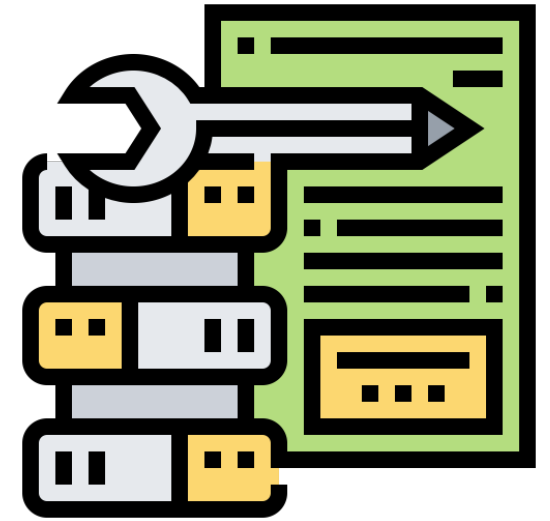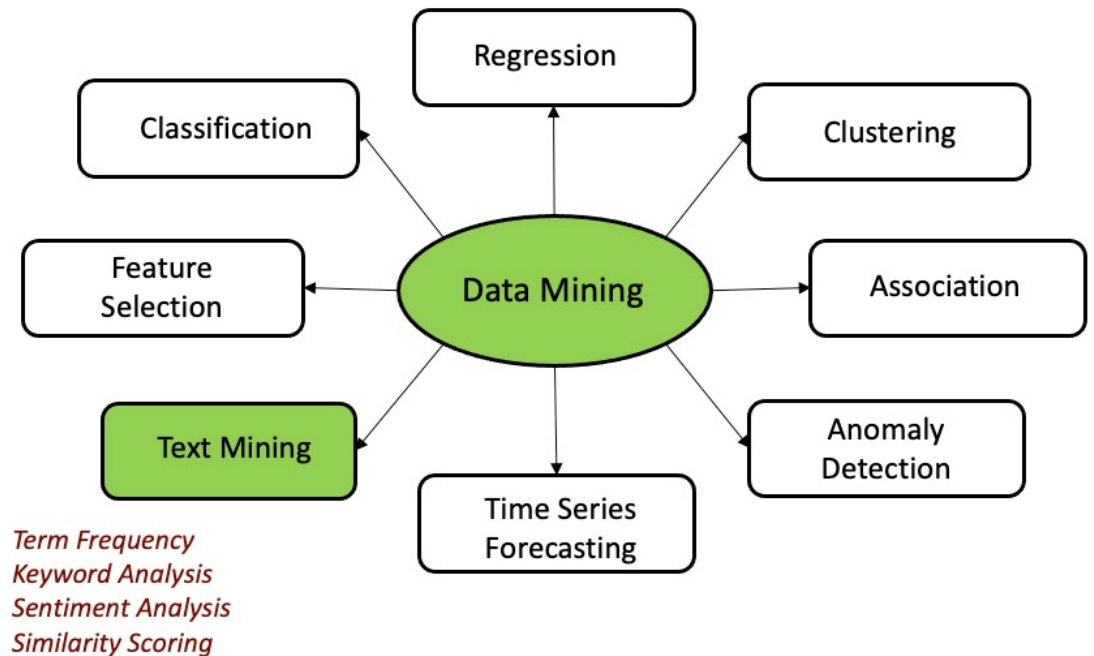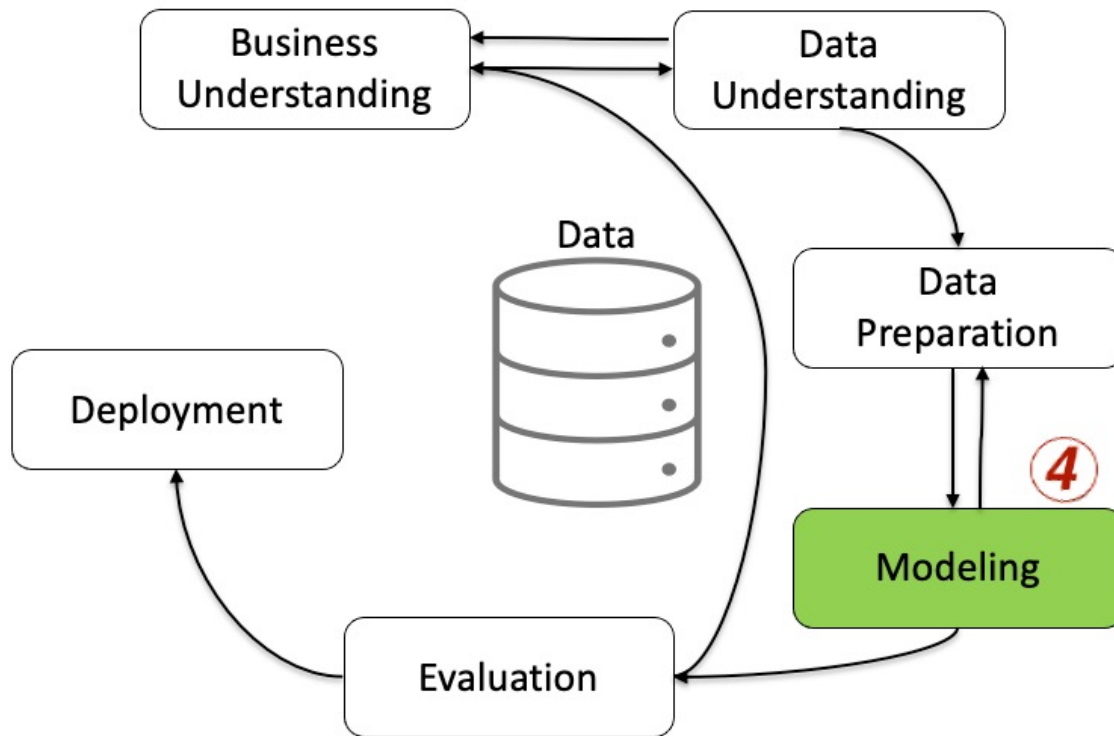# Text Data Mining For Business Decisions
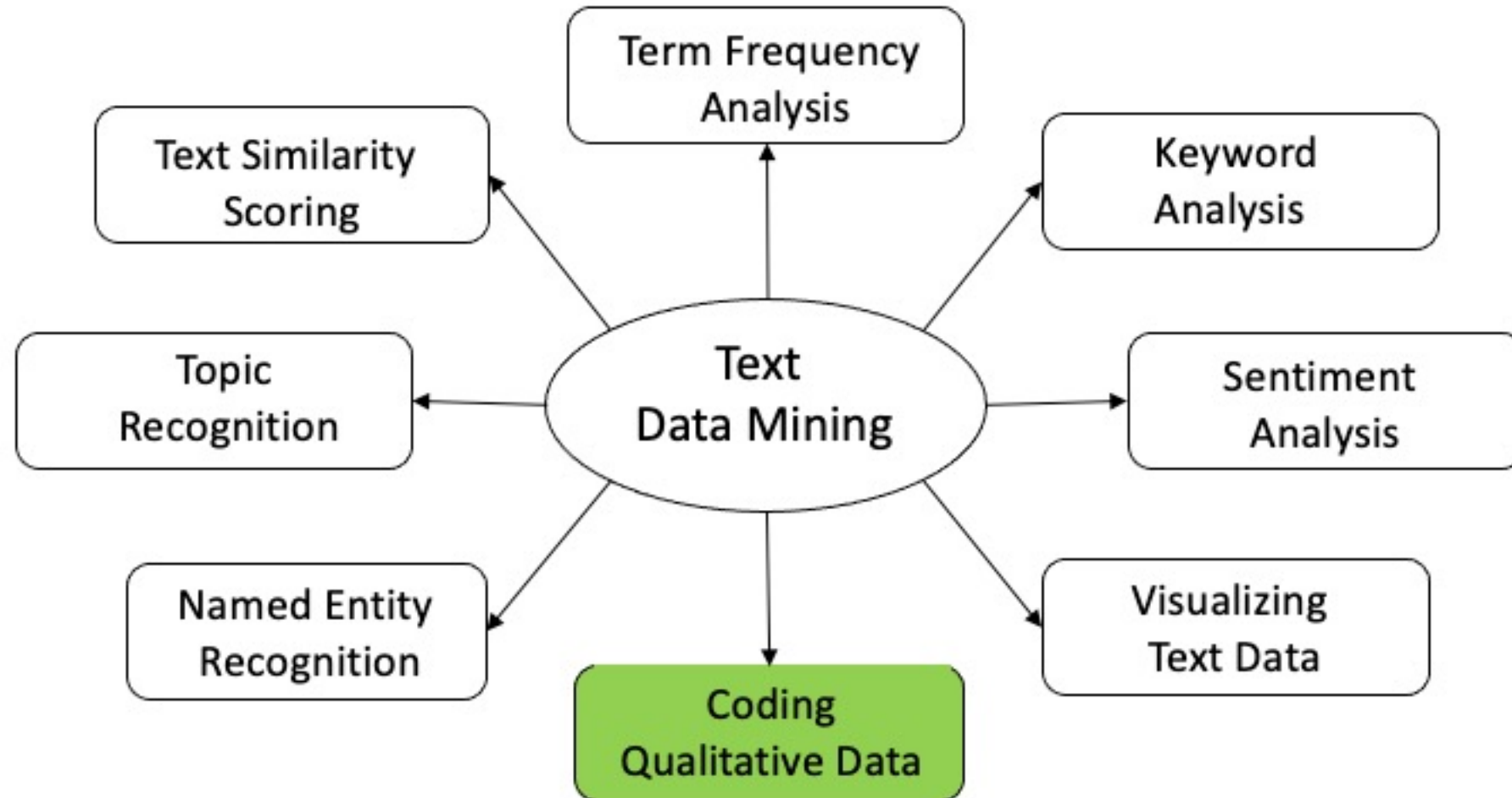
Module 7
Coding Qualitative Data
How do we analyze qualitative data using coding techniques?

Data Mining- Continuing Model-Making

Coding Qualitative Data Analysis is the Original Process for Text Data Mining

What is Coding Qualitative Data analysis?

- In this session we analyze text data using a traditional approach.

- Text data is called qualitative data, as opposed to the quantitative data that we collect as numerical or categorical data.

- Researchers have developed a sophisticated technique to analyze qualitative data, which is referred to as coding.

- It is a way of translating text that is difficult to enumerate and characterizing it using an *analyst- based scheme*, the coding, into something that can be tabulated by quantizing the text data.

What are the two kinds of Coding?

- There are two kinds of coding.

- One is *inductive*, where the analyst extracts basic categories using a close reading of the text.

  - For example, consider reading many social media postings, like tweets, that have not been categorized by adding hashtags. Inductive coding is essentially the process of adding those hashtags, which in your opinion, categorize each tweet.

- The other form of coding is *deductive* coding.

  – In that case, we start with a preconceived notion of what the codes are and use them to classify each text.

Practice in this class

- We will use survey responses to inductively create codes, categorize each survey response, and tabulate the responses.

- We do the same thing for customer feedback of products.

- For the deductive coding practice, we employ a well-known code system for categorizing books, the Dewey decimal system, and categorize books according to that scheme

- *Analysis coding is the critical link between data collection and an explanation of meaning.*

FORTINO
Global Education

What is a Code?

- In qualitative inquiry, a code is most often a *word* or *short phrase* that symbolically assigns a summative, salient, essence-capturing, or evocative attribute for a portion of language-based data.

- We do this all the time - it's like adding a hashtag to a tweet

- What codes would you use to categorize for the following survey response?

  "I like taking my classes online. I feel the online class will be the future trend, we can have our class in any places of the world, and we can review the class materials any time online."

FORTINO
Global Education

A Manual Tagging/Machine Computation Process

- A human reads the texts, identify opinions, and detect nuances,
- It is costly and time consuming
  - Manually tagging opinions can be arduous, given that the amount of data businesses receive is constantly growing.
  - Researchers who collected qualitative data (interviews, open ended questions) routinely use this method
- Not good for large amounts of text - a few survey responses at the most
- But is very accurate in that it yields results focused on the information need (research question)

How does Inductive Coding Work?

1. Break your qualitative dataset into smaller samples (randomly select a few of the survey responses). This works well if you have thousands of survey responses.
2. Read the samples of the data.
3. Create codes for the sample data.
4. Reread the sample and apply the codes.
5. Read a new sample of data, applying the codes you created for the first sample
6. Note where codes don't match or where you need additional codes.
7. Create new codes based on the second sample.
8. Go back and recode all responses again. This is the step where you can use the codes as keywords and do a preliminary classification of the responses by keyword analysis. A human- based classification can interpolate close matches by affinity coding.
9. Repeat the process from step 5 until you have coded all of your text data.
10. Tabulate by Code

What is Deductive Coding?

- Deductive coding means you start with a predefined set of codes, then assign those codes to new qualitative data.

- These codes might come from previous research, or you might already know what themes you're interested in analyzing.

- Deductive coding is also called *concept-driven* coding.

  – For example, let's say you're conducting a survey on customer experience. You want to understand the problems that arise from long call wait times, so you choose to make "wait time" one of your codes before you start examining the data.

FORTINO
Global Education