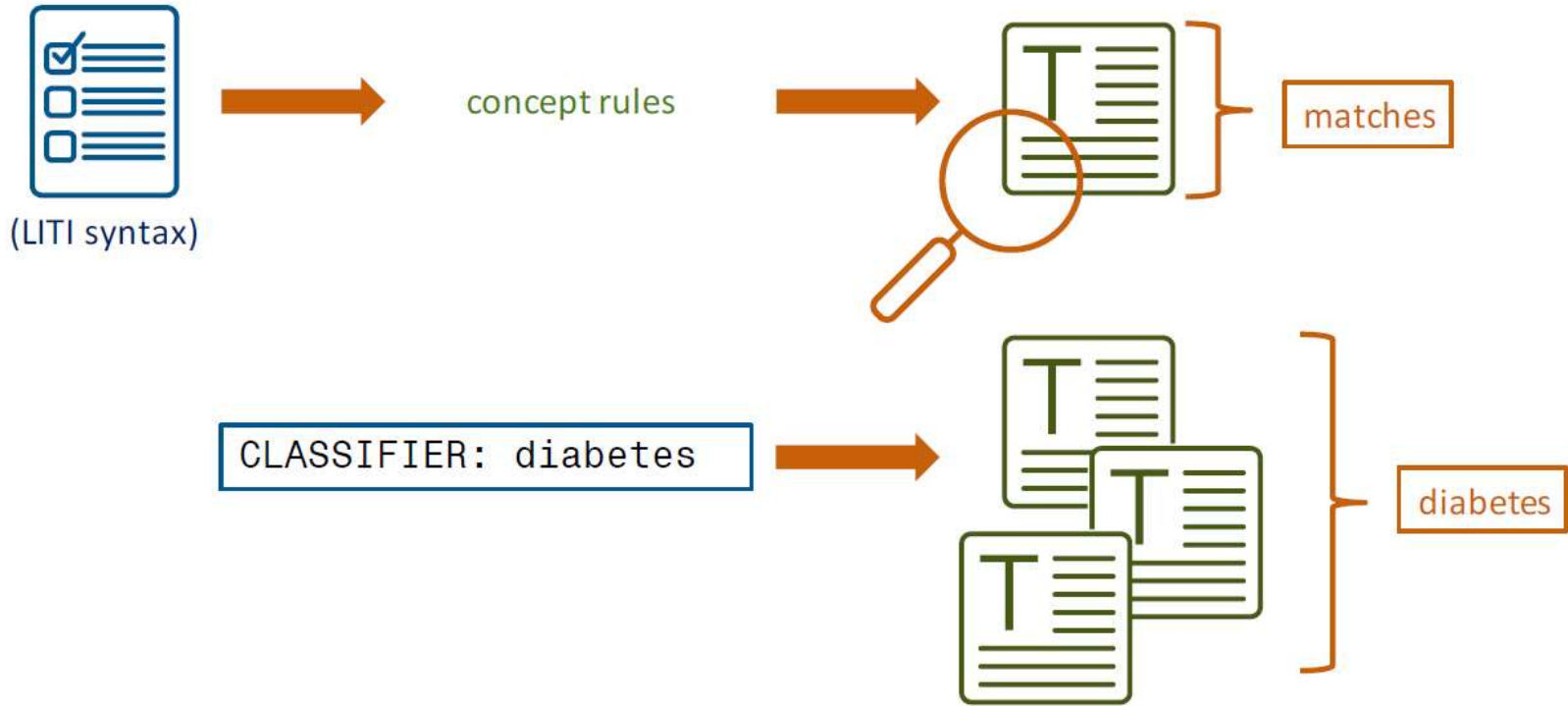


Research and analysis

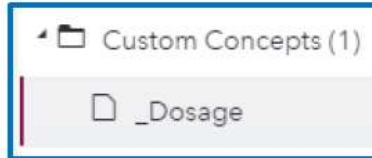
## Concepts

- Measures – 50 mg...
- Persons – John...
- Dates – Monday...
- Organizations – SAS...

# Introduction to Concept Rules



# Text Analytics Components

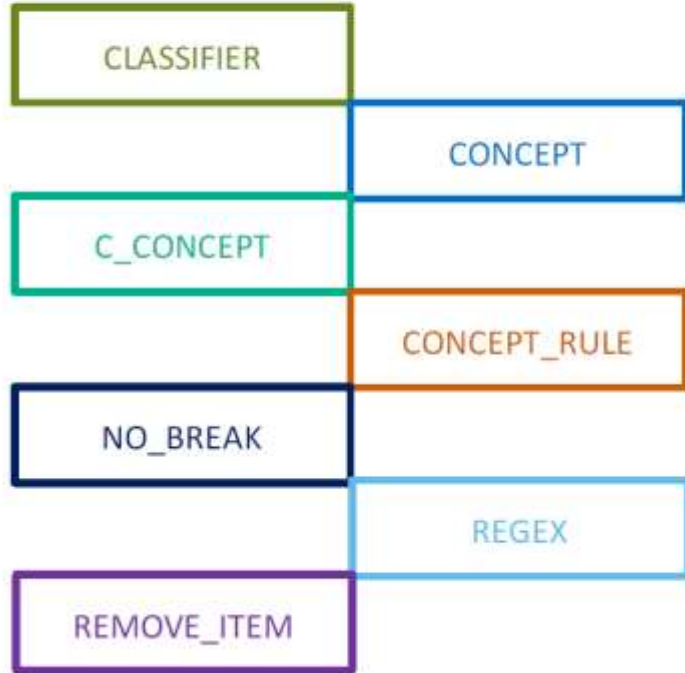


|   |                                  |
|---|----------------------------------|
| 1 | REGEX: [0-9]+[\.][0-9]+\s?mg\.-? |
| 2 | REGEX: [\d]+\s?mg\.-?            |

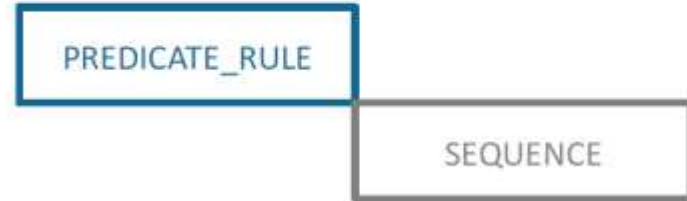
I have been on **10mg** prexifan for 6 months along with **2000mg** noderall **1mg** gemulex and fortifex when needed.

# Concept Rules

Extracting Concepts:

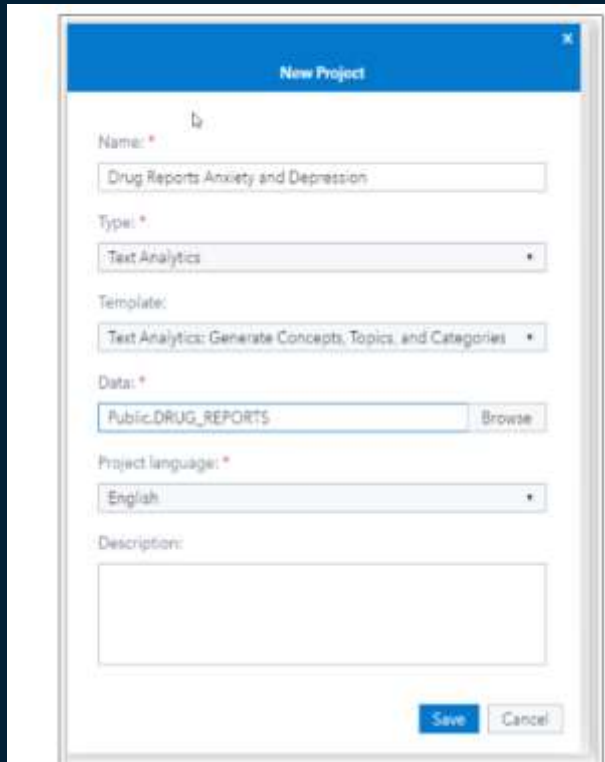


Extracting Facts:



5

# SAS Visual Text Analytics Project creation

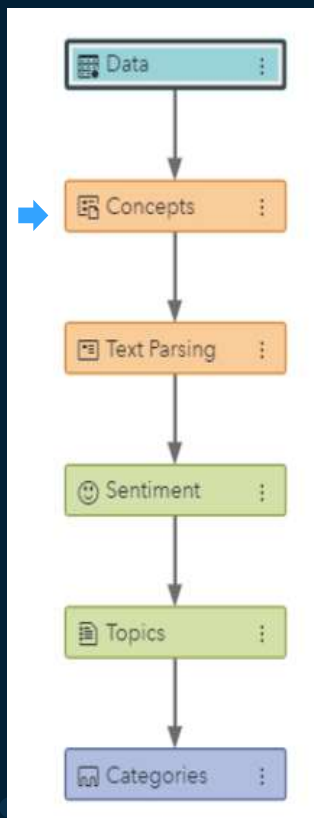


The image shows a 'New Project' dialog box in SAS Visual Text Analytics. The dialog has a blue header with the title 'New Project' and a close button. Below the header, there are several fields for project configuration:

- Name:** A text input field containing 'Drug Reports Anxiety and Depression'.
- Type:** A dropdown menu with 'Text Analytics' selected.
- Template:** A dropdown menu with 'Text Analytics: Generate Concepts, Topics, and Categories' selected.
- Data:** A text input field containing 'Public.DRUG\_REPORTS' and a 'Browse' button to the right.
- Project language:** A dropdown menu with 'English' selected.
- Description:** A large empty text area.

At the bottom right of the dialog, there are two buttons: 'Save' (highlighted in blue) and 'Cancel'.

# SAS Visual Text Analytics Steps – Concept extraction

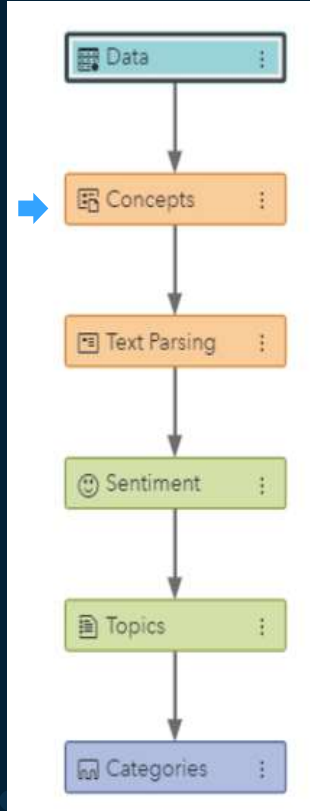


The display below illustrates the `_DOSAGE` concept rules and the matched documents.

The screenshot shows the SAS Visual Text Analytics interface for a project titled 'Drug Report Anxiety and Depression'. The 'Concepts' tab is active, showing a list of concepts: 'Predefined Concepts (9)' and 'Custom Concepts (4)'. The 'Custom Concepts' list includes `_DOSAGE`, `_MEDICATION`, `_PRESCRIPTION`, and `_SIDE_EFFECTS`. The `_DOSAGE` concept is selected, and its rules are displayed in the 'Edit Concept' pane. Two rules are shown: `1 REGEX: [0-9]+\.[\.[0-9]+\s?mg\%?` and `2 REGEX: \d+\s?mg\%?`. The first rule is highlighted. Below the rules, a 'Documents' section shows 'Text Sample Text' with a search bar and a 'Matched (300 of 1414)' indicator. A sample document is displayed, showing text with dosage values like '225mg' and '150mg' highlighted in blue. A blue arrow points to the 'Concepts' tab, and numbered callouts (1-5) highlight specific elements: 1 points to 'Predefined Concepts', 2 to 'Custom Concepts', 3 to the `_DOSAGE` concept, 4 to the first regex rule, and 5 to the 'Matched' indicator.



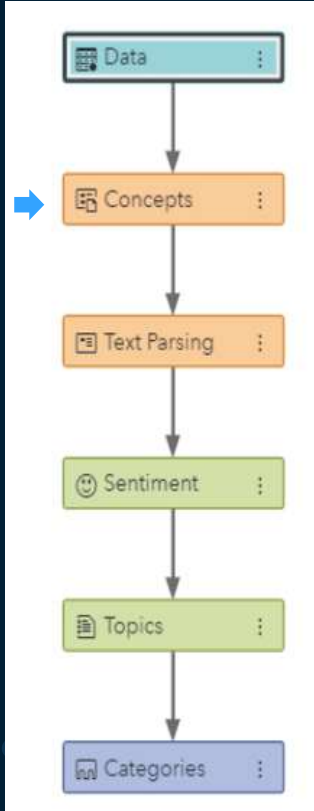
# Open Concept Node



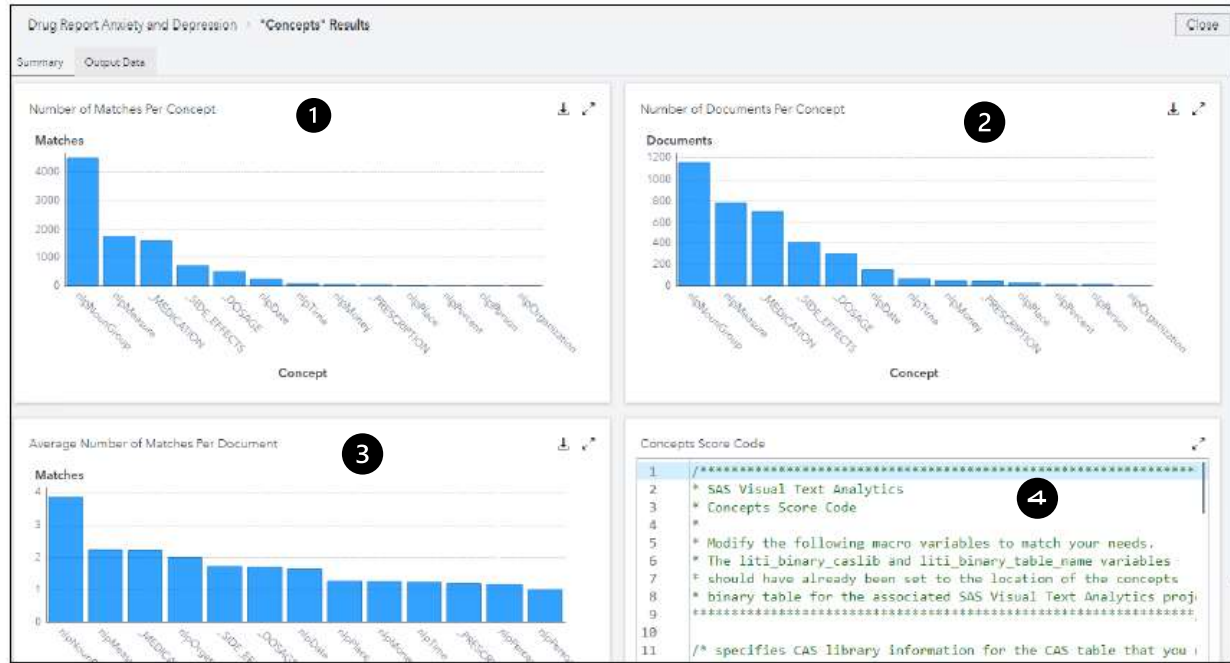
The screenshot shows the SAS Concepts interface. On the left, a list of predefined concepts includes `_DOSAGE`, `_MEDICATION`, `_PRESCRIPTION`, and `_SIDE_EFFECTS`. A blue arrow points to the `_MEDICATION` concept. On the right, a search results window displays a list of concepts with their classifier codes, such as `CLASSIFIER:Abidal`. Below the list, a search bar shows the query `Abidal, Ecstapin` and the results `All (1414) Matched (562 of 704)`. A search bar at the bottom of the window shows the query `Abidal, Ecstapin` and the results `Highlight Concept matches Search matches`.

You see that 562 out of 704 documents matched the terms *Abidal* or *Ecstapin*.

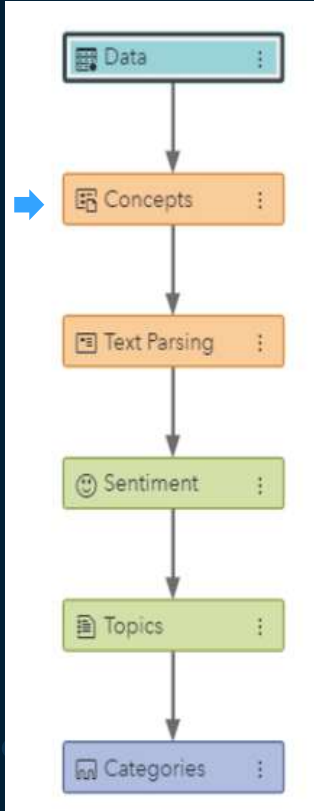
# Concept node results



Select the **Concept** node and click **Results** option to see the results.



The Concept results window displays three bar charts: Number of Matches Per Concept, Number of Documents Per Concept, and Average Number of Matches Per Document.



Click **Output Data** (to the right of the Summary at the upper left). Select **Output Tables** ⇨ **Transactional** ⇨ **View Output Data**.

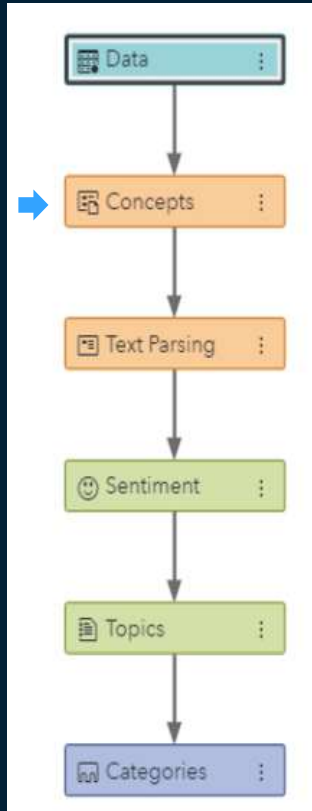
Drug Report Anxiety and Depression - "Concepts" Results

Summary Output Data **1**

**2** Output Tables: Transactional **3**

| __uniqueid__ | Text  | Full Path     | Concept Name  | Keywords   | Concept Type | ID   |
|--------------|---|---------------|---------------|--|--------------|------|
| 401          | this med does perk you up, but after 4 weeks of   | nlpMeasure    | nlpMeasure    | 5-10 days   4 weeks   4-5 days   a month                   | Concept      | 0004 |
| 401          | this med does perk you up, but after 4 weeks of   | nlpNounGroup  | nlpNounGroup  | family support   cautious going   even though i            | Concept      | 0004 |
| 401          | this med does perk you up, but after 4 weeks of   | _MEDICATION   | _MEDICATION   | fortifex   prelixan  | Concept      | 0004 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | _SIDE_EFFECTS | _SIDE_EFFECTS | dizziness  | Concept      | 1094 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | _DOSAGE       | _DOSAGE       | 90mg   30mg   60mg   60mg   90mg   30mg.                   | Concept      | 1094 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | nlpMeasure    | nlpMeasure    | 30mg   60mg   a day   90mg   3 days   90mg   9             | Concept      | 1094 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | nlpNounGroup  | nlpNounGroup  | usual side   60mg daily   slower rate   brain zaps   final | Concept      | 1094 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | nlpPlace      | nlpPlace      | My Dr.   | Concept      | 1094 |
| 902          | I was on this for 9 months, up to 90mg a day. Had | _MEDICATION   | _MEDICATION   | Abidal   | Concept      | 1094 |

# Exploring and visualize the scored data in SAS VA



Click Explore and Visualize (🔍).

Click the right arrow by cas-shared-default, and select library Public.

### Explore and Visualize Output Data

Filter

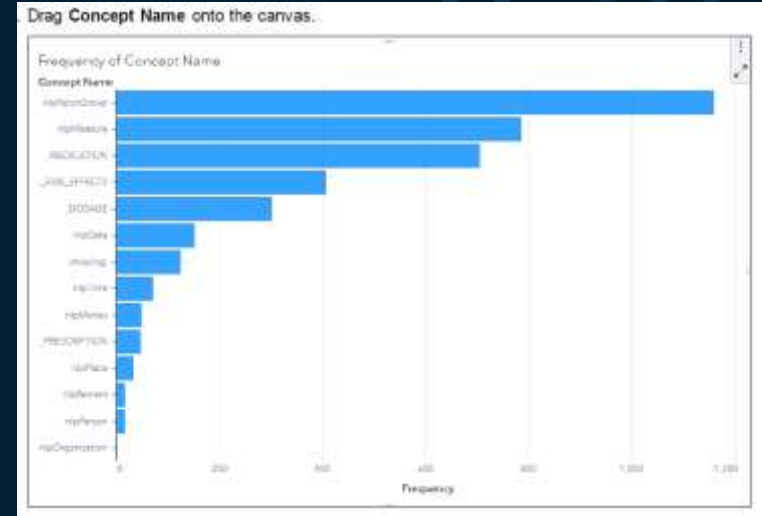
Data Sources: cas-shared-default

- modelPerformanceData PATH
- Models PATH
- Public PATH**
- Samples PATH

Table name: \*

Drug Report Anxiety and Depression\_Concepts\_Transactional

Explore and Visualize Cancel





# Scoring concept models

```
*drug_conceptscore.sas *
Submit | Cancel | History | Address Snippet | Nov 9, 2011, 5:52:55 PM

Code | Lng | Width
-----|---|-----
1 | | |
7 | * SAS Visual Text Analytics
3 | * Concepts Score Code
4 | *
5 | * Modify the following macro variables to match your needs.
6 | * The liti_binary_caslib and liti_binary_table_name variables
7 | * should have already been set to the location of the concepts
8 | * binary table for the associated SAS Visual Text Analytics project.
9 | *
10 |
11 | /* cas library information for cas table containing the data set you would like to score */
12 | %let caslib_name="PUBLIC";
13 |
14 | /* the cas table you would like to score */
15 | %let input_table_name = "ORIGINALPORT";
16 |
17 | /* the column in the cas table that contains the unique id */
18 | %let key_column = "filename";
19 |
20 | /* the column in the cas table that contains the text data to score */
21 | %let document_column = "content";
22 |
23 | /* cas library information for output cas tables to produce */
24 | %let output_caslib_name = "CASUSER";
25 |
26 | /* the concepts output cas table to produce */
27 | %let output_concepts_table_name = "drug_out_concepts";
28 |
29 | /* the facts output cas table to produce */
30 | %let output_facts_table_name = "out_facts";
31 |
32 | /* cas library information for liti binary table... should have been set to your Text Analytics project's cas library */
33 | %let liti_binary_caslib = "Analytics_Project_d8d41c77-e7b4-42ef-9a31-1075995e34ed";
34 |
35 | /* cas table name for liti binary table... should have been set to your Text Analytics project's concept node's model table */
36 | %let liti_binary_table_name = "da72c06e58f2c740165b411c910017_CONCEPT_BINARY";
37 |
38 | /* hostname for cas server */
39 | %let cas_server_hostname = "server.demo.sas.com";
40 |
41 | /* port for cas server */
42 | %let cas_server_port = 5570;
43 |
44 | /* create a session */
45 | %cas session host=%cas_server_hostname port=%cas_server_port uid=alex-hawkes3 uidid ;
46 | %libname sasccas1 cas %cas_server_hostname %cas_server_port uid=alex-hawkes3 uidid ;
47 |
```

# SAS Visual Text Analytics: Term Extraction

Concepts



Parse defined concepts

Text Parsing



General language elements

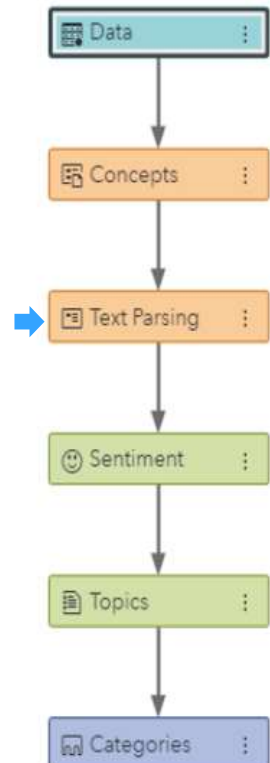
## NLP Tools

- parsing
- tokenization
- part-of-speech (POS) tagging
- synonym detection
- stemming



# Text parsing output

Select the Text Parsing node and select Open.



Drug Report Anxiety and Depression - Text Parsing - Manage Terms

Kept Terms (1511)

| Term                                | Role | Documents | Frequency |
|-------------------------------------|------|-----------|-----------|
| <input type="checkbox"/> not        | ADV  | 658       | 1174      |
| <input type="checkbox"/> take       | V    | 676       | 1104      |
| <input type="checkbox"/> depression | N    | 492       | 816       |
| <input type="checkbox"/> year       | N    | 436       | 561       |
| <input type="checkbox"/> feel       | V    | 379       | 541       |
| <input type="checkbox"/> work       | V    | 379       | 487       |
| <input type="checkbox"/> drug       | N    | 342       | 486       |
| <input type="checkbox"/> day        | N    | 344       | 479       |

Dropped Terms (6711)

| Term                          | Role | Documents | Frequency |
|-------------------------------|------|-----------|-----------|
| <input type="checkbox"/> I    | PRO  | 1081      | 5003      |
| <input type="checkbox"/> be   | V    | 1127      | 4155      |
| <input type="checkbox"/> have | V    | 1069      | 2849      |
| <input type="checkbox"/> and  | CONJ | 1028      | 2747      |
| <input type="checkbox"/> the  | DET  | 903       | 2558      |
| <input type="checkbox"/> to   | PPOS | 895       | 2383      |
| <input type="checkbox"/> it   | PRO  | 878       | 2081      |
| <input type="checkbox"/> a    | DET  | 792       | 1655      |

Documents: All (1414) Matched

**DrugReport**

This medication made me gain 40 pounds it has been 2 years and I have only lost 10 pounds. Beware and watch your weight.

causing extreme anger, to the point that my family has become afraid of me. Doing things I would have considering doing before, like challenging the police, starting arguments, really, really wanting to beat the crap out of someone. Enough to the point that it scares me. But my doctor at this time refuses to take me off this medication. I am scared I will end up in jail for severely hurting someone.

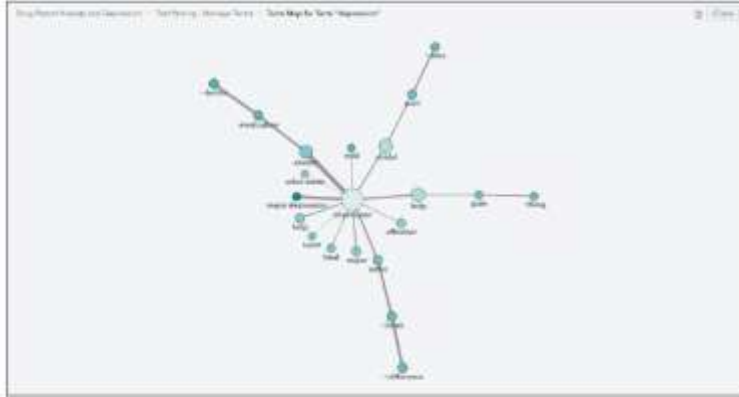
with several questions are extremely uncomfortable if medicine doesn't work.

Document 1 of 1414



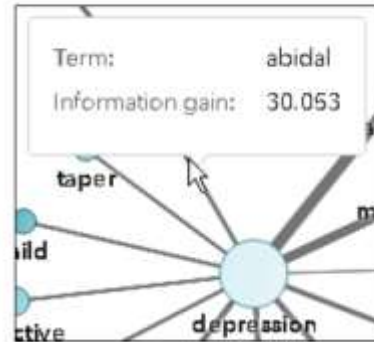
# Term map in the text parsing node

Select the kept term **depression**. Click  (Term map).



A total of 492 documents contain the term *depression*.

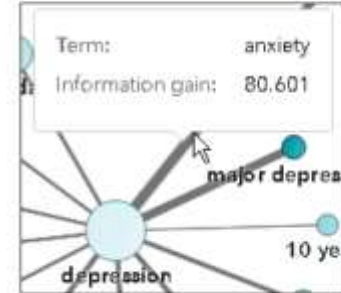
The line that connects *depression* and *abidal* has a thickness that is relative to the information gain (IG). You can right-click the line to display the IG value.



A total of 292 documents contain the drug name *abidal*, and 142 of those *abidal* documents also contain the word *depression*.



The largest information gain is associated with *anxiety*.



# Sentiment Analysis



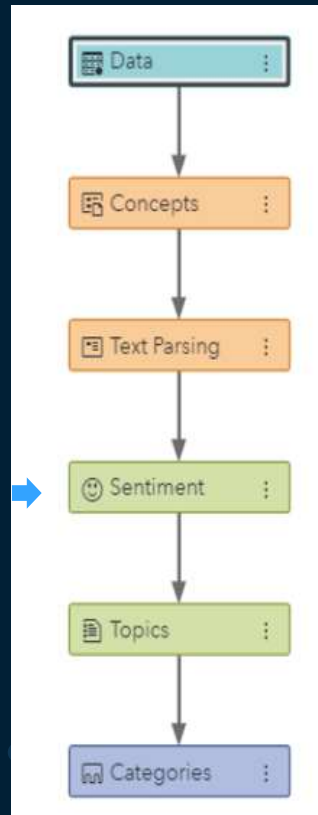
use proprietary rules  
to imply sentiment

accommodate user-created  
sentiment models

assign sentiment score to reflect  
positive, neutral, or negative

transfer sentiment scores  
to successor nodes

# System generated sentiment scores



The new anxiety topic identifies 204 documents.

The screenshot displays the SAS interface. On the left, a list of topics is shown, with 'anxiety, anxious, anxiety' selected. On the right, a detailed view of the 'anxiety' topic is shown, including a table of terms and their relevance scores, and a list of documents with their sentiment scores.

| Terms   | Relevance | Role | Documents | Frequency |
|---------|-----------|------|-----------|-----------|
| anxious | 1.000     | A    | 16        | 16        |
| anxiety | 1.000     | N    | 4         | 4         |
| anxiety | 1.000     | N    | 187       | 231       |

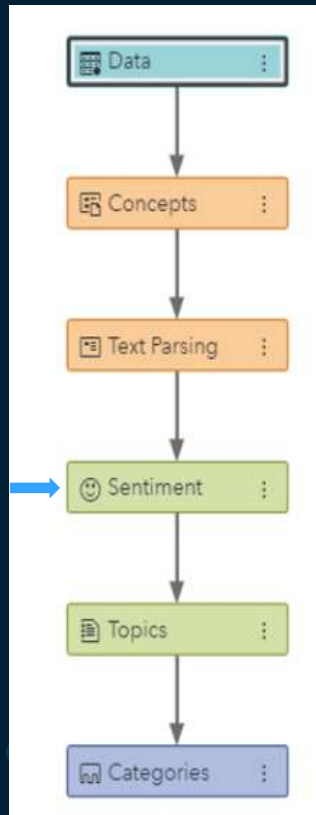
  

| Documents   | Sentiment | Relevance |
|---|-----------|-----------|
| ...fight depression and anxiety   | 😊         | 0.917     |
| ...the time from anxiety to irregularity line, however I wonder if I should be taking pills for depression and anxiety for daily anxiety, something like Quiescal or Xanax for the anxiety. | 😊         | 0.710     |
| ...more depression more anxious causing more anxiety attacks  | 😞         | 0.658     |
| ...major depression and anxiety until I started taking Abidal. I did put on a few pounds because my appetite increased but it was better than feeling depressed and anxious.                | 😊         | 0.610     |

Finally, a user-defined, combined depression-and-anxiety-based topic is created by merging the three depression-based and anxiety-based user-defined topics. Select the three User topics and click **Merge topics**. A fourth User topic is created.

Click **Run Node**.

# System generated sentiment scores

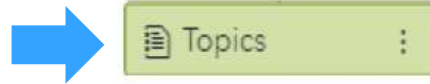


Examine each promoted topic category rules.

The screenshot shows the SAS Text Analytics interface for 'Drug Reports Anxiety and Depression'. The 'Categories' panel on the left lists several categories, including 'depression, anxiety, -depress, depressed, -depr...', 'anxiety, anxious, anxy', and 'depression, anxiety, -depress, depressed, -depr...'. The main window displays a list of documents with sentiment scores and relevancy scores. A blue arrow points to the 'Sentiment' column header.

| Document   | Text Service Text   | Sentiment | Relevancy |
|------------|---|-----------|-----------|
| All (1414) | Matched (557 of 1414)   |           |           |
| DrugReport | ... fit with some <b>depression</b> few months ago. Wow, that's no fun. I will never say to someone they snap out of it. Life works in funny ways sometimes. Anyway, was forced to get help with the <b>depression</b> . Started with low-dose 37.5 for 10 days, then 75mg for 14 days and to date at 75 + 37.5 = 112.5mg. The hardest part was waiting for that split. It took almost 6th weeks. I wanted to quit many times, but stuck with it. Weird side effects along the way but he...              | 😊         | 12,000    |
| ...        | ... to treat severe <b>depression</b> and anxiety with heart palpitations. My doc also put me on a steroid to control the palpitations, which is a nice combo because it reduces your blood pressure and prevents migraines. After year I feel better. Like many of you who complained about wanting to get off it, and slowly tapered off. What I have learned in the last year of horrible <b>depression</b> trying to do it on my own with diet, exercise, journaling, and the steroid is that...      | 😊         | 9,000     |
| ...        | ... grew for my <b>depression</b> , but the thing that worked best for me was keeping busy and getting involved in a project out of the home. Productivity increases your confidence immensely. Also good doctors that know the newest drugs on the market that are found to allow <b>depression</b> . My doctor if I converted that MIBIPLEX, a anti-calcium medicine, helped sustain a better mood in patients. A combination of the right medicines will help alleviate <b>depression</b> , but you... | 😊         | 8,000     |
| ...        | ... for poor part in <b>depression</b> that escalated to <b>depression</b> I tried 3 other types before escitalopram 150. This one is the best so far I am no longer <b>depressed</b> but I have not been able to lose weight like I want to. I have not had any side effects and I have been taking it about a year now. I do get tired more now but not like I used when I was <b>depressed</b> .   | 😊         | 8,000     |

# Automatically Discover Hidden Topics: Machine Learning



automatically  
discovers core topics

relevant documents ranked  
per topic

topics merged or split

document-level  
sentiment included

derived topics modified  
to create custom

# Modifying Topics discovery settings

The image illustrates the process of modifying Topics discovery settings. It starts with a workflow diagram on the left, showing a sequence of steps: Data, Concepts, Text Parsing, Sentiment, Topics, and Categories. A blue arrow points from the 'Topics' step to the first screenshot of the 'Topics' configuration window.

The first screenshot shows the 'Topics' configuration window with the following settings:

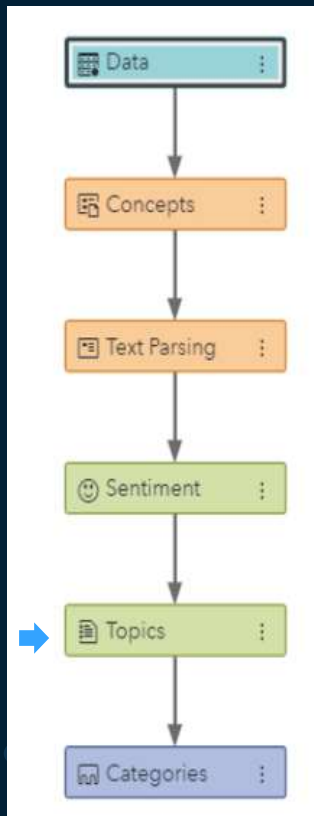
- Description: Assigns documents to topics.
- Topic Discovery:
  - Automatically determine number of topics
  - Maximum topics: 25
  - Term density: 1 (slider from 0 to 10)
  - Document density: 1 (slider from 0 to 10)

A second blue arrow points from the 'Topics' step to the second screenshot of the 'Topics' configuration window, showing the modified settings:

- Description: Assigns documents to topics.
- Topic Discovery:
  - Automatically determine number of topics
  - Maximum topics: 20
  - Term density: 3 (slider from 0 to 10)
  - Document density: 2 (slider from 0 to 10)

Blue arrows also point to the 'Maximum topics', 'Term density', and 'Document density' input fields in both screenshots, indicating the specific settings being modified.

# System generated Topics



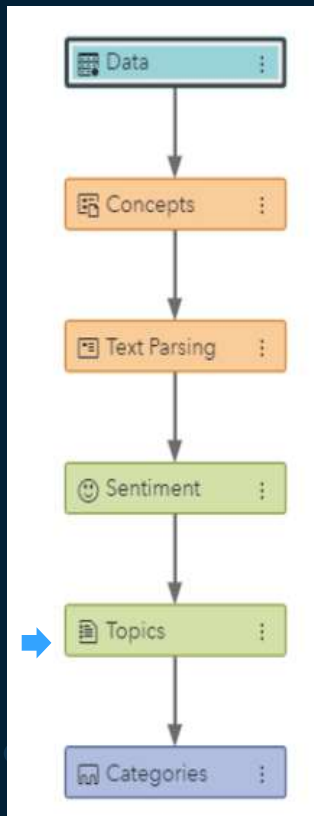
Topics (20)

| <input type="checkbox"/> | Topic  | Created by | Documents ↓ |
|--------------------------|--|------------|-------------|
| <input type="checkbox"/> | side, +side effect, -effect, no, +make             | System     | 90          |
| <input type="checkbox"/> | +withdrawal, +symptom, horrible, off of, +drug     | System     | 79          |
| <input type="checkbox"/> | abidal, +pain, +make, also, +medicine              | System     | 70          |
| <input type="checkbox"/> | mg, +day, a day, very, +dose                       | System     | 70          |
| <input type="checkbox"/> | +try, best, many, +other, meds                     | System     | 64          |
| <input type="checkbox"/> | +life, +save, +change, +drug, back                 | System     | 64          |
| <input type="checkbox"/> | +medication, very, anxiety, +good, +well           | System     | 64          |
| <input type="checkbox"/> | +day, +sleep, prexfan, only, +night                | System     | 62          |
| <input type="checkbox"/> | weight, gain, weight gain, +cause, med             | System     | 61          |
| <input type="checkbox"/> | +pound, +gain, +lose, +month, +drug                | System     | 60          |
| <input type="checkbox"/> | +side, +effect, sexual, no, sexual side            | System     | 60          |
| <input type="checkbox"/> | +depress, still, +medicine, +think, +mood          | System     | 55          |
| <input type="checkbox"/> | +depression, anxiety, ecstapin, panic, +suffer     | System     | 53          |
| <input type="checkbox"/> | +problem, +find, vary, +medicine, effective        | System     | 53          |
| <input type="checkbox"/> | +feel, better, abidal, +make, again                | System     | 51          |
| <input type="checkbox"/> | +help, really, +drug, not, +happy                  | System     | 49          |
| <input type="checkbox"/> | +work, not, great, well, +well                     | System     | 49          |
| <input type="checkbox"/> | suicidal, +thought, +suicidal thought, +make, very | System     | 45          |
| <input type="checkbox"/> | +start, just, +week, med, +know                    | System     | 39          |
| <input type="checkbox"/> | blood, pressure, blood pressure, anxiety, +high    | System     | 35          |





# Creating user defined topics by merging other topics



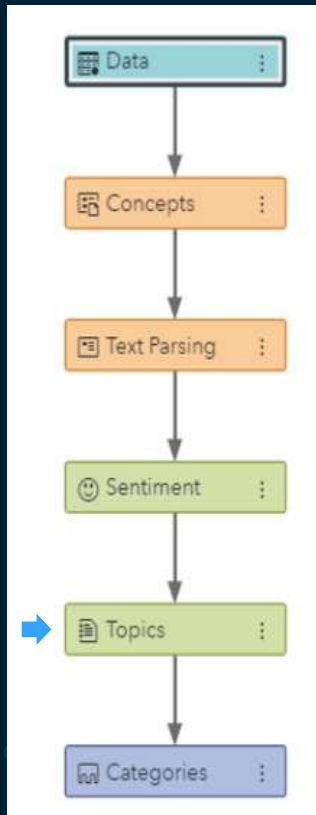
Increasing the maximum number of topics to 20 resulted in a new depression-based topic (**+depress, still, +medicine, +think, +mood**) with 55 matched documents and 22 matched terms. Thus, increasing the number of maximum topics did not help identify more depression- and anxiety-related documents.

Select the two depression topics and click **Merge topics**. A new topic is added as a User (rather than System) topic.

Click **Run Node**.

| Topics (21)                         |   |            |             |
|-------------------------------------|---|------------|-------------|
| <input type="checkbox"/>            | Topic   | Created by | Documents ↓ |
| <input type="checkbox"/>            | side, +side effect, +effect, no, +make          | System     | 90          |
| <input type="checkbox"/>            | +withdrawal, +symptom, horrible, off of, +drug  | System     | 79          |
| <input type="checkbox"/>            | abidal, +pain, +make, also, +medicine           | System     | 70          |
| <input type="checkbox"/>            | mg, +day, a day, very, +dose                    | System     | 70          |
| <input type="checkbox"/>            | +try, best, many, +other, meds                  | System     | 64          |
| <input type="checkbox"/>            | +life, +save, +change, +drug, back              | System     | 64          |
| <input type="checkbox"/>            | +medication, very, anxiety, +good, +well        | System     | 64          |
| <input type="checkbox"/>            | +day, -sleep, prexifan, only, +night            | System     | 62          |
| <input type="checkbox"/>            | weight, gain, weight gain, +cause, med          | System     | 61          |
| <input type="checkbox"/>            | +pound, +gain, +lose, +month, +drug             | System     | 60          |
| <input type="checkbox"/>            | +side, +effect, sexual, no, sexual side         | System     | 60          |
| <input checked="" type="checkbox"/> | +depression, still, anxiety, +depress, ecstapin | User       | 57          |

# Creating a custom topic from terms



Deselect all topics. In the term pane, click the **All** tab, and sort by **Term**. Select all terms that start with *depress*. There should be 10 terms selected.

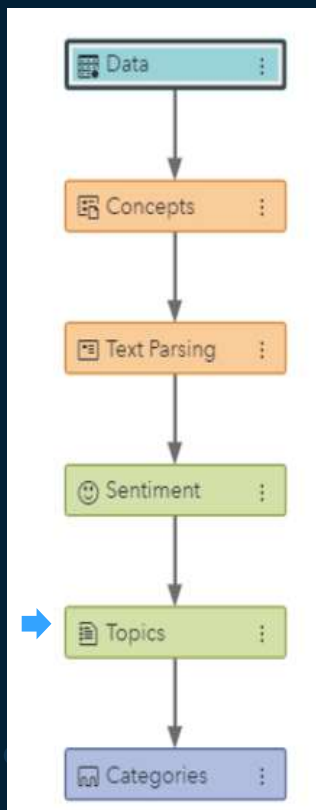
Click **+** (**Create a topic from terms**). This action creates a second User topic. There are now 20 System topics and two User topics.

Click **Run Node**.

The new topic captures 586 documents.

| Topics (22)                         |   |            |             |
|-------------------------------------|---|------------|-------------|
| <input type="checkbox"/>            | Topic   | Created by | Documents ↓ |
| <input checked="" type="checkbox"/> | +depress, +depressant, depressed, depression, +depression | User       | 586         |
| <input type="checkbox"/>            | side, +side effect, +effect, no, +make                    | System     | 90          |
| <input type="checkbox"/>            | +withdrawal, +symptom, horrible, off of, +drug            | System     | 79          |
| <input type="checkbox"/>            | abidal, +pain, +make, also, +medicine                     | System     | 70          |
| <input type="checkbox"/>            | mg, +day, a day, very, +dose                              | System     | 70          |
| <input type="checkbox"/>            | +try, best, many, +other, mede                            | System     | 64          |
| <input type="checkbox"/>            | +life, +save, +change, +drug, back                        | System     | 64          |
| <input type="checkbox"/>            | +medication, very, anxiety, +good, +well                  | System     | 64          |
| <input type="checkbox"/>            | +day, +sleep, prefixan, only, +night                      | System     | 62          |
| <input type="checkbox"/>            | weight, gain, weight gain, +cause, med                    | System     | 61          |
| <input type="checkbox"/>            | +pound, +gain, +lose, +month, +drug                       | System     | 60          |

# Creating user defined topics by merging user defined topics



Finally, a user-defined, combined depression-and-anxiety-based topic is created by merging the three depression-based and anxiety-based user-defined topics. Select the three User topics and click **Merge topics**. A fourth User topic is created.

Click **Run Node**.

The newest User topic identifies 647 documents.

| Topic  | Created by | Documents |
|--|------------|-----------|
| <input type="checkbox"/> +depression, -anxiety, -depress, -depressi                        | User       | 37        |
| <input type="checkbox"/> -depress, -depression, depressed, depression, -depression         | User       | 386       |
| <input type="checkbox"/> anxiety, anxious, anxiety   | User       | 204       |
| <input checked="" type="checkbox"/> +depression, anxiety, +depress, depressed, -depressant | User       | 647       |
| <input type="checkbox"/> +help, -help, -drug, -not, +happy                                 | System     | 49        |
| <input type="checkbox"/> weight, gain, weigh, gain, -cause, -med                           | System     | 51        |
| <input type="checkbox"/> side, -side effect, -effect, -to, -male                           | System     | 30        |
| <input type="checkbox"/> +withdraw, -symptom, -horrible, -off, -drug                       | System     | 79        |
| <input type="checkbox"/> +start, -pass, -week, -med, -review                               | System     | 36        |
| <input type="checkbox"/> -try, -best, -many, -other, -made                                 | System     | 34        |
| <input type="checkbox"/> -life, -take, -change, -drug, -back                               | System     | 64        |

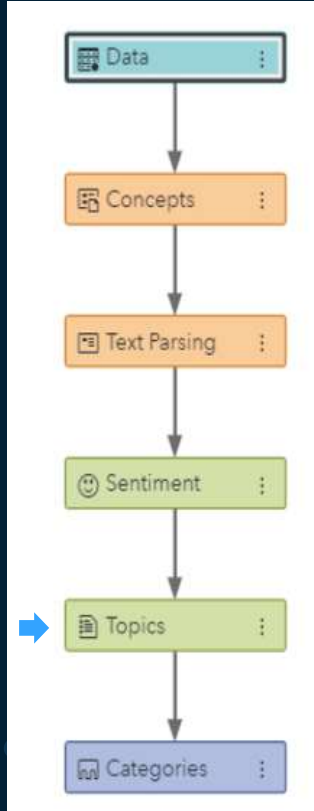
  

| Term  | Relevance | Role        | Documents | Frequency |
|---|-----------|-------------|-----------|-----------|
| <input type="checkbox"/> + depression         | 0.431     | N           | 492       | 478       |
| <input type="checkbox"/> anxiety              | 0.295     | N           | 187       | 231       |
| <input type="checkbox"/> + depress            | 0.081     | V           | 41        | 63        |
| <input type="checkbox"/> + depressed          | 0.233     | N           | 9         | 9         |
| <input type="checkbox"/> - depressive         | 0.233     | N           | 4         | 4         |
| <input type="checkbox"/> - depressive         | 0.233     | PN          | 4         | 4         |
| <input type="checkbox"/> + depressive symptom | 0.233     | N NounGroup | 4         | 4         |
| <input type="checkbox"/> - depressive         | 0.233     | A           | 17        | 18        |
| <input type="checkbox"/> - depressive         | 0.233     | PN          | 9         | 10        |
| <input type="checkbox"/> + depressive episode | 0.233     | N NounGroup | 4         | 7         |
| <input type="checkbox"/> - anxious            | 0.233     | A           | 16        | 16        |

| Document  | Sentiment | Relevance |
|---|-----------|-----------|
| depression                                      | 😊         | 0.431     |
| <depression                                     | 😊         | 0.431     |
| ...helped my fight depression and anxiety       | 😊         | 0.381     |
| ...has helped my depressive symptoms both times | 😊         | 0.335     |
| ...cleaning up my depression                    | 😊         | 0.332     |

# Creating another custom topic



Add another User topic by selecting all terms related to anxiety. The terms **anxiety**, **anxious**, and **anxity** are selected.

Terms

All (1511) Matched Filter

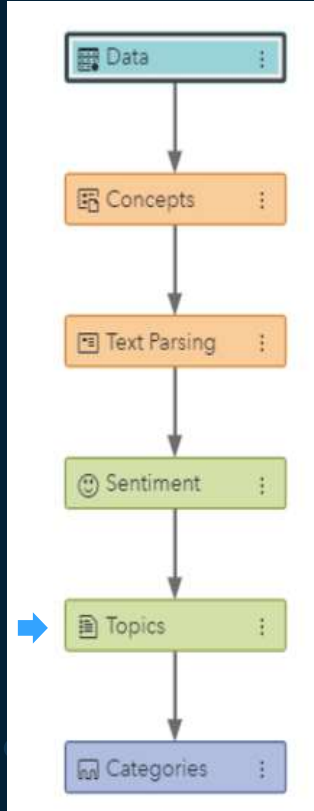
|                                     | Term              | Role | Documents | Frequency |
|-------------------------------------|-------------------|------|-----------|-----------|
| <input type="checkbox"/>            | ▶ anti-depressant | N    | 51        | 57        |
| <input type="checkbox"/>            | ▶ antidepressant  | N    | 52        | 59        |
| <input checked="" type="checkbox"/> | anxiety           | N    | 187       | 231       |
| <input checked="" type="checkbox"/> | anxious           | A    | 16        | 16        |
| <input checked="" type="checkbox"/> | anxity            | N    | 4         | 5         |
| <input type="checkbox"/>            | anymore           | ADV  | 45        | 47        |
| <input type="checkbox"/>            | anything          | N    | 34        | 37        |
| <input type="checkbox"/>            | anyway            | ADV  | 7         | 8         |
| <input type="checkbox"/>            | anywhere          | ADV  | 5         | 6         |
| <input type="checkbox"/>            | appetite          | N    | 20        | 20        |

Table  
Description automatically generated

Click the icon for **Create a topic from terms**.

Click **Run Node**.

# Creating another custom topic



The new anxiety topic identifies 204 documents.

The screenshot shows the SAS interface. On the left, a list of topics is displayed with columns for 'Topic', 'Created by', and 'Documents'. The 'anxiety, anxious, anxiety' topic is selected, showing it was created by 'User' and has 204 documents. Below this, a 'Documents' section shows a search for 'anxiety' with 204 matches. On the right, a 'Terms' table is visible, showing terms like 'anxious', 'anxiety', and 'anxiety' with their respective 'Relativity', 'Role', 'Documents', and 'Frequency'.

| Topic  | Created by | Documents |
|--|------------|-----------|
| -depress, -depressed, depressed, depression, +depression | User       | 154       |
| anxiety, anxious, anxiety                                | User       | 204       |
| side, side effect, -effect, no, -nause                   | System     | 65        |
| Anticholin: -symptom, horrible, off of, +drug            | System     | 79        |
| abridal, -pain, -nause, also, -medicine                  | System     | 70        |
| mg, +day, a day, very, -dose                             | System     | 70        |
| 4ty, best, many, -other, medi                            | System     | 64        |
| -life, -save, -change, +drug, back                       | System     | 64        |
| +medication, very, anxiety, -good, -well                 | System     | 64        |
| 4day, -sleep, prefxin, only, -right                      | System     | 62        |
| weight, gain, weight, gain, -cause, mad                  | System     | 61        |

| Terms   | Relativity | Role | Documents | Frequency |
|---------|------------|------|-----------|-----------|
| anxious | 1.000      | A    | 1c        | 1d        |
| anxiety | 1.000      | N    | 4         | 1         |
| anxiety | 1.000      | N    | 107       | 231       |

Finally, a user-defined, combined depression-and-anxiety-based topic is created by merging the three depression-based and anxiety-based user-defined topics. Select the three User topics and click **Merge topics**. A fourth User topic is created.

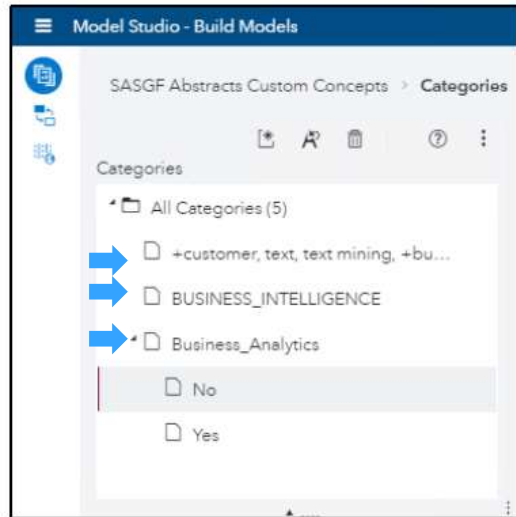
Click **Run Node**.

# A sample Topic scored data file

The output table **ASRS\_out\_documents** is created and stored in the PUBLIC library.

| PUBLIC ASRS_OUT_DOCUMENTS |                  |                  |                  |                  |                  |                   |                  |                  |                  |                  |
|---------------------------|------------------|------------------|------------------|------------------|------------------|-------------------|------------------|------------------|------------------|------------------|
| Columns: 32 of 32         |                  |                  |                  |                  |                  |                   |                  |                  |                  |                  |
| Total rows: 21519         |                  |                  |                  |                  |                  |                   |                  |                  |                  |                  |
| Rows 1 to 200             |                  |                  |                  |                  |                  |                   |                  |                  |                  |                  |
| Enter expression          |                  |                  |                  |                  |                  |                   |                  |                  |                  |                  |
|                           | @_Col1_          | @_Col2_          | @_Col3_          | @_Col4_          | @_Col5_          | @_Col6_           | @_Col7_          | @_Col8_          | @_Col9_          | @_Col10_         |
| 1                         | 0.19847354<br>63 | 0.00643179<br>44 | 0.08643419<br>02 | 0.19534131<br>19 | 0.01030578<br>2  | 0.00679471<br>97  | 0.09100080<br>65 | 0.07681528<br>22 | -0.0113818<br>8  | -0.0394757<br>02 |
| 2                         | 0.01061343<br>45 | 0.01193305<br>31 | 0.01314013<br>28 | 0.00962410<br>33 | 0.07654393<br>57 | 0.01444503<br>21  | 0.06485029<br>94 | 0.01410577<br>68 | 0.00628118<br>96 | 0.01111935<br>87 |
| 3                         | 0.05846577<br>84 | 0.00494122<br>73 | 0.04873458<br>16 | 0.04605954<br>64 | -0.0144930<br>23 | 0.04485330<br>82  | 0.06051552<br>53 | 0.00377150<br>36 | 0.08580722<br>04 | 0.05197937<br>26 |
| 4                         | -0.0111955<br>26 | 0.01486958<br>18 | 0.10565853<br>77 | 0.05208292<br>53 | 0.05097773<br>61 | -0.00826112<br>82 | 0.00465248<br>86 | -0.0156777<br>25 | 0.00489825<br>24 | 0.11555344<br>99 |
| 5                         | 0.04617909<br>26 | -0.0362952<br>86 | 0.02473544<br>15 | 0.00207705<br>78 | -0.0326571<br>68 | -0.0227175<br>62  | 0.54798575<br>16 | 0.25292577<br>28 | 0.04759192<br>15 | -0.0414928<br>31 |
| 6                         | 0.02979279<br>32 | 0.02245643<br>63 | 0.05213735<br>1  | 0.02768965<br>32 | 0.12076104<br>34 | 0.21113382<br>08  | 0.01544792<br>62 | 0.02208903<br>29 | 0.01810194<br>79 | -0.0100048<br>02 |
| 7                         | 0.02135490<br>8  | 0.07013431<br>4  | 0.00213814<br>42 | 0.06251029<br>28 | -0.0301033<br>07 | 0.00207380<br>85  | 0.03137579<br>06 | -0.0249775<br>06 | -0.0039019<br>03 | 0.01432084<br>82 |
| 8                         | -0.0317092<br>51 | 0.02580648<br>25 | 0.02948603<br>27 | 0.16134114<br>11 | 0.00498050<br>61 | 0.02108605<br>43  | 0.03294550<br>17 | -0.0264215<br>38 | 0.00283823<br>92 | 0.01281520<br>99 |
| 9                         | 0.04669131<br>47 | 0.01753994<br>86 | 0.02462550<br>27 | 0.03384969<br>97 | -0.0041408<br>24 | -0.00132597<br>49 | 0.18782173<br>35 | 0.14231479<br>71 | 0.01314466<br>53 | 0.02318901<br>54 |
| 10                        | 0.00145617<br>33 | 0.15270084<br>39 | 0.01704378<br>97 | 0.04772194<br>2  | 0.00941408<br>55 | 0.02288804<br>83  | 0.04561390<br>18 | 0.00340108<br>39 | -0.0048414<br>81 | 0.04136030<br>5  |
| 11                        | 0.07536212<br>51 | 0.00256392<br>7  | 0.08370938<br>69 | 0.23151136<br>25 | -0.0480711<br>39 | -0.0290987<br>75  | 0.30556481<br>59 | 0.03134610<br>72 | 0.03438437<br>91 | -0.0329465<br>74 |
| 12                        | 0.05113851<br>08 | 0.05279136<br>14 | 0.02084400<br>18 | 0.09924056<br>69 | 0.03714441<br>85 | 0.02160616<br>21  | 0.05749636<br>91 | 0.06281262<br>14 | 0.02495957<br>03 | 0.00738071<br>77 |

# Categorization Rule Generation



Boolean linguistic rules automatically derived

easy to understand

can be edited



# Category Rules

## Boolean Rules

CATEGORY

- Boolean and proximity operators
- arguments
- modifiers

- ✓ more precise than linguistic rules
- ✓ term located for structured input documents (for example, XML)
- ✓ qualifiers modify the effect of the rule



# Category rules

## Boolean Rule Syntax

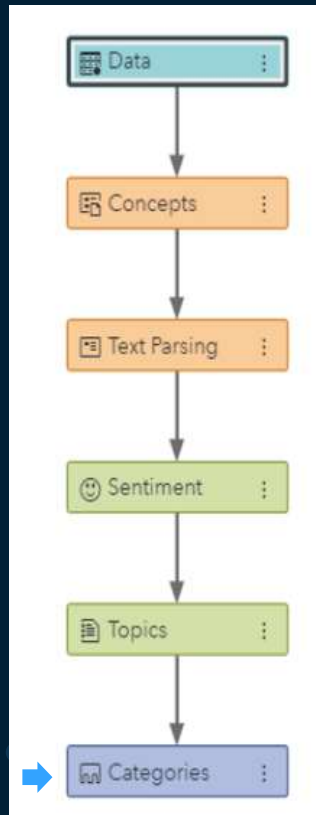
- Boolean operator names are specified in uppercase.
- Boolean operators and their arguments are enclosed in parentheses.
- Arguments are separated by commas.
- Terms are enclosed in double quotation marks.

Example, when displayed in text view:

```
(AND, "Amazon", (OR, "online retailer", "e-commerce"))
```

This rule is true when the term *Amazon* appears with either of the terms *online retailer* or *e-commerce*.

# Categories are generated from topics



Select the three User topics at the top based on the sort order of the number of documents.

| Topics (24)   |            |           |
|---|------------|-----------|
| <input type="checkbox"/> Topic  | Created by | Documents |
| <input type="checkbox"/> +depression, still, anxiety, +depress, ecstapin                      | User       | 57        |
| <input checked="" type="checkbox"/> +depress, +depressant, depressed, depression, +depression | User       | 586       |
| <input checked="" type="checkbox"/> anxiety, anxious, anxiety                                 | User       | 204       |
| <input checked="" type="checkbox"/> +depression, anxiety, +depress, depressed, +depressant    | User       | 647       |
| <input type="checkbox"/> +help, really, +drug, not, +happy                                    | System     | 49        |
| <input type="checkbox"/> weight, gain, weight gain, +cause, med                               | System     | 61        |
| <input type="checkbox"/> side, +side effect, +effect, no, +make                               | System     | 90        |
| <input type="checkbox"/> +withdrawal, +symptom, horrible, off of, +drug                       | System     | 79        |

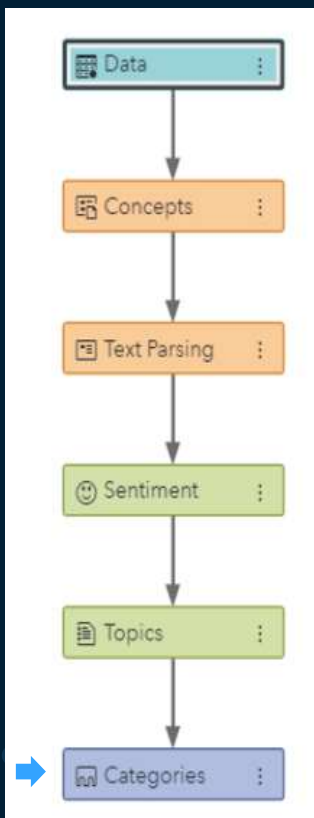
Select **Add topics as categories**.

| Topics (24)   |            |           |
|---|------------|-----------|
| <input type="checkbox"/> Topic  | Created by | Documents |
| <input type="checkbox"/> +depression, still, anxiety, +depress, ecstapin                      | User       | 57        |
| <input checked="" type="checkbox"/> +depress, +depressant, depressed, depression, +depression | User       | 586       |
| <input checked="" type="checkbox"/> anxiety, anxious, anxiety                                 | User       | 204       |
| <input checked="" type="checkbox"/> +depression, anxiety, +depress, depressed, +depressant    | User       | 647       |
| <input type="checkbox"/> +help, really, +drug, not, +happy                                    | System     | 49        |
| <input type="checkbox"/> weight, gain, weight gain, +cause, med                               | System     | 61        |
| <input type="checkbox"/> side, +side effect, +effect, no, +make                               | System     | 90        |
| <input type="checkbox"/> +withdrawal, +symptom, horrible, off of, +drug                       | System     | 79        |

Close the **Topics** node.

Run the pipeline.

# Category rules



Examine each promoted topic category rules.

**Categories**

- All Categories (3)
  - depress, +depress, depressed, depression, +de...
  - depression, anxiety, -depress, depressed, -depre...
  - anxiety, anxious, anxiety

**Text Sample Text**

41 (1414) **Matched (157 of 1414)** Search

**Drug Report**

...fit with some **depression** few months ago. Well, that's no fun. I will never say to someone they stand out of 4 Life worlds in funny ways sometimes. Anyway, was forced to get help with the **depression**. Started with low-dose 37.5 for 10 days, then 75mg for 14 days and to date at 75 + 27.5 = 112.5mg. The hardest part was waiting for it to work. It took almost 6/8 weeks. I wanted to quit many times, but stuck with it. Weird side effects along the way but he...

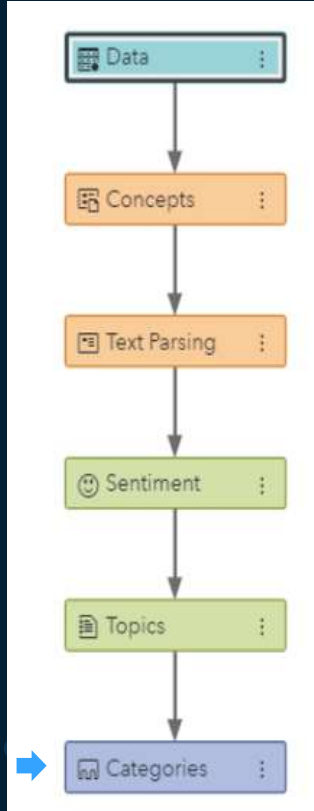
...to treat severe **depression** and anxiety with heart palpitations. My doc also put me on steroids to control the palpitations, which is a nice combo because it reduces your blood pressure and prevents migraines. After you I felt better. Like many of you who complained about wanting to get off it, and slowly tapered off. What I have learned in the last year of horrible **depressive** trying to do it on my own with diet, exercise, sunlight, and the steroids is that...

...great for my **depression**, but the thing that worked best for my was keeping busy and getting involved in a project out of the home. Productivity increases your confidence immensely. Also good doctors that know the newest drugs on the market that are found to all over **depression**. My doctor discovered that NERFLEX, a prescription medicine, helped sustain a better mood in patients. A combination of the right medicines will help alleviate **depression**, but you...

...for post-partum **depression** that escalated to **depression** I tried 3 other types before escitalopram 150. This one is the best so far I am no longer **depressed** but I have not been able to lose weight like I want to. I have not had any side effects and I have been taking about a year now. I do get tired more now but not like later when I was **depressed**.

| Sentiment | Relevancy |
|-----------|-----------|
| ☹️        | 12,000    |
| ☹️        | 7,000     |
| 😊         | 3,000     |
| 😊         | 3,000     |

# Scoring Category model



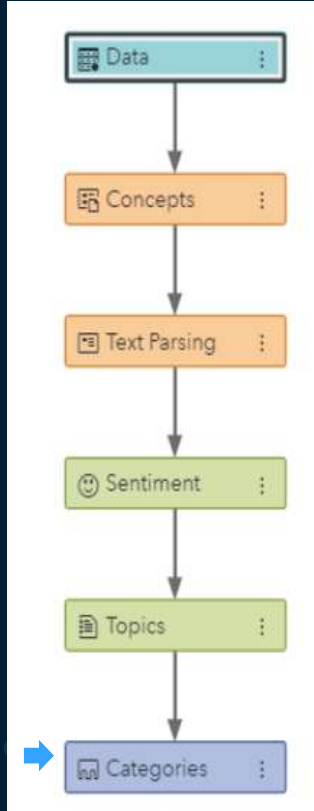
Upload the `ScoreCode_Categories_NYC_Low.sas` file to the server and open the file in the SAS Studio editor. Make the following changes to the score code:

```
%let input_caslib_name = "PUBLIC";
%let input_table_name = "NYCDOHMH_A";
%let key_column = "RECORD_ID";
%let document_column = "VIOLATION_DESCRIPTION";
%let output_caslib_name = "PUBLIC";
%let output_categories_table_name = "nyc_out_categories";
%let output_matches_table_name = "nyc_out_matches";
%let output_modeling_ready_table_name = "nyc_out_modeling_ready";
```

Run the program. Three output tables are written to the PUBLIC library.

| Code                                     | Log                    | Results |                |                   |
|--|------------------------|---------|----------------|-------------------|
| The CAS Procedure                        |                        |         |                |                   |
| textRuleScore.applyCategory              |                        |         |                |                   |
| Output CAS Tables                        |                        |         |                |                   |
| Results from textRuleScore.applyCategory |                        |         |                |                   |
| Output CAS Tables                        |                        |         |                |                   |
| CAS Library                              | Name                   | Label   | Number of Rows | Number of Columns |
| Public                                   | nyc_out_categories     |         | 56298          | 4                 |
| Public                                   | nyc_out_matches        |         | 206241         | 5                 |
| Public                                   | nyc_out_modeling_ready |         | 20530          | 12                |

# Scored Output data from category node



| <input type="checkbox"/> | Variable Name         | ↑ | Label                                | Type      | Role ←   | Level    |
|--------------------------|-----------------------|---|--------------------------------------|-----------|----------|----------|
| <input type="checkbox"/> | category_1            |   | _VerminCategory                      | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_10           |   | _FoodPrepCategory                    | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_11           |   | _FacilitiesCategory                  | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_2            |   | _DocumentationCategory               | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_3            |   | SCOREGRADE/C                         | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_4            |   | _WorkersCategory                     | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_5            |   | vermin, +facility, present, live, 's | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_6            |   | SCOREGRADE/A                         | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_7            |   | SCOREGRADE/B                         | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_8            |   | LowGrade/0                           | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | category_9            |   | LowGrade/1                           | Numeric   | Input    | Binary   |
| <input type="checkbox"/> | LowGrade              |   | Low Grade Indicator                  | Numeric   | Target ← | Binary   |
| <input type="checkbox"/> | RECORD_ID             |   | Document ID                          | Character | ID       | Nominal  |
| <input type="checkbox"/> | SCORE                 |   |                                      | Numeric   | Rejected | Interval |
| <input type="checkbox"/> | SCOREGRADE            |   | Score Grade                          | Character | Rejected | Nominal  |
| <input type="checkbox"/> | VIOLATION_DESCRIPTION |   |                                      | Character | Rejected | Nominal  |

# Summary

---

*Introduced Visual Text Analytics in SAS Viya*

*Text analytics steps*

*Scoring new text data for text features*



# Thank you

## Questions / Comments

**Who should attend**  
Text analysts, business and marketing analysts, web analysts, BI professionals, customer intelligence professionals, social media analysts, and statement translators.

**Formats available**  
 Classroom  
 Live Web Classroom  
 e-Learning

**Standard Duration**  
 3 days  
 4 half-day sessions  
 14 hours/100-day format

**System Requirements**  
[System Requirements](#)

[View course details](#)

[View learning path](#) [Suggest course date](#) [See program details](#)

**Self-Paced e-Learning**

| Title   | Duration     | Access Period            | Language | Price    | Add to Cart |
|---|--------------|--------------------------|----------|----------|-------------|
| SAS Visual Text Analytics in SAS Viya (2020) (1-2) virtual self-paced | 14 (1) hours | 100 days from order date | English  | \$70 USD |             |
| SAS Visual Text Analytics in SAS Viya (2020) (3-5) virtual self-paced | 14 (1) hours | 100 days from order date | English  | \$70 USD |             |

**Classroom and Live Web Schedule** View: United States | Nearby | Americas

| Date           | Location | Time                  | Language | Price     | Add to Cart |
|----------------|----------|-----------------------|----------|-----------|-------------|
| 24-27 JUL 2022 | Live Web | 07:00 PM-04:30 PM EDT | English  | 1,490 USD |             |
| 27-28 SEP 2022 | Live Web | 01:50 PM-04:30 PM EDT | English  | 1,490 USD |             |
| 30-02 DEC 2022 | Live Web | 01:00 PM-04:30 PM EDT | English  | 1,490 USD |             |

# Q&A

Please submit your questions using the Q&A icon located in the menu at the bottom of your screen





# Explore Helpful Resources

## [Ask the Expert](#)

View other user webinars that provide insights into using SAS products to make your job easier.

## [FREE Training](#)

Learn from home – free for 30 days. Get software labs to practice and online support if needed.

## [SAS Support Communities](#)

Ask questions, get answers and share insights with SAS users.

## [SAS Analytics Explorers](#)

An exclusive platform to collaborate, learn and share your expertise. Gain access to a diverse network to advance your career. Special rewards and recognition exclusively for SAS users.

## [SAS Users YouTube Channel](#)

A plethora of videos on hundreds of topics, just for SAS users.

## [Newsletters](#)

Get the latest SAS news plus tips, tricks and more.

## [Users Groups](#)

Meet local SAS users, network and exchange ideas – virtually.

## [SAS Profile](#)

If you haven't already done so, create your SAS Profile to access free training, SAS Support Communities, technical support, software downloads, newsletters and more.



Thank you  
for joining us for  
this SAS webinar