Topic Modeling Tool

The Topic Modeling tool identifies and categorizes topics in a body of text.

This example uses tools that require the Alteryx Intelligence Suite be installed and licensed before it can be run.

1) Run the workflow (Ctrl+R).	2) Select a tool to view its output in the Results window.

Enable the Tool Container to run this example

Without Text Pre-processing

To configure the Topic Modeling tool, connect your text data to the input anchor. Use the drop down to select the column that contains the text data to be analyzed, then set the number of topics to classify. Finally, choose whether to return the Interactive Chart or the Word Relevance Summary.

The Interactive Chart option returns an interactive visualization of the model that you can view with a Browse tool. The interactive chart has 2 parts, a map with the distance between the topics, and some metrics for evaluation. The Intertopic Distance map shows us how similar the identified topics are.

The Word Relevance Summary returns the words included in the topic model as well as Relevance and Saliency metrics. Saliency is how prominent the word is in the overall text. Relevance is a metric used to order words within topics and helps us to identify the most appropriate words for each topic. The higher the value for a given topic, the more important that word is for that topic.

There are 3 output anchors for the Topic Modeling tool. The D anchor returns your original data with the additional columns for the topics added. The R anchor returns the report selected in the configuration menu. A Browse tool is required to view this report. The M anchor returns the model object that can be used down stream with a Predict tool.

Additional configuration options are available to customize the model. Additional configurations include the Dictionary Options Min and Max Frequency, and Max words. Alpha and Eta settings are also available in the LDA Options section. Select the ? icon next to any of the options to learn more about what that option does and recommended values.

With Text Pre-processing

A best practice is to use the Topic Modeling tool with the Text Pre-processing tool. The Text Pre-processing tool helps to clean up the text data by converting words to their roots and removing unwanted punctuation, digits, and stop words.

With the exception of the Text Pre-processing tool, this example uses the same set up as the one above. The only difference is the report selected to show the different options.

More Info

You can use the Part-of-Speech Tagger tool to remove nouns to further enhance the Topic Modeling results.

After training the model, use the Predict tool in the Machine Learning tool category to predict topic relevance in a new body of text.