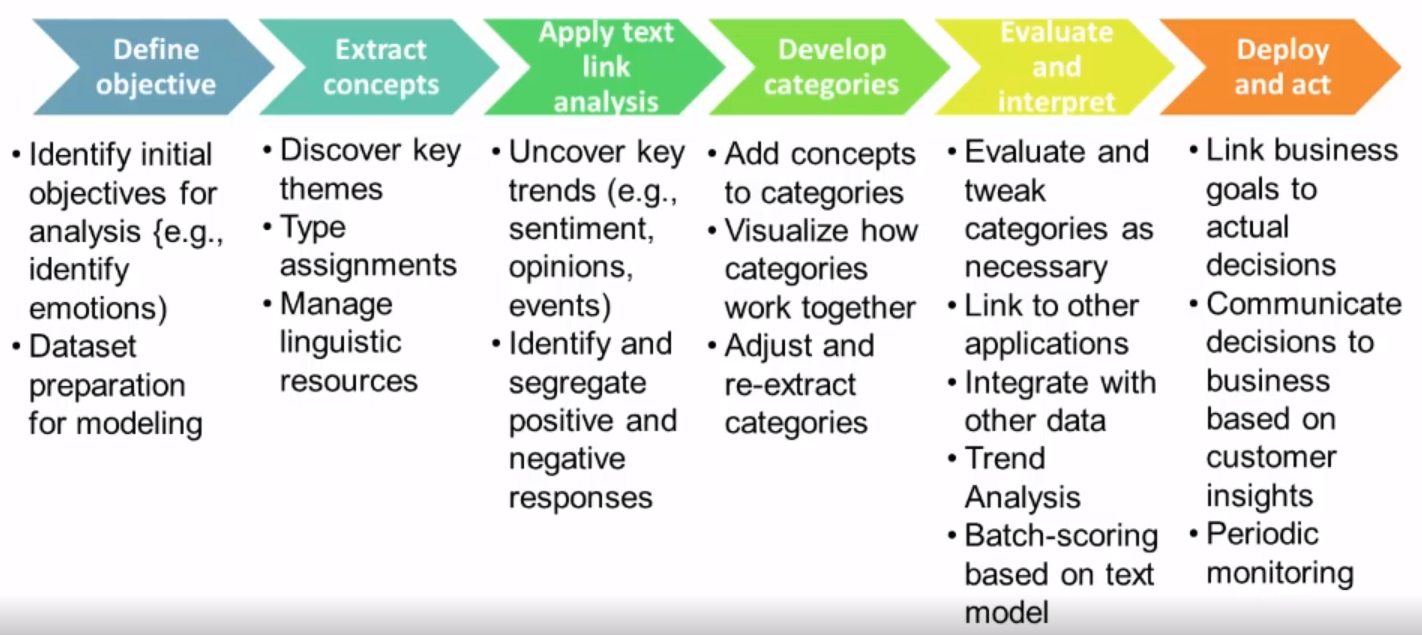
Text Mining

**Text Analytics**

Natural Language Processing (NLP) getting data from text

CRISP – DM Methodology applies to text mining

Text analytics Framework



**Data mining vs Text Mining**

In traditional data mining application you can either train the model on target variable (supervisd modeling), or let the model find natural patterns in the data (ie unsupervised clustering)

In text mining, the work remains very time consuming and manual:  
 SW helps extract most frequent concepts  
 apply one of the available dictionaries to map to (limited) pre-defined categories  
 Extract sentiment  
 Extract behavioral sentiment

**Data Mining (Predictive Analytics) & Text Mining**

While both Data Mining and Text Mining aim at extracting patterns in data, Data Mining uses only structured data as input while Text Mining can also work with information stored in an unstructured collection of documents.

Before Data Mining tools can used to find patterns in free text data the information contained therin must first be converted into structured data called concepts, types and categories.

Once categories have been identified, they can be combined with other data source and analyzed using Data Mining techniques using analytical software packages

**Text Mining – “Structuring” unstructured information**

Selecitng appropriate dictionary  
 May need customized dictionaries

What kinds of concepts and categories should we expect to find?  
 Automatic concept discovery will help understand data  
 This information must be augmented based on the business objective and insights from domain experts

Categories are usually the final outcome of the text mining process, they are new fields which will then be used by the data mining process to enrich predictive models  
 The categories allow us to apply numerical values to the text data (usually frequencies)  
 We can combine categories with sentiment to get more granular frequencies and patterns

**Social Media Analytics is a subset of Natural Language Processing**

While both data mining and text mining aim at extracting patterns in data, data mining uses only structured data as input, while Text Mining can also work with information stored in an unstructured collection of documents

**SMA measures a combination of volume and sentiment of a customer’s discussions**

**Buzz:** Measure of buzz: How much customers are talking about a particular product or brand  
Share of buzz: How much of the discussion surrounds the product vs the competitor

**Sentiment:** Classification of sentiment: What is the customer’s reaction? Is the sentiment positive, negative or neutral?  
How sentiment changes with new product announcements or marketing messages?

**Personality profiling:**  enables targeted response to capitalize on the positive messages and counteract the negative ones

Profiling enables analysts to segment the comments coming from Users to Recommenders and Perspective Users

The insights can lead to critical information to enable the enterprise to capitalize on the positive messages and counteract the negative ones.

Data Aggregators – aggregates information from multiple sources.

Most Effective social media mining goes beyond text mining and sentiment analysis

The secret to SMA is not limited to just effective text processing or the ability to structure customer feedback or even to measure sentiment

**Augment with other sources of information**: In order to be effective SMA information must be overlaid with the variety of traditional legacy sources of information, including customer sales data, call centers and traditional marketing research

**Most effective when used over time**: SMA is most effective when used over time to gain insights into how the buzz is changing depending on what is happening in the marketplace and with the competitive environment

The results of Social Media Analytics can be effectively delivered on mobile devices, for example using Social Analytics iPad App