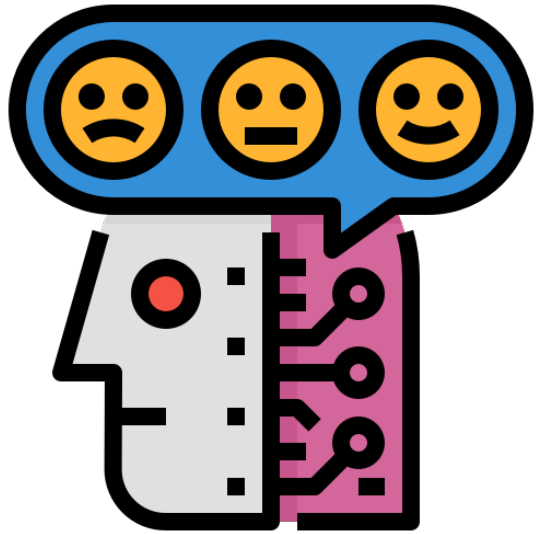
Text Data Mining
For Business Decisions

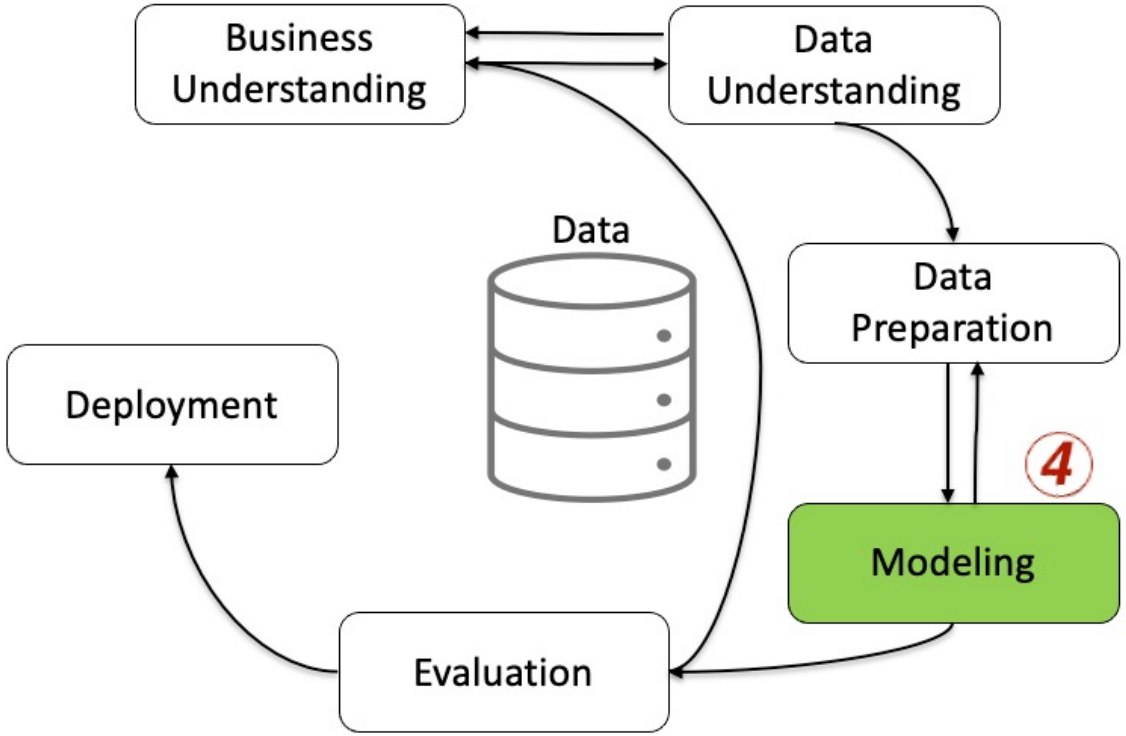
Module 6

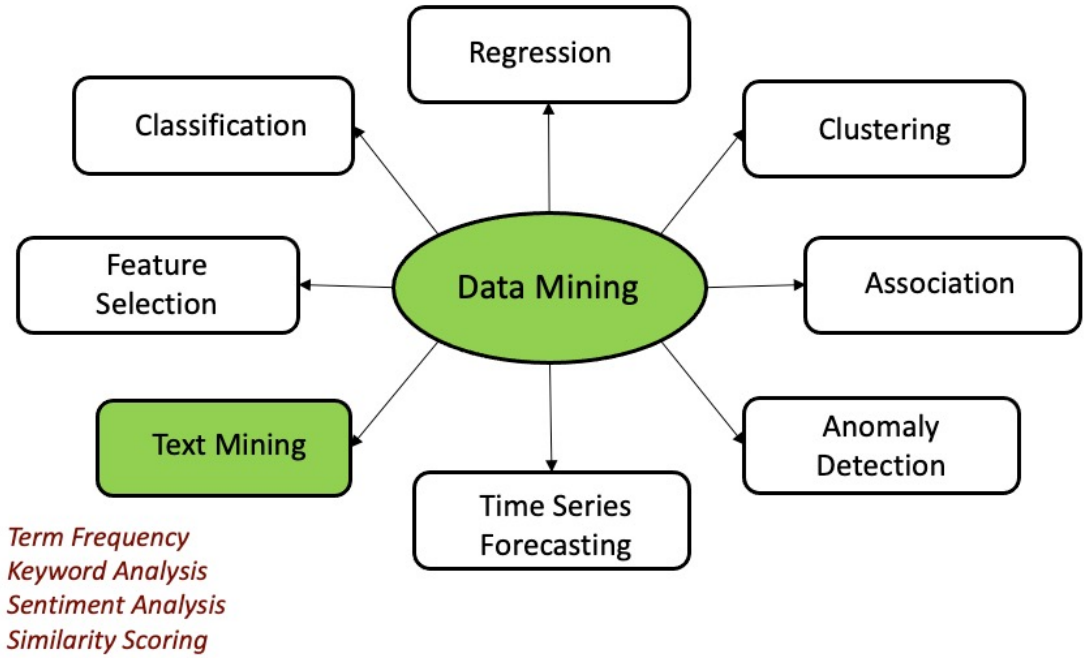
Sentiment Analysis

How do we discover what people are telling us in their texts?

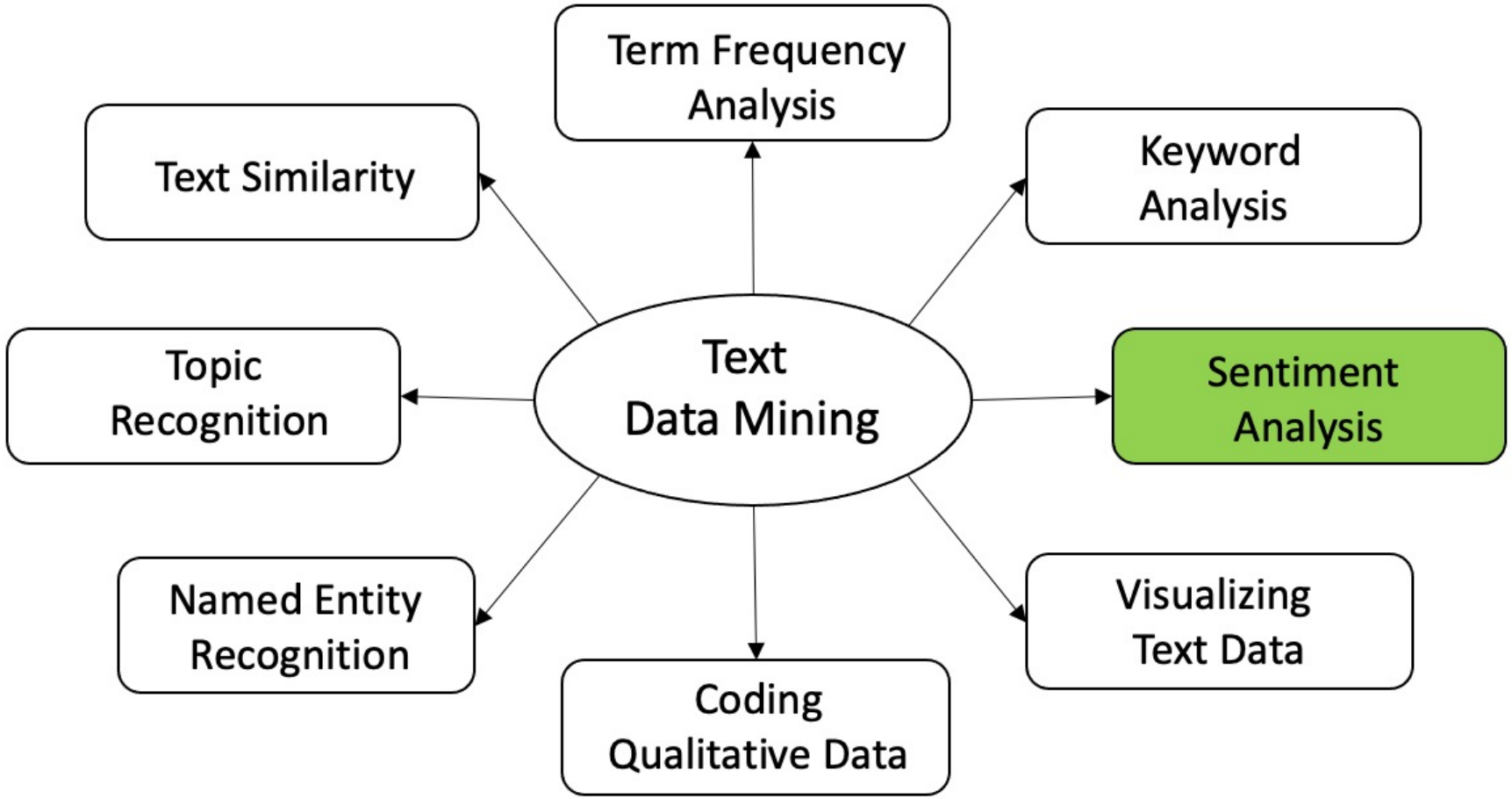


Data Mining- Continuing Model-Making

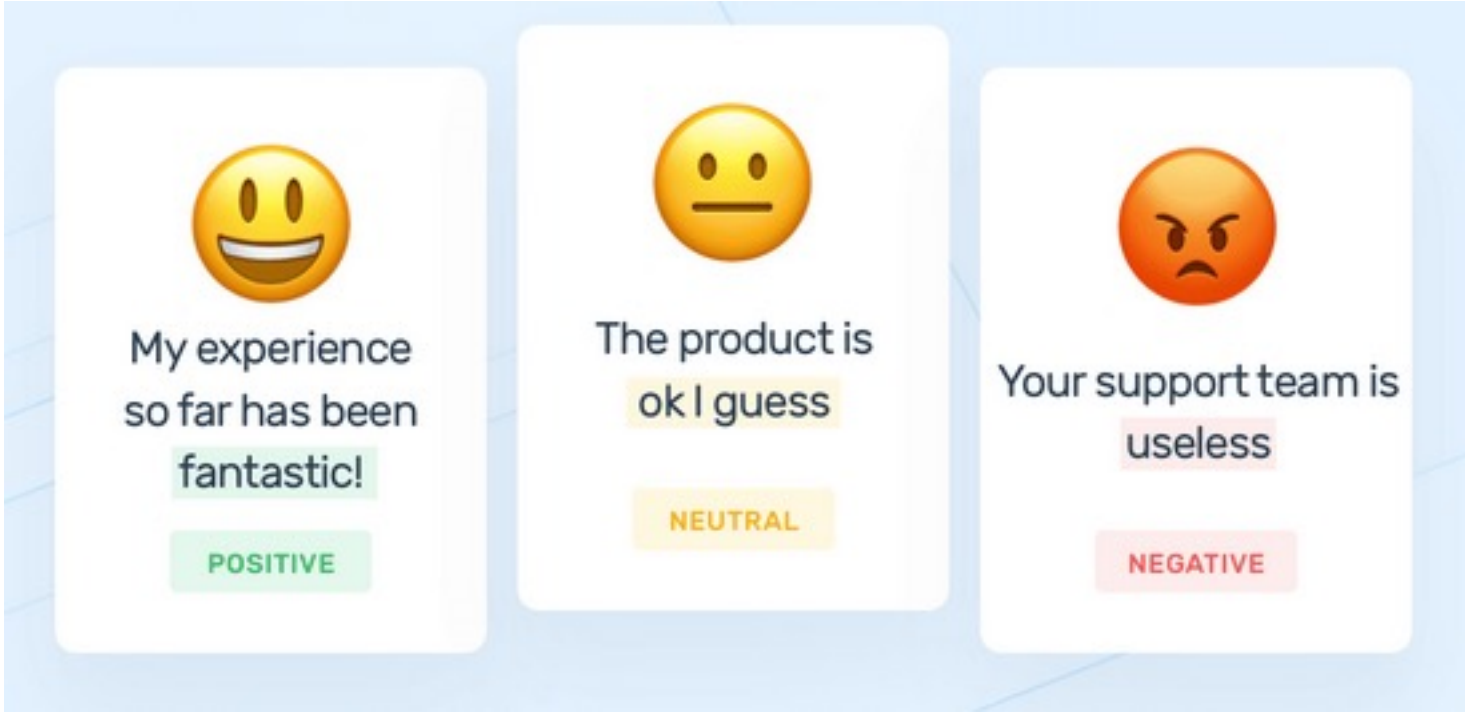




Sentiment Analysis is one of the Basic Text Mining Semantic Algorithms



What is sentiment analysis?



- Sentiment analysis (or opinion mining) uses NLP to determine whether data is positive, negative or neutral.

What is sentiment analysis?

- *Sentiment analysis* is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback and understand customer needs.
- *Sentiment analysis* is the process of detecting positive or negative sentiment in text.
  - It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Applications of Sentiment Analysis in Business

- Some of the most popular applications of sentiment analysis:
  - Social media monitoring
  - Customer support
  - Customer feedback
  - Brand monitoring and reputation management
  - Voice of customer (VoC)
  - Voice of employee
  - Product analysis
  - Market research and competitive research

Sentiment Analysis

- *Sentiment analysis* is the automated process of analyzing text to determine the sentiment expressed (positive, negative or neutral).
  - Some popular sentiment analysis applications include social media monitoring, customer support management, and analyzing customer feedback.
- In the background of sentiment analysis, advanced AI algorithms apply language deconstruction techniques, like tokenization, part-of-speech tagging, parsing, and lemmatization to break down and make sense of text.
  - Only then can machine learning software classify unstructured text by emotion and opinion.

Manual Computation

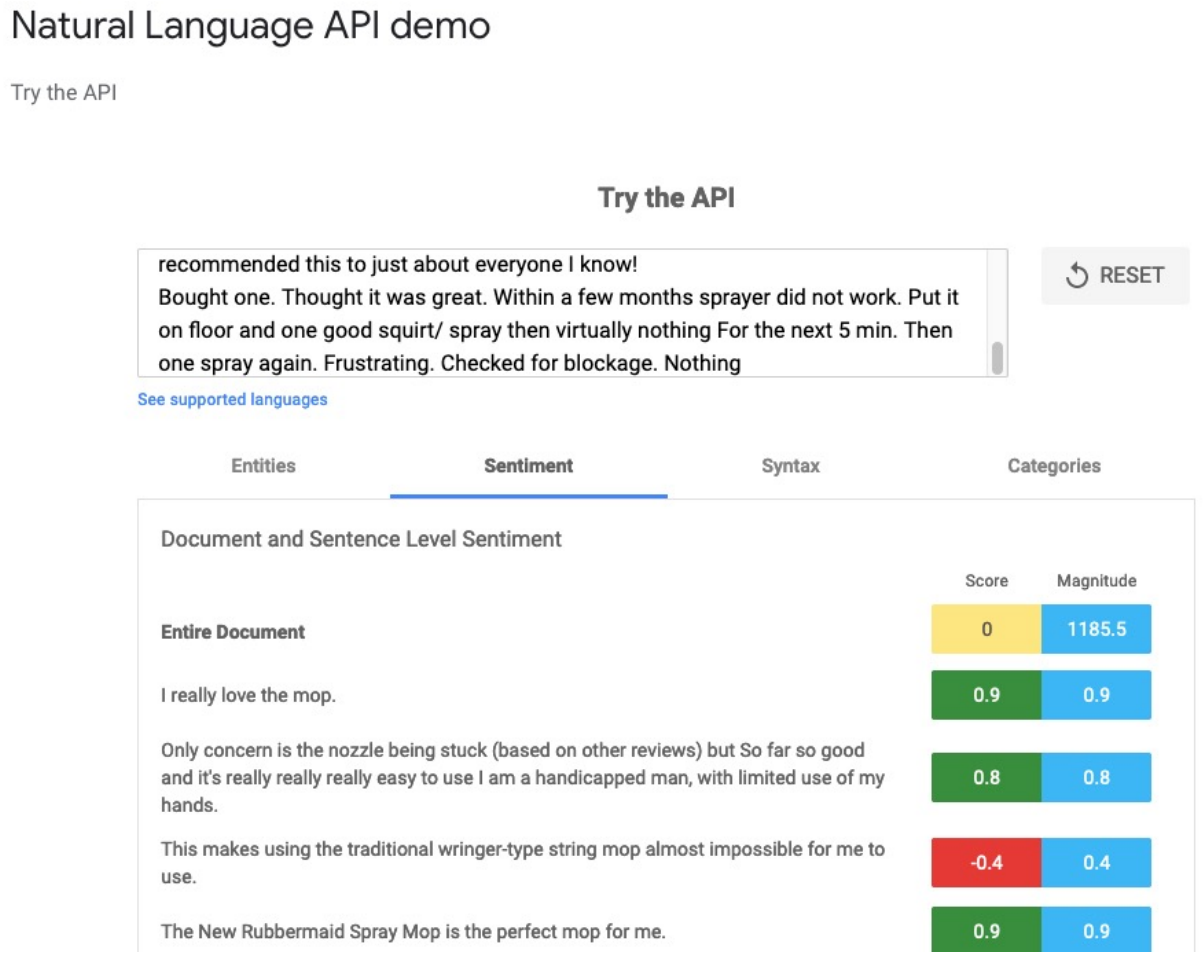
- Of course, a human can read texts, identify opinions, and detect nuances, but at what cost?
  - Manually tagging opinions can be arduous, given that the amount of data businesses receive is constantly growing.
  - Researchers who collected qualitative data (interviews, open ended questions) would use this method
  - We will do this in Session 8 under “Qualitative Data Analysis”

Commercial Products

- Which automated tools are available?
  - MonkeyLearn | Build custom, no-code sentiment analysis tools
  - IBM Watson | Discover APIs for sentiment analysis
  - Lexalytics | Uncover context-rich patterns and insights
  - MeaningCloud | Try out multilingual sentiment analysis
  - Rosette | An API for sentiment analysis in 30 languages
  - Repustate | Customizable sentiment analysis API
  - Clarabridge | Perform sentiment analysis on audio data
  - Aylien | Try a simple sentiment analysis API

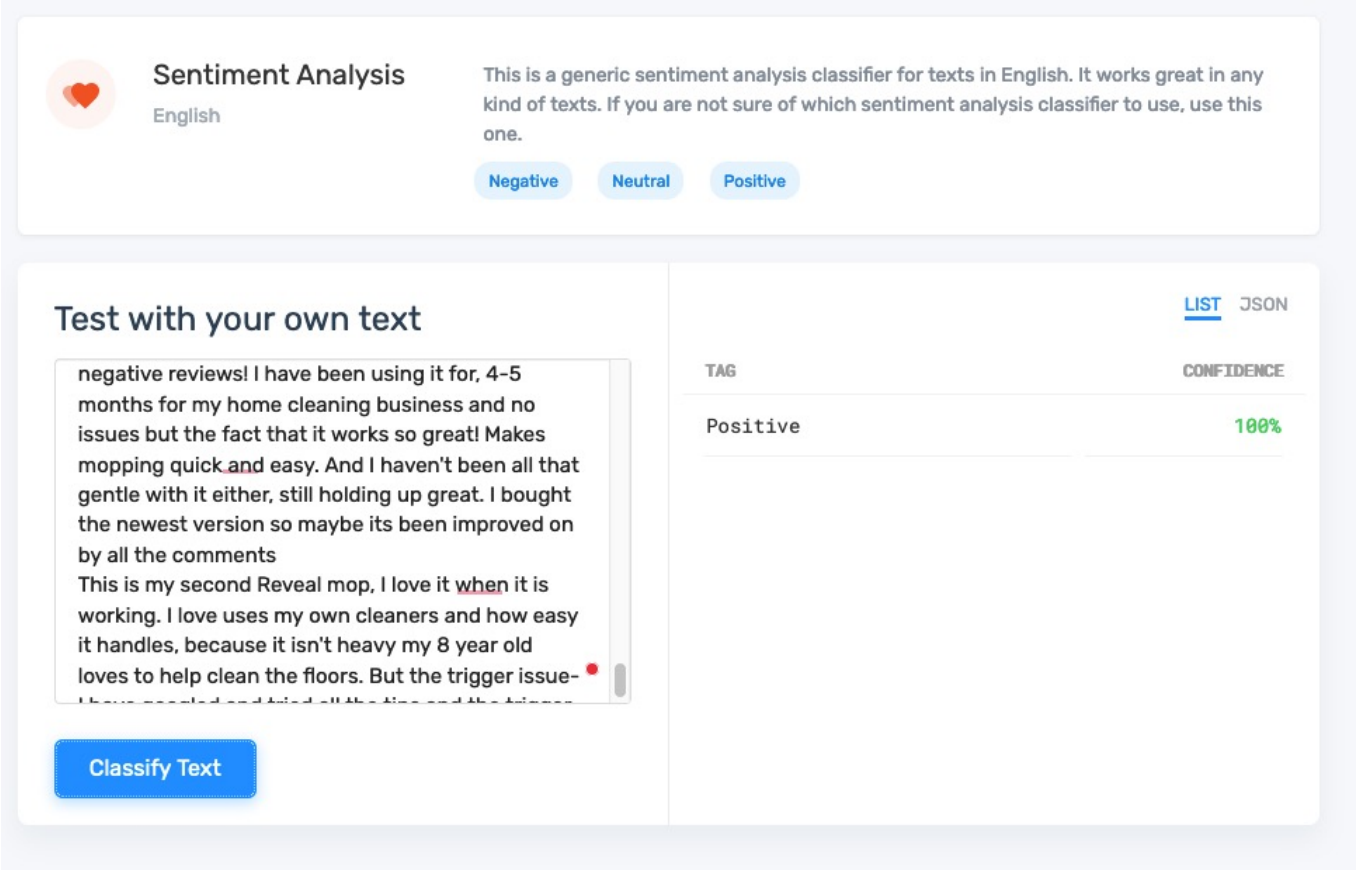
Demo A

- Google NLP API
  - https://cloud.google.com/natural-language#section-2



Demo B

- MonkeyLearn
  - https://app.monkeylearn.com/main/classifiers/cl_pi3C7JiL%EF%BB%BF/



Fine Grained Sentiment Analysis – The Most Popular

- Sentiment analysis models focus on polarity (*positive, negative, neutral*) but also on feelings and emotions (*angry, happy, sad*, etc), urgency (*urgent, not urgent*) and even intentions (*interested v. not interested*).
- Here is the most popular types of sentiment analysis:
- Fine-grained Sentiment Analysis
  - Very positive
  - Positive
  - Neutral
  - Negative
  - Very negative
- This is usually referred to as fine-grained sentiment analysis, and could be used to interpret 5-star ratings in a review, for example:
  - Very Positive = 5 stars
  - Very Negative = 1 star

The Process

- We will do a sentiment analysis using Excel by hand to see what goes into this process.
- Open *Waffle Iron Reviews Sentiment Analysis.xlxs* file
- Read through some of the reviews to get the tone of the language and how it is used
- Come up with a list of positive and negative words (10)
- Use COUNTIFF to count how many rows have the positive words and how many rows have the negative words in them

How to do it – Example A

- Essentially, you want to answer the questions:

  Are customer who bout this cast-iron waffle maker happy with the product?

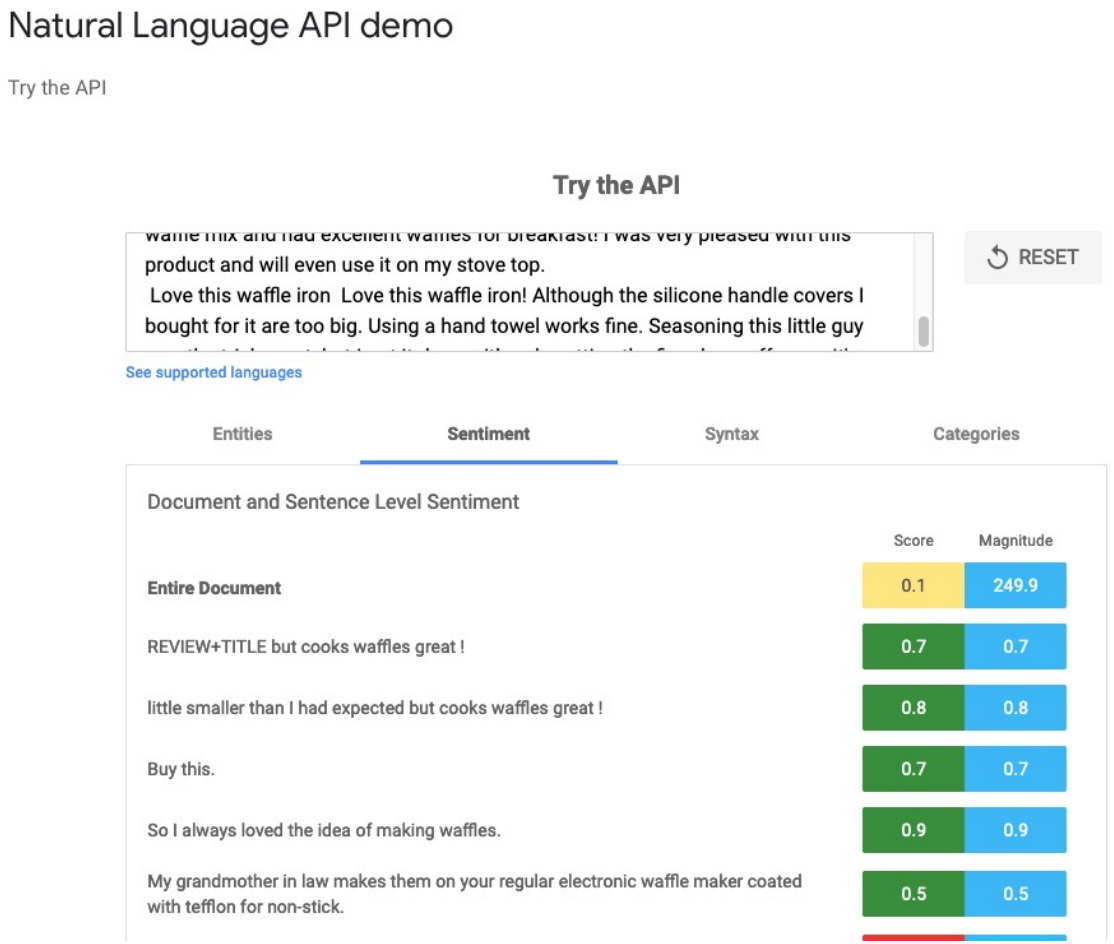
  Does the average rating score for the product match the sentiment analysis?

- Sentiment = (Positive - Negative)/(Positive + Negative)

Total Positive	83	
Total Negantive	14	
Difference	69	
Sum	97	
Sentiment	0.711	Positive
Average Rating	3.856	Positive

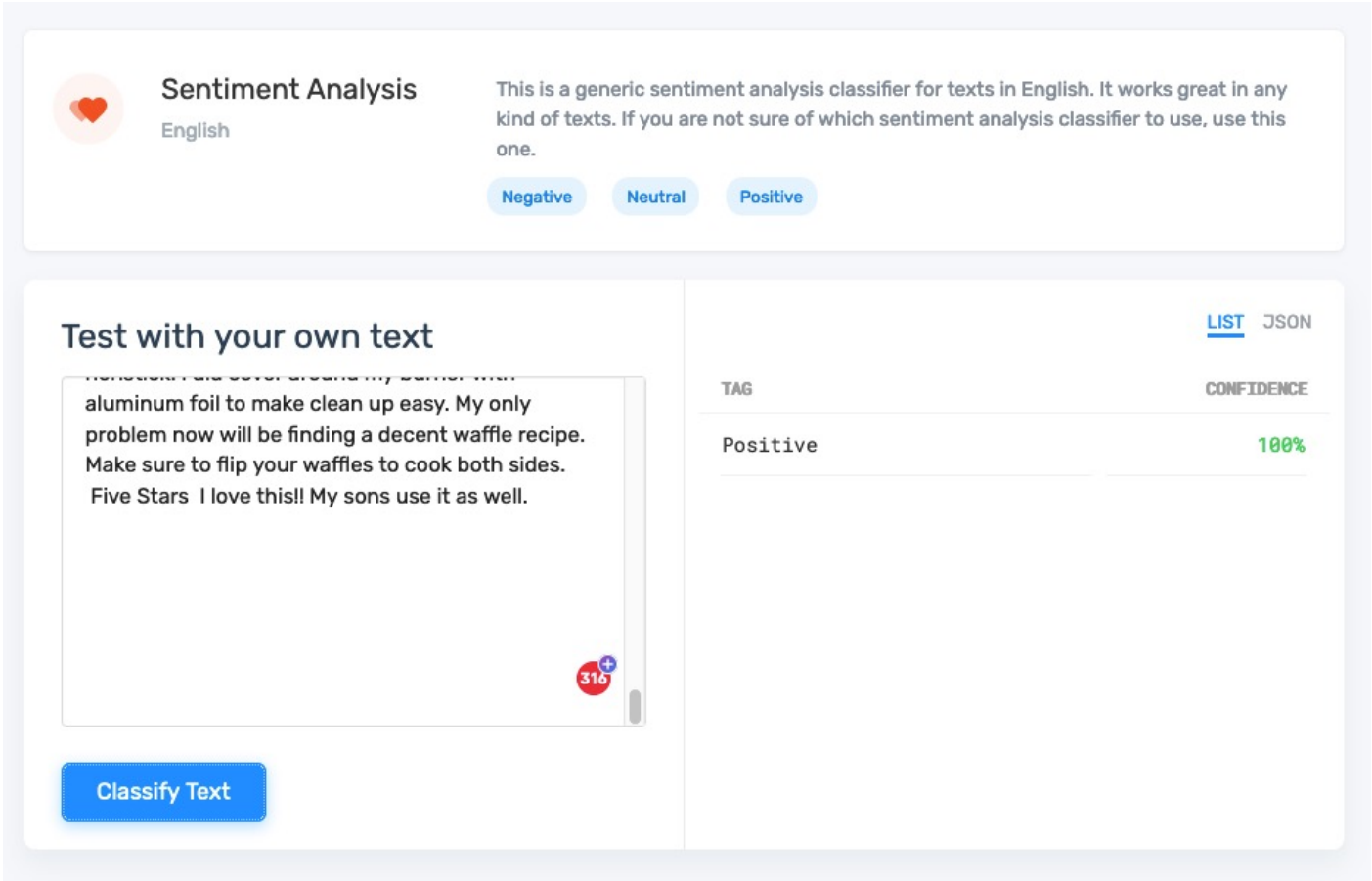
Demo A - Waffle Iron Reviews

- Google NLP API
  - https://cloud.google.com/natural-language#section-2



Demo B - Waffle Iron Reviews

- MonkeyLearn
  - https://app.monkeylearn.com/main/classifiers/cl_pi3C7JiL%EF%BB%BF/



Case Study A

- The MASY program director is interested in making sure the MASY students are satisfied with the program. There were major changes during the pandemic in all classes online. Did that cause a lot of dissatisfaction?
- She ran a survey polling students on how they feel about online modality. She did not quantize her survey but used open ended questions.
- She hires you to analyze the survey results
- You have one tool to use, sentiment analysis suing Excel, how would use it to give her an answer? Compare the results to what a commercial analyzer would do (use Google NLP API in demo mode)

Case Study A

- Use the response to the survey collected by your instructor.
- Your team should create a table of the results in Excel.
- Read over the responses – brainstorm with your team at least 10 positive and 10 negative single word terms to use for sentiment analysis
- Use Excel and the COUNTIF function to generate a sentiment score using these 20 keywords
- Use question 1 for sentiment analysis.
- You may repeat for question 2 and see if you obtain the same result but it is not required.
- Use the question 1 data and the Google NLP API demo function to generate a similar score – compare the two. Explain any discrepancies
- What would you report to the program director about hwo students feel about online classes?