That FinTech Feeling

FinTech

Lesson 12.2

Class Objectives

By the end of the class, you will be able to:

Identify the applications of sentiment analysis for FinTech.

Retrieve data from news feeds to analyze sentiments and tone.

Perform data preparation techniques oriented to sentiment analysis.

Apply natural language processing through `NLTK` and `VADER` to classify news sentiment

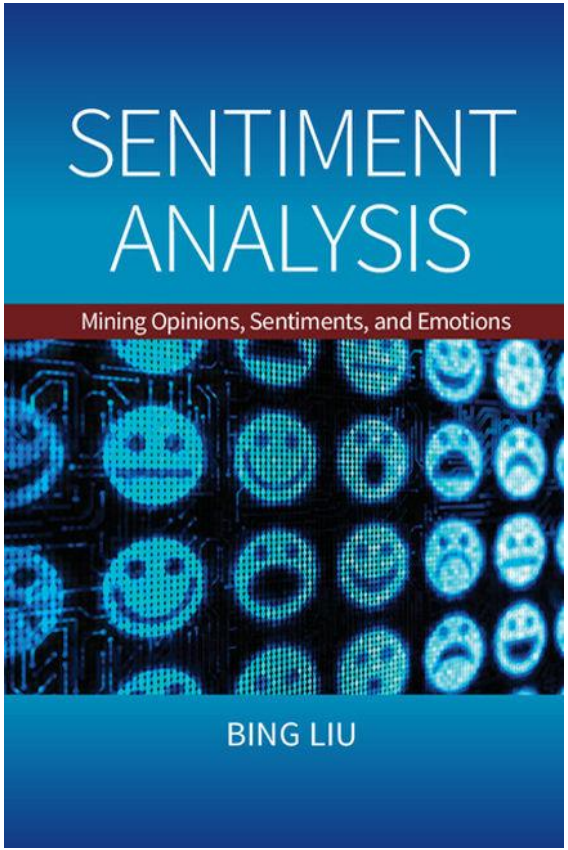
Conduct tone analysis using the IBM Tone Analyzer cloud service.

Sentiment Analysis

Sentiment Analysis

Sentiment analysis is a field within NLP, it's defined as "the computational study of people's opinions, sentiments, emotions, and attitudes."

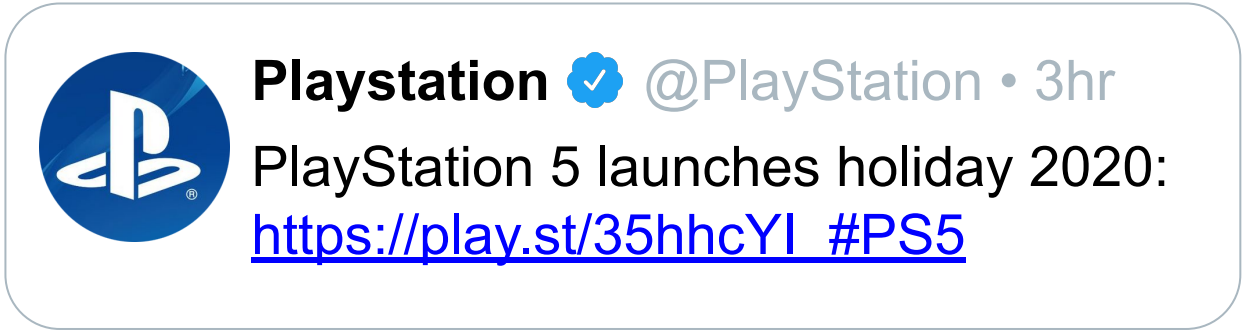
Book Suggestion



Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, 1st Edition
by Bing Liu
(Cambridge University Press, 2015)

Sentiment Analysis

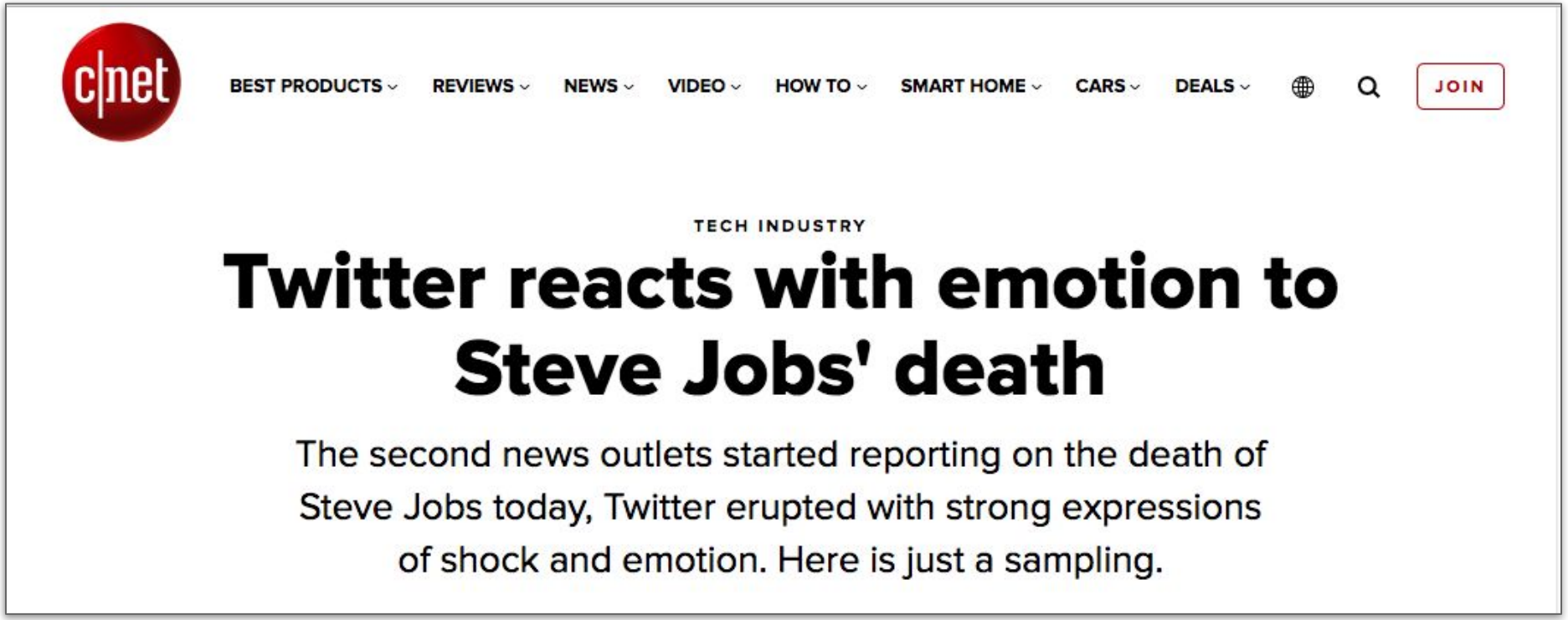
Opinions of people over the Internet are constantly growing, a large number of texts expressing points of view are available in review sites, forums, blogs, and social media.



What kind of sentiments could be analyzed using algorithms?

Polarity (positive, neutral, negative)

Emotions (angry, happy, sad, etc.)



Intention (detects what people want to do)

Example: "Bitcoin price is rising, I'll call my broker"

How does Sentiment Analysis Work?

Ruled-Based Analysis

Perform sentiment analysis based on a set of manually crafted rules using NLP techniques such as stemming, tokenization, part of speech tagging and parsing.

Basic Operation Example:

Most people dislike cloudy days, but I think they are beautiful so I love them

Ruled-Based Analysis

Define two lists of polarized words (e.g. positive words and negative words).

Positive Words

Acclaimed

Beneficial

Celebrated.

Distinguished

Effective

Genuine

Impressive

Optimistic

Negative Words

Abysmal

Belligerent

Contradictory

Deplorable

Gruesome

Horrendous

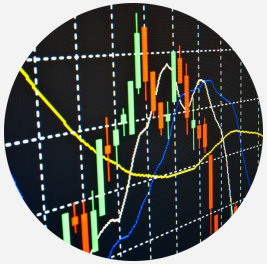
Insidious

Nonsense

Ruled-Based Analysis

Given a text:

1. Count the number of positive words that appear in the text.



jones121

$AAPL My favorite semiconductor stock, Nvidia, was just upgraded by the Royal Bank of Canada (RBC), rallying the stock 1.3% in today's trade. RBC moved its price target from $190 to $217, which would represent a more than 17% upside from NVDA's current share price of $185.

Ruled-Based Analysis

Given a text:

2. Count the number of negative words that appear in the text.

CNBC

$TSLA Tesla's stock is down nearly 35% for the year as of Tuesday's close of $217.10 a share but the stock has slowly come back after hitting a low of $179 a share last week.

Ruled-Based Analysis

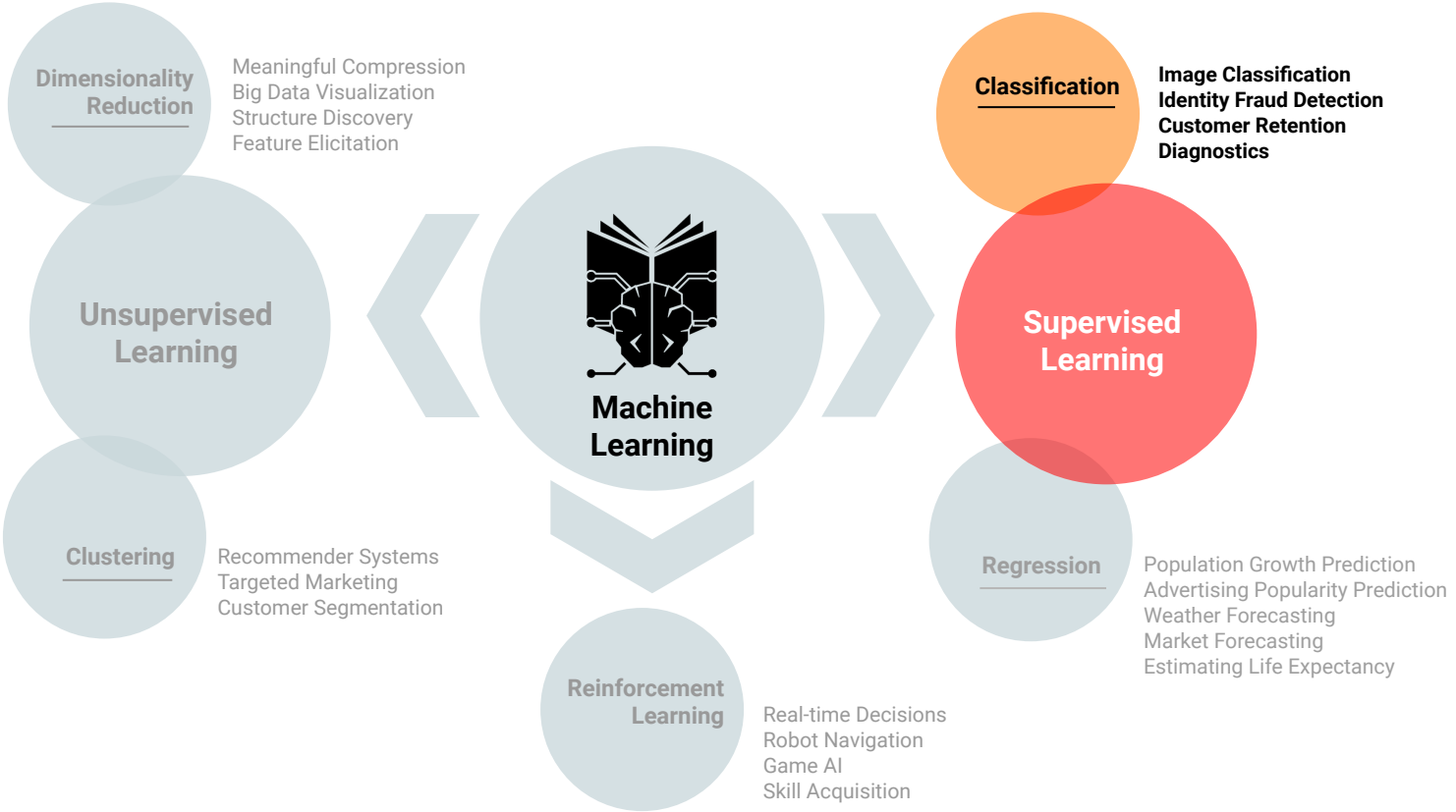
If the number of positive word appearances is greater than the number of negative word appearances return a positive sentiment, conversely, return a negative sentiment. Otherwise, return neutral

CNBC

Samsung Electronics is one of the world's worst-performing technology stocks this year with billions wiped of its value as factors from increased smartphone competition to worries about the health of the semiconductor market have weighed on the company's performance.

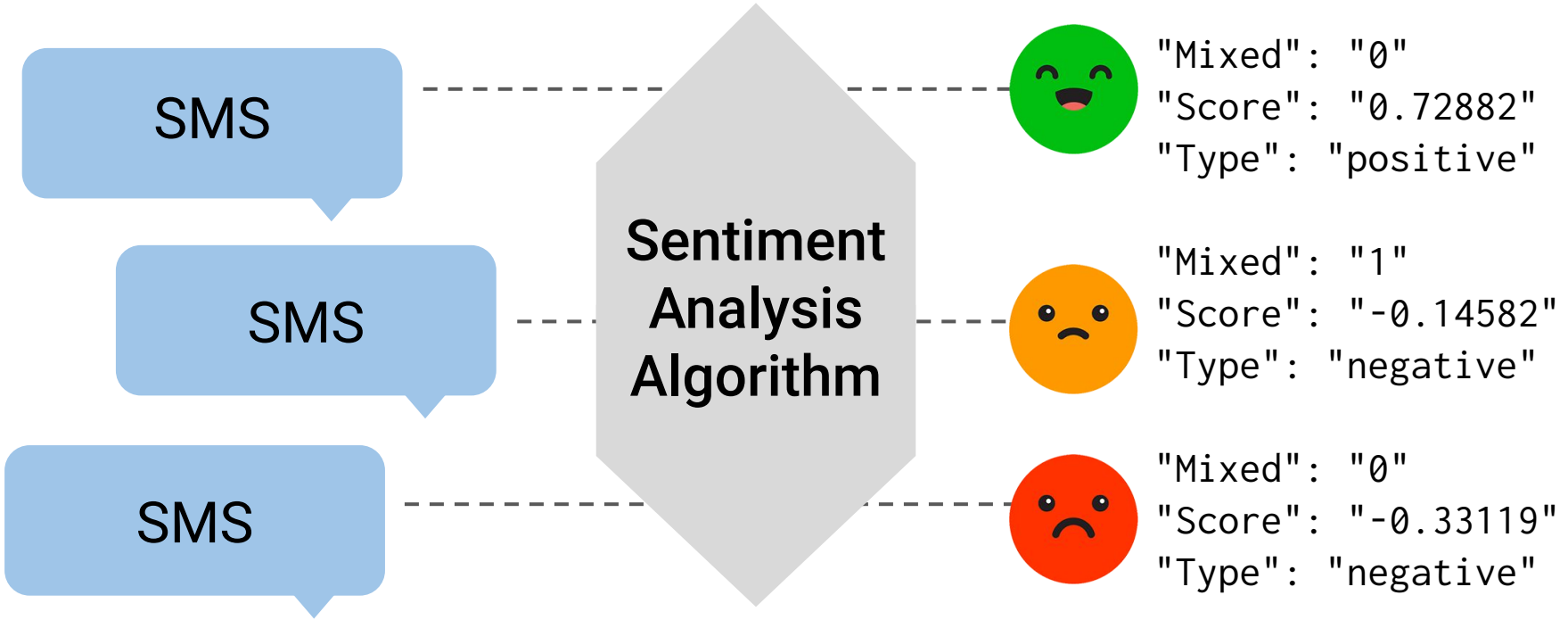
Shares of the South Korean giant are down over 11 percent this year, more than major U.S. and other Asian technology stocks. Over $39.4 billion has been erased from Samsung's market capitalization year-to-date. It's a far cry from the 41 percent rally Samsung shares saw in 2017.

Automatic sentiment analysis is normally done using supervised machine learning techniques using classification algorithms such as decision trees, support vector machines, naive Bayes, or neural networks.



Automatic Sentiment Analysis

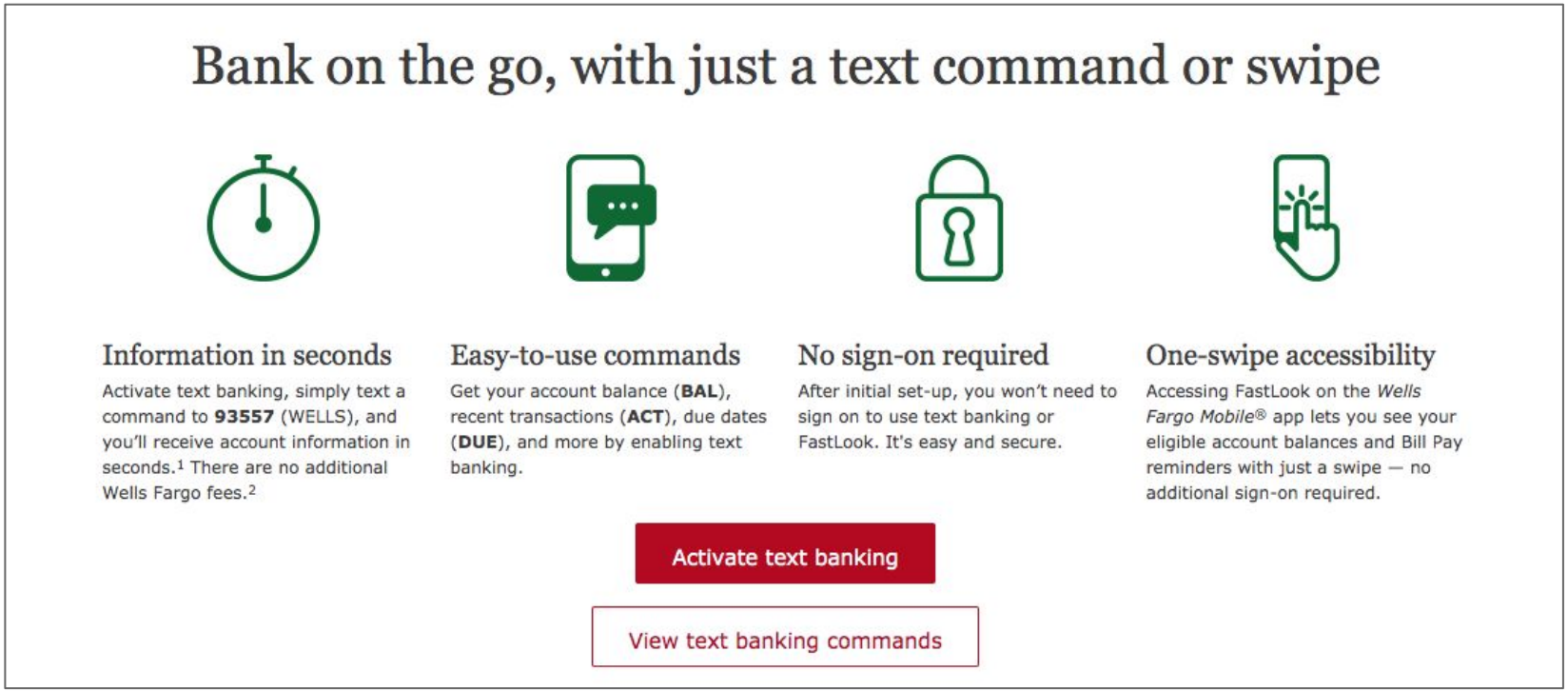
On this approach the classification model is fed with a text and returns the corresponding category as the sentiment or tone, e.g. positive, neutral or negative.



How Sentiment Analysis is Used Today in FinTech?

Sentiment Analysis

Banks are using instant messaging to improve customer service and improve response time to common questions and claims made by Clients.



Sentiment Analysis

Brokers are enhancing investment decision by listening to the news and social media.

Sentiment Analysis

C Level executives are using sentiments as additional data to make business decisions and growing plans.

Example: Arby's team discovered that their customers mentioned the brand on social media quite a lot. Their love for Arby's sauces was not only noticeable, but also measurable, thanks to Oracle's social listening tools included in their Marketing Cloud platform. Arby's team decided to capitalize and create a new advertising campaign. Using visuals in the style of 1950's movie posters, Arby's brand marketing team created stories of what would happen if there was no sauce while preparing the BBQ.

Challenges of Sentiment Analysis

- Multilingual sentiment analysis is quite complex since local and cultural aspects impact the sentiment.
- People express opinions in complex ways.
- In opinion texts, lexical content alone can be misleading.
- Voice tone is not considered.

Can a stock feel sadness if it opens with losses Today?

It's difficult for a stock to have feelings, but definitely we can use sentiment analysis to understand what are the sentiments expressed about a stock on the news headlines to support investment decisions.

Is it possible for your bank account to suggest a travel destinations based on your tweets from last year?

If you allow your bank to follow you on Twitter, it's possible to score the sentiment of your tweets, even the tone you express on your posts. Your bank can give you some travel recommendations to boost your mood.

Julie Simon @jsimon53 • December 26, 2017

Would love to wake up on this beach in Thailand!

Can an auction website prevents fraud from the conversation between a potential buyer and a dealer?

Using tone analysis it's possible to follow the utterances between a potential buyer and a dealer, you can score the emotions they are expressing on their conversation to prevent a fraud.

Robert85

I would love to purchase your vintage suitcase. This suitcase is perfect! I will send you the money via paypal and then you will send me this amazing suitcase with beautiful leather and wheels.

What about your privacy on this sentiment analysis revolution?

This is definitely a major concern, you should have consent from people to analyse the sentiment of their writing communications, as well as, to have an storage and security infrastructure in compliance to user privacy regulations.

Terms relevance

Terms relevance

Term relevance is quite important on sentiment analysis since leads to a better understanding of human speech.

What is a Corpus?

Corpus

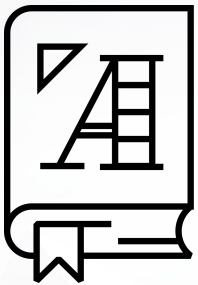
A corpus (corpora in plural) is a large, structured and organized collection of text documents that normally verses on a specific matter

A corpus may contain texts in a single language (monolingual corpus) or text data in multiple languages (multilingual corpus).

TF-IDF

Term Frequency

Inverse Document Frequency



TF-IDF a weighting factor intended to measure how important a word is to a document in a collection of documents or corpus.

TF-IDF

A measure intended to reflect the importance of a word in a text.

- TF—Term Frequency: A count of the word in a document
- IDF—Inverse Document Frequency:

$$\log\left(\frac{\textit{total number of documents}}{\textit{number of documents containing target word}}\right)$$

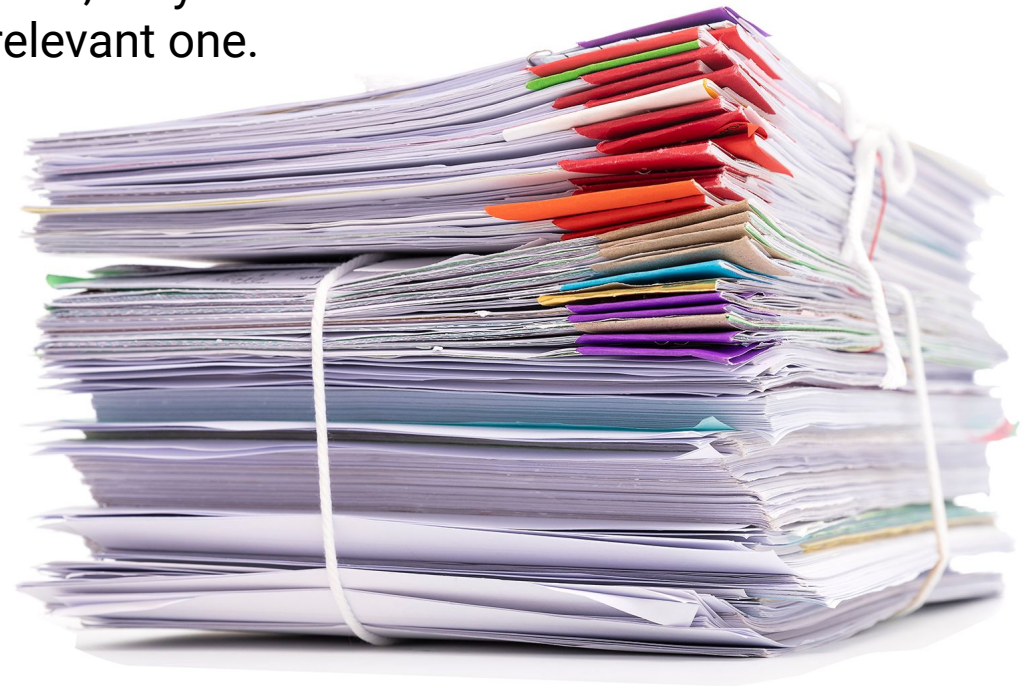
- IDF: The more documents that include the term, the lower the IDF score.
- TF–IDF is the product of the two. TF drives up the score, but IDF will bring it down if the word occurs in all or many documents.

The Rationale Behind TF-IDF

TF indicates that if a word appears multiple times in a document, it can be concluded that it's relevant and should be more meaningful than other words in the same text.

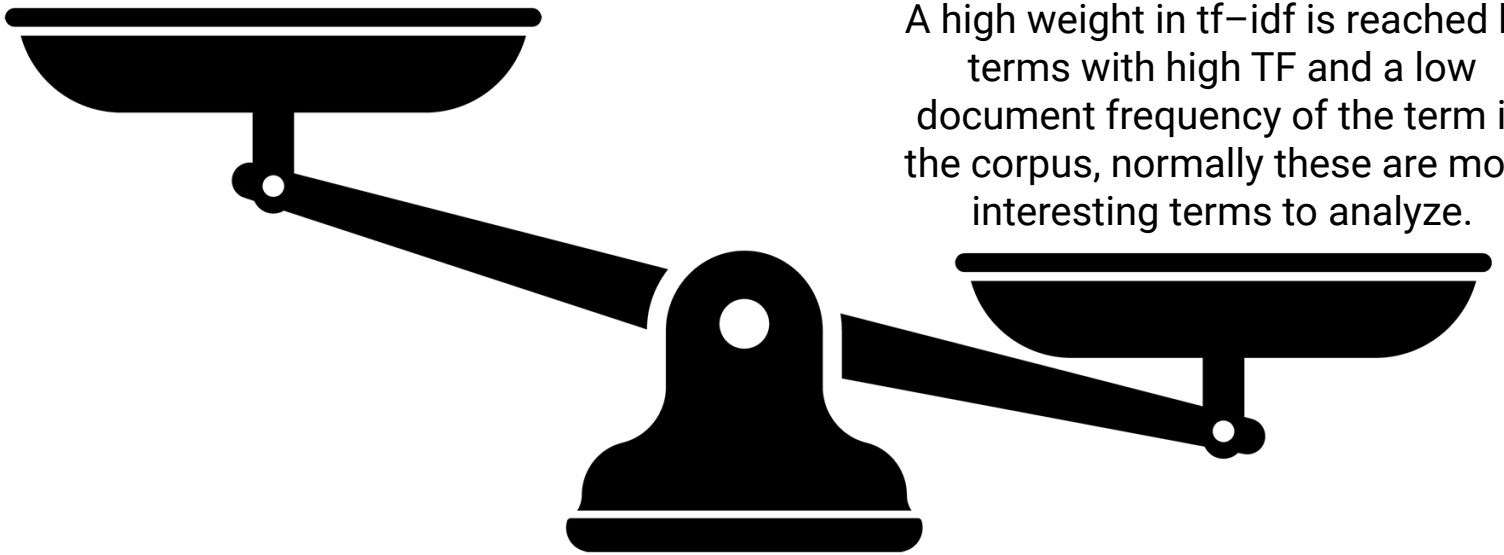
The Rationale Behind TF-IDF

IDF comes to action when you're analyzing several documents. If a word also appears many times along a collection of documents, maybe it's just a frequent word and not a relevant one.



The Weights of TF-IDF on a Corpus

A low weight in TF-IDF is reached by terms with low TF and a high document frequency of the term in the corpus, normally these are quite common terms across the corpus that could be less interesting to analyze.



A high weight in tf–idf is reached by terms with high TF and a low document frequency of the term in the corpus, normally these are more interesting terms to analyze.

TF: Bag of Words

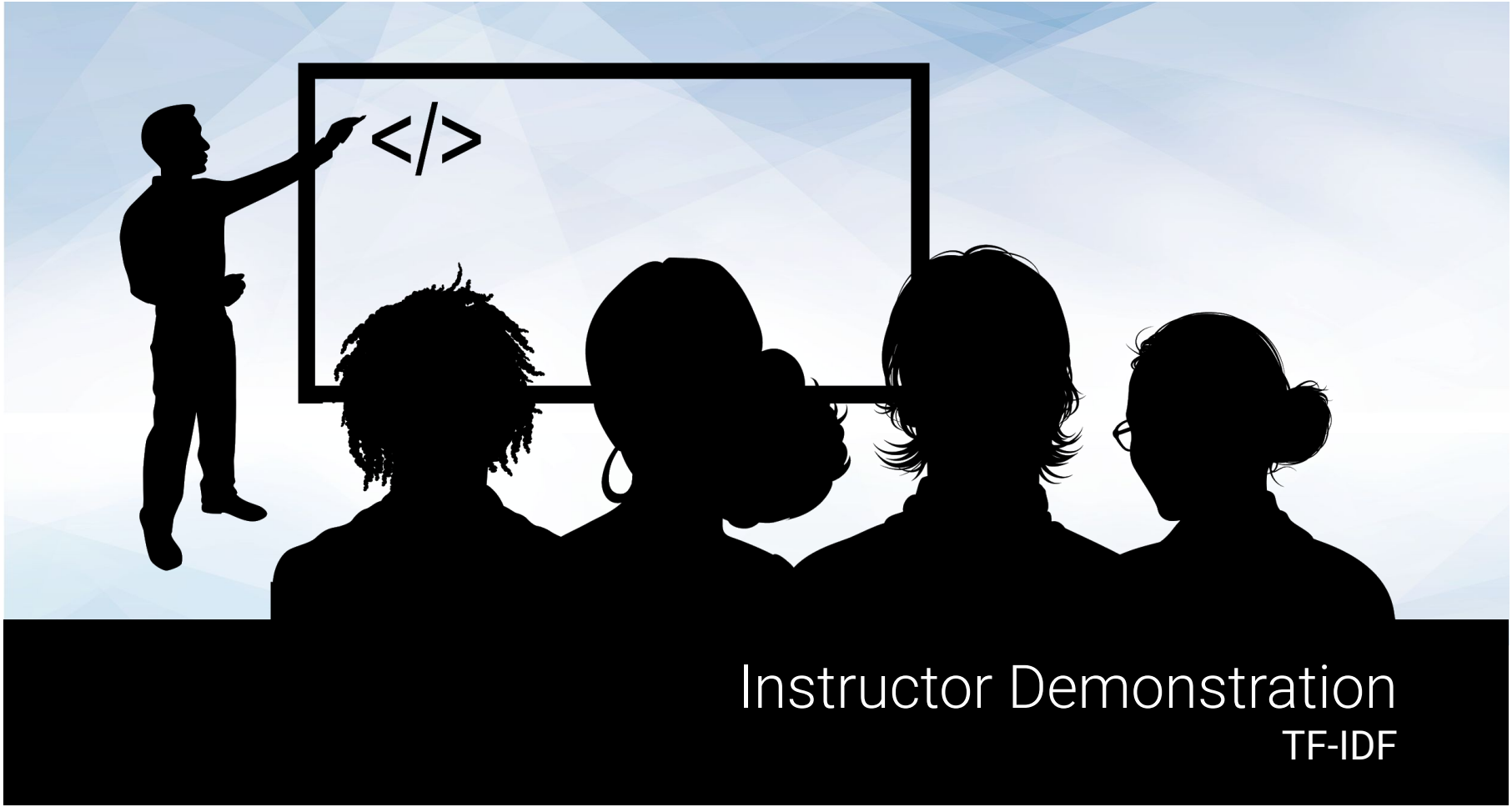
The methods used for term frequency use a “bag of words” approach.

Each document is represented by a “bag of words” where grammar and word order are disregarded, but multiplicity is kept.

The Bag of Words Representation

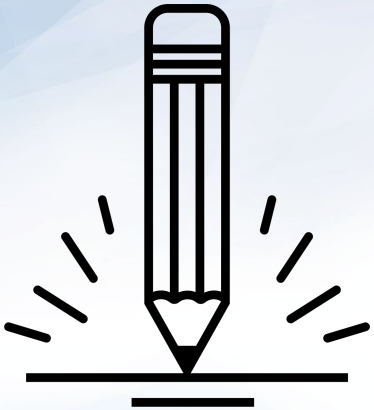
I love this movie! It’s sweet but with satirical humor. The dialogue is great and the adventure scenes are fun...It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I’ve seen it several times, and I'm always happy to see it again whenever I have a friend who hasn’t seen it yet!

It	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1



Instructor Demonstration

TF-IDF

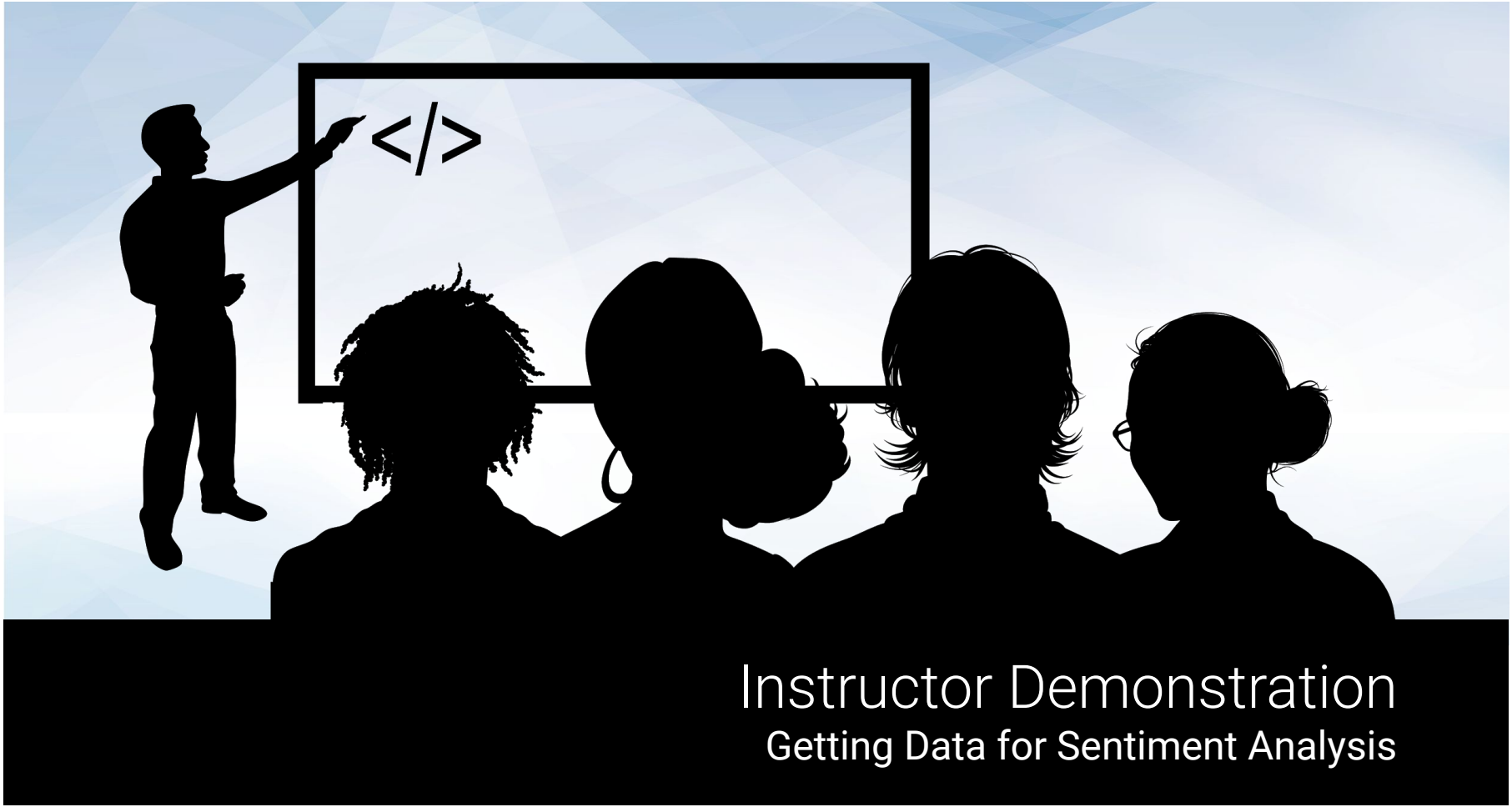


Activity: Bossy Words

In this activity, you will create a word cloud based on TF-IDF weights.

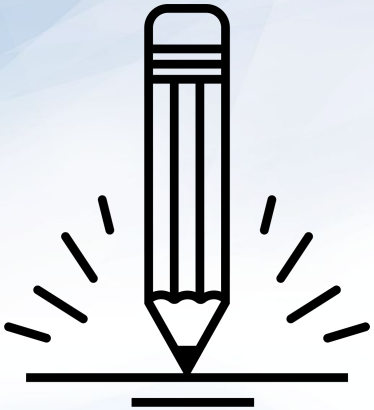
Suggested Time:
20 minutes

Time’s Up! Let’s Review.



Instructor Demonstration

Getting Data for Sentiment Analysis



Activity: The Voice of the Crisis

In this activity you will use the News API to retrieve news articles on English and French about the financial crisis of 2008.

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

VADER Sentiment

VADER Sentiment

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a tool used to score the sentiment polarity of human speech as positive, neutral or negative based on a set of rules and a lexicon (a kind of list of words).

		
"Mixed": "0" "Score": "0.72882" "Type": "positive"	"Mixed": "1" "Score": "-0.14582" "Type": "negative"	"Mixed": "0" "Score": "-0.33119" "Type": "negative"

How does VADER works?

VADER Sentiment

In general terms, VADER calculates the polarity of a text based on a combination of rules and a list of words that where manually tagged as positive or negative according to their semantic orientation.

		
"Mixed": "0" "Score": "0.72882" "Type": "positive"	"Mixed": "1" "Score": "-0.14582" "Type": "negative"	"Mixed": "0" "Score": "-0.33119" "Type": "negative"

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

VADER Sentiment

VADER not only gets the positive, negative or neutral score but also tells us about how positive or negative a sentiment is.

		
"Mixed": "0" "Score": "0.72882" "Type": "positive"	"Mixed": "1" "Score": "-0.14582" "Type": "negative"	"Mixed": "0" "Score": "-0.33119" "Type": "negative"

VADER Sentiment

VADER generates four scores for each analyzed text:

01 Positive (pos)

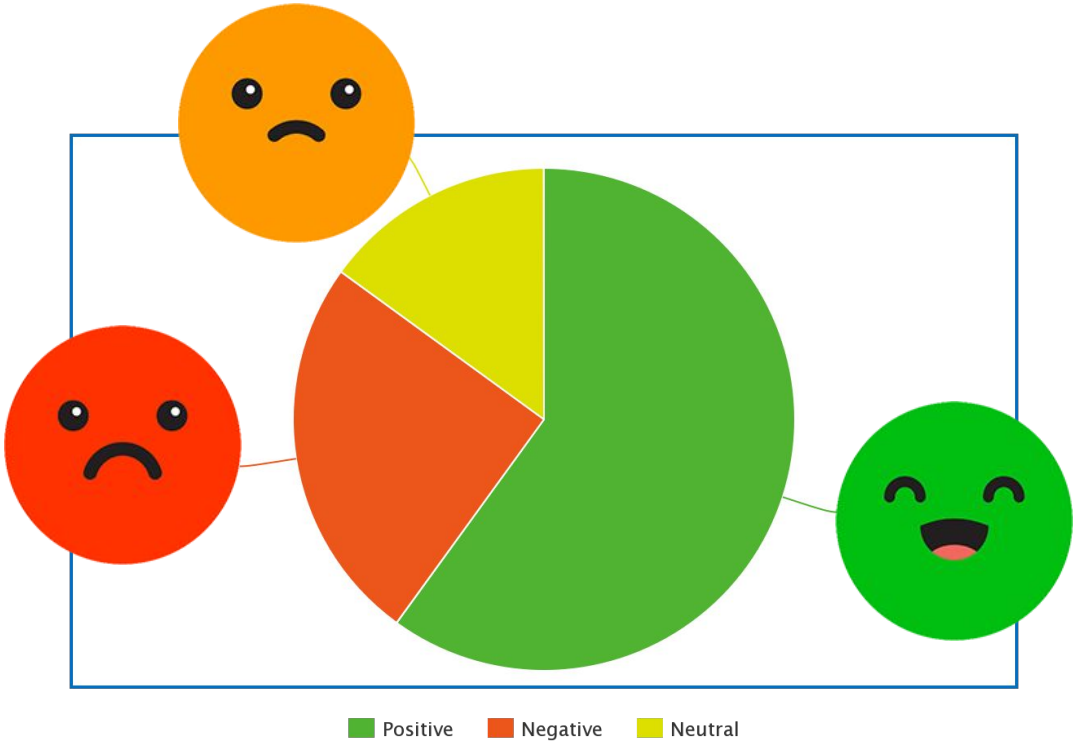
02 Negative (neg)

03 Neutral (neu)

04 Compound

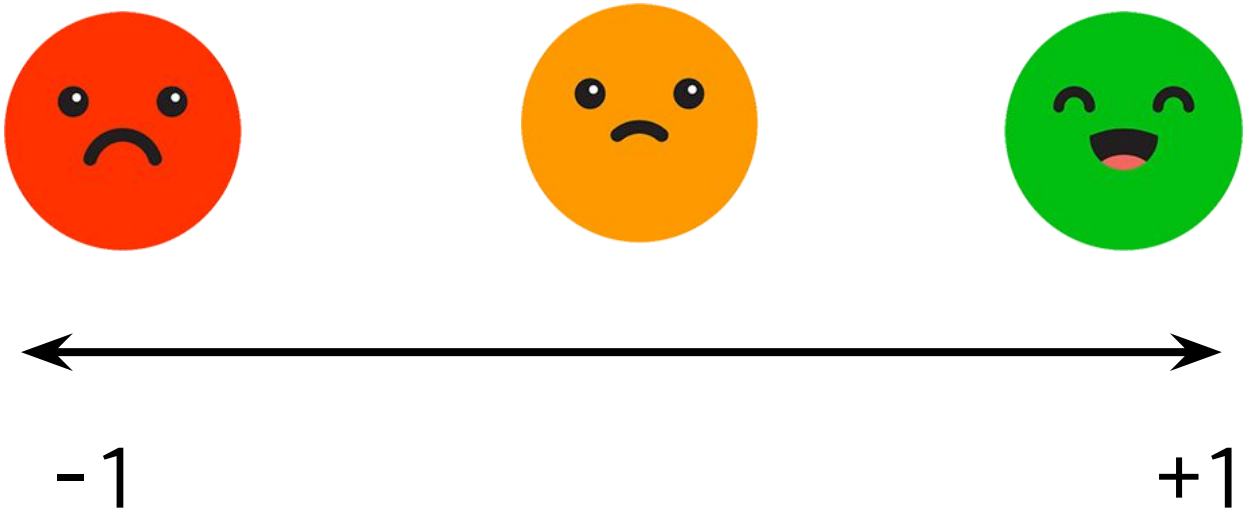
VADER Sentiment

Pos, neg and neu range from 0 to 1.



VADER Sentiment

Compound is a normalized score and goes from -1 (most extreme negative) and +1 (most extreme positive)



VADER Sentiment

positive sentiment:

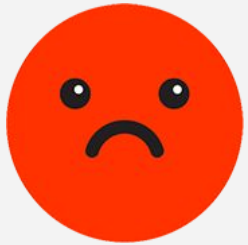
compound score
>= 0.05

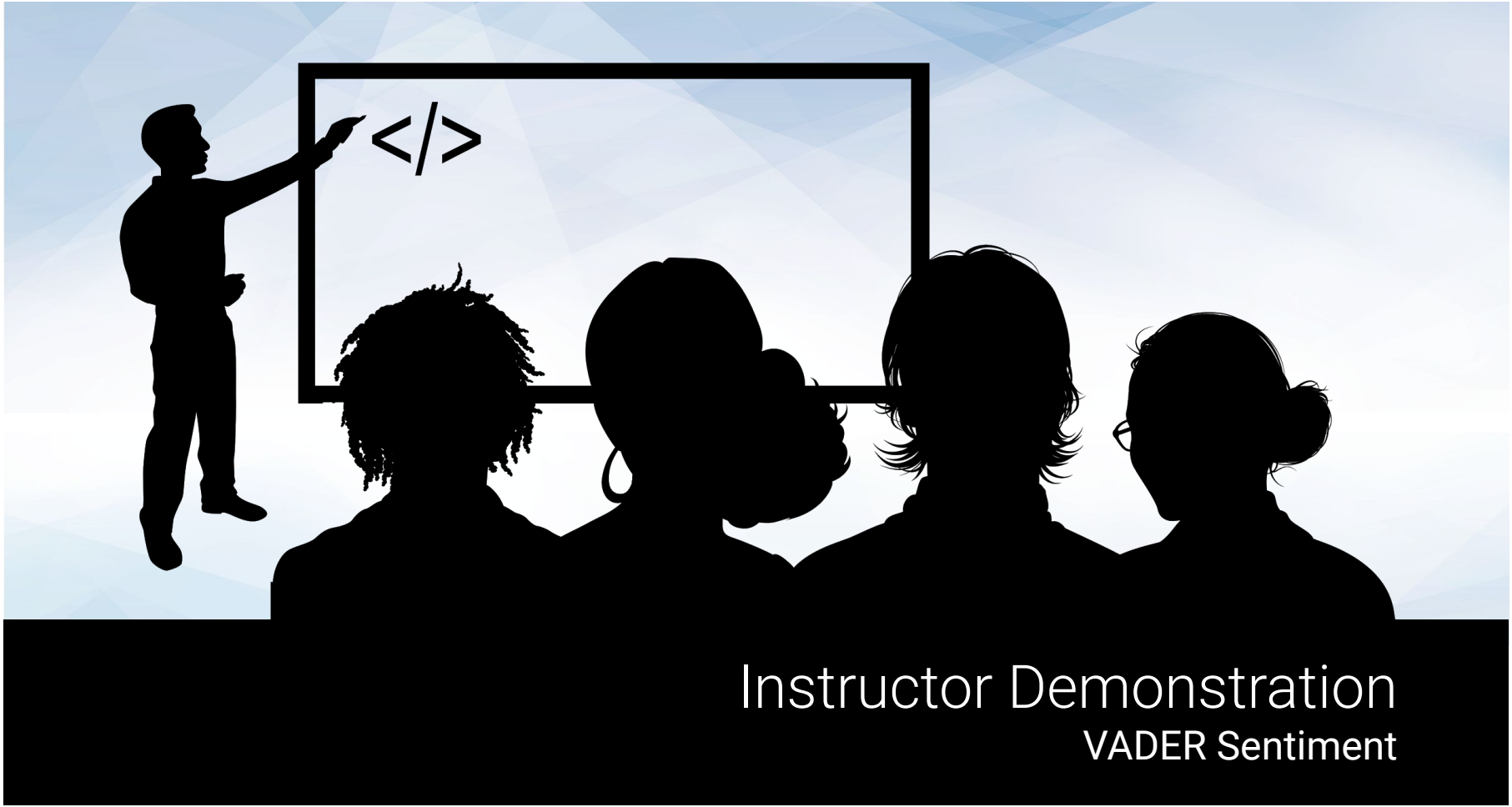
neutral sentiment:

compound score
between
-0.05 and 0.05

negative sentiment:

compound score
<= -0.05





Instructor Demonstration

VADER Sentiment


Break

Activity:

The Feelings of the Crisis

In this activity, you will use VADER to score the sentiment of news' title and text to verify if they have the same sentiment..

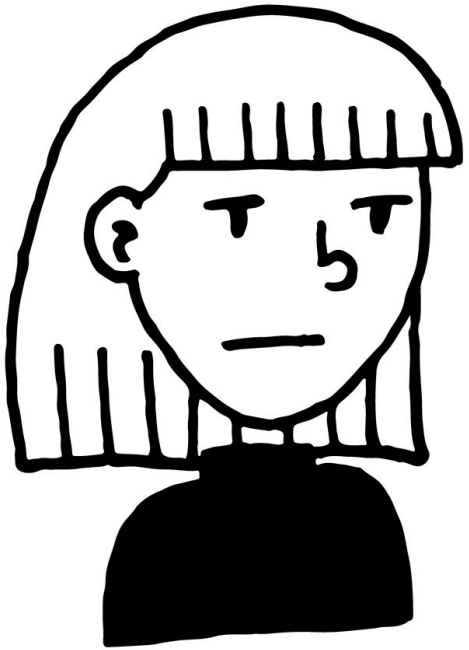
Suggested Time:
20 minutes

Time’s Up! Let’s Review.

Tone Analysis

The Complexity of Human Speech

Human speech goes beyond being positive, negative or neutral, as humans we express emotions that can be felt on the tone.

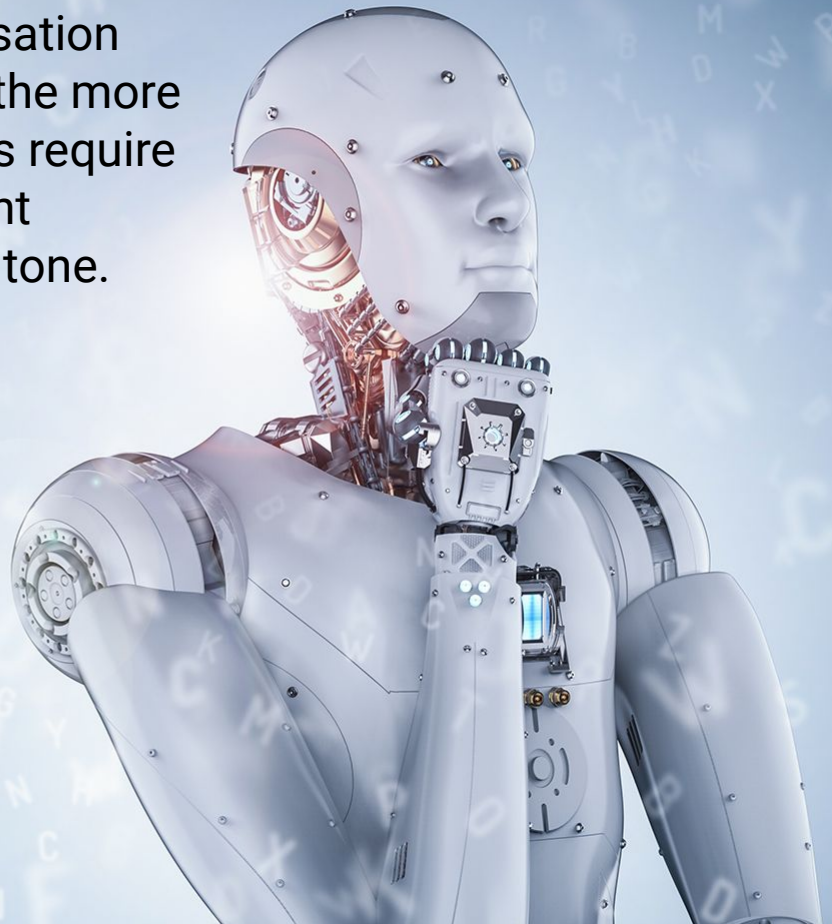


VERY IMPRESSIVE. CAN'T YOU SEE MY EXCITEMENT?

How does an algorithm identify tone?

Tone Analysis

Defining the tone of a text or conversation depends on several aspects, one of the more complex is the context so algorithms require to analyze tons of texts from different contexts to learn how to identify the tone.



IBM Watson Tone Analyzer

Tone Analyzer is a cloud service from IBM Watson that is able to measure the tone of written text. This service is able to analyze tone in English and French conversations and you can used in Python via its API.

GO TO:

cloud.ibm.com/catalog/services/tone-analyzer

Questions?