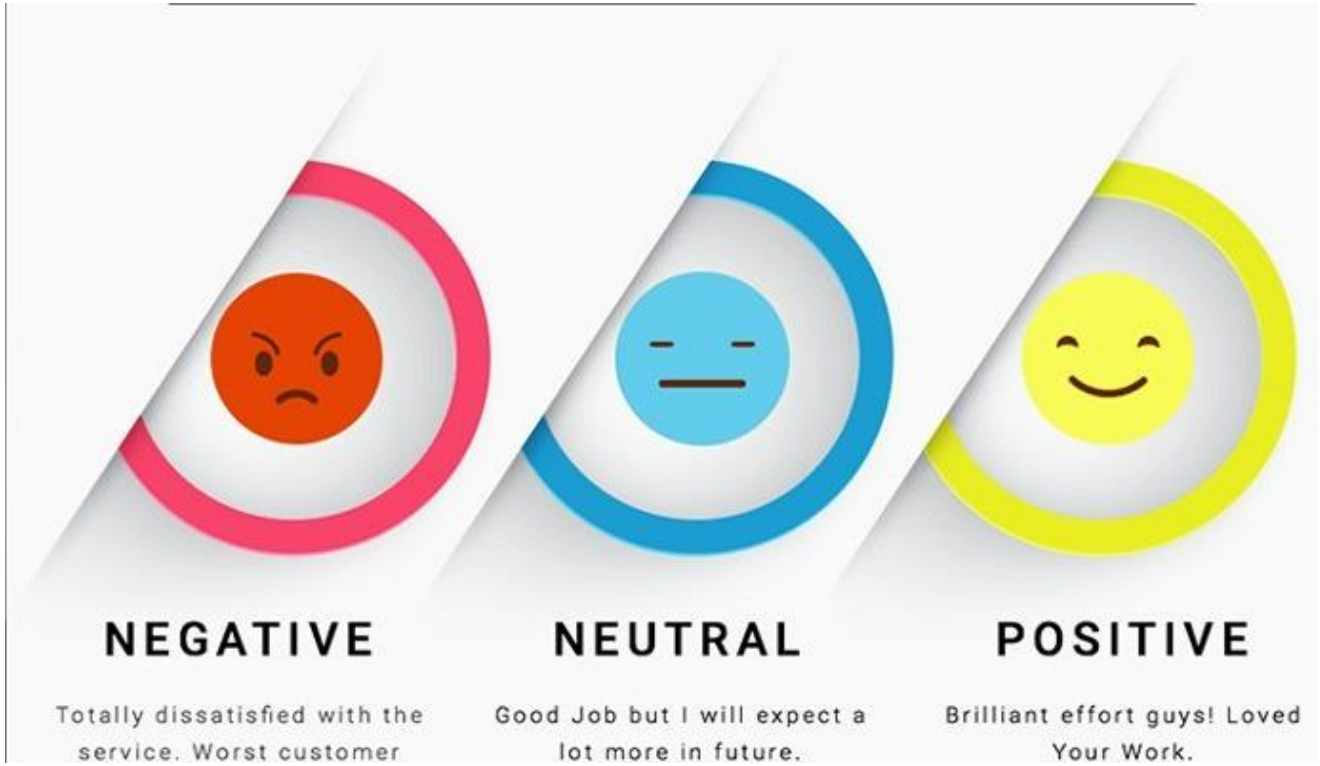
Sentiment Analysis using VADER

- Kathirmani Sukumar

Sentiment Analysis

“The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.”


NEGATIVE
Totally dissatisfied with the service. Worst customer
NEUTRAL
Good Job but I will expect a lot more in future.
POSITIVE
Brilliant effort guys! Loved Your Work.

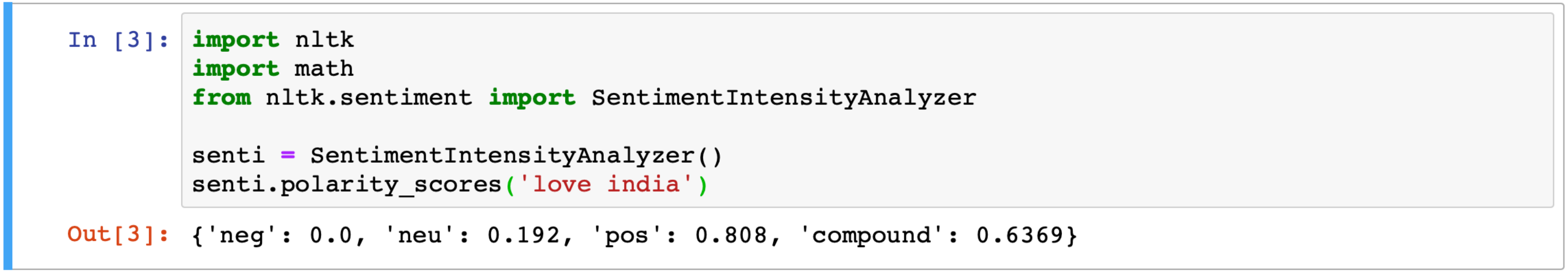
Source: https://www.kdnuggets.com/images/sentiment-fig-1-689.jpg

Types

- Supervised Models
  - If we have labelled datasets, where against each review we have labels like positive, negative and neutral
  - Pros
    - Can handle regional languages
    - Many a times, accuracy might out perform unsupervised analysis
  - Cons:
    - We need to have labels before starting the analysis. This is time consuming process
- VADER analysis:
  - If we have do not have labelled data, we can use sentiment analysis using VADER package
  - Pros
    - Labels not required
    - Sentiment intensity can be obtained
  - Cons
    - Support to all regional languages is not be available currently
    - Does not work well for long documents
    - Cannot identify sarcasm

VADER

- VADER: Valence Aware Dictionary and sEntiment Reasoner
- Gives intensity to most of the commonly used positive and negative words
- Intensity of each positive or negative word ranges between -4 to +4
- Gives overall compound score for any sentences. Ranges between -1 to +1



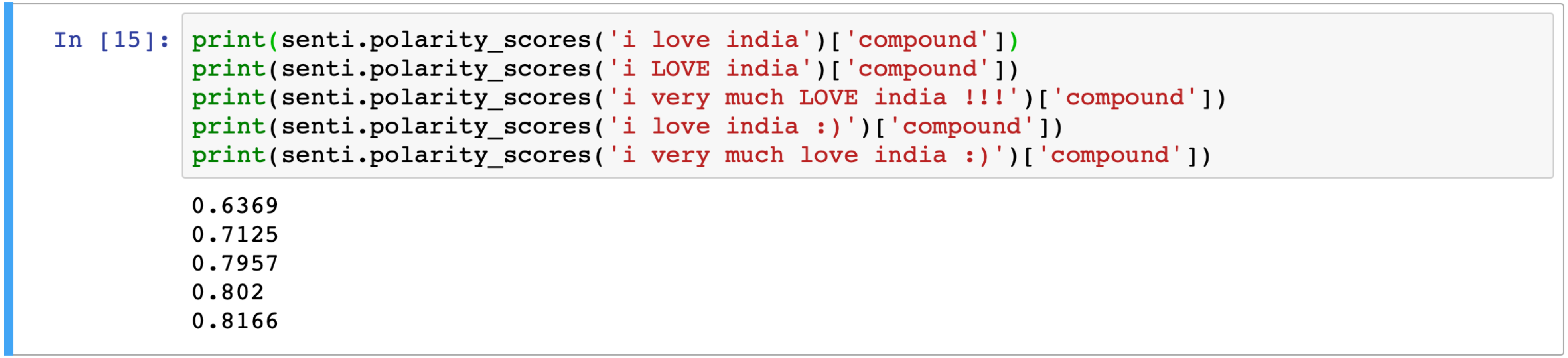
```
In [3]: import nltk
        import math
        from nltk.sentiment import SentimentIntensityAnalyzer

        senti = SentimentIntensityAnalyzer()
        senti.polarity_scores('love india')

Out[3]: {'neg': 0.0, 'neu': 0.192, 'pos': 0.808, 'compound': 0.6369}
```

VADER

- Gives importance to the following scenarios
  - Presence of positive and negative words
  - If positive or negative words are expressed in capital letters, intensity increases
  - Presence of emoticons
  - Presence of boosting or negate words one or two words before positive or negative words alters the intensity
  - Punctuations marks are consider for boosting the intensity



```
print(senti.polarity_scores('i love india')['compound'])
print(senti.polarity_scores('i LOVE india')['compound'])
print(senti.polarity_scores('i very much LOVE india !!!')['compound'])
print(senti.polarity_scores('i love india :)')['compound'])
print(senti.polarity_scores('i very much love india :)')['compound'])
```

```
0.6369
0.7125
0.7957
0.802
0.8166
```

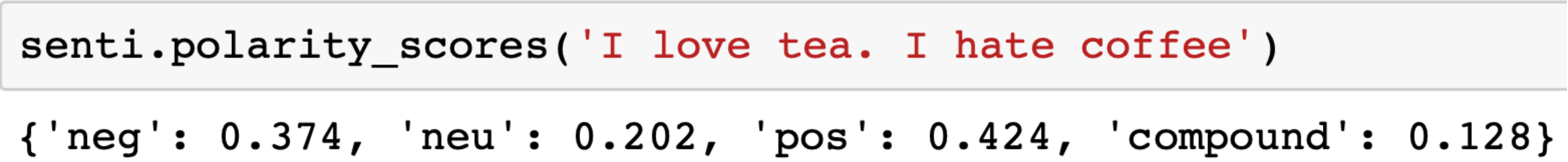
Percentage of Positive, Negative & Neutral

Problem Statement: Calculate the percentage of positive score, negative score and neutral score

Input Sentence: I love tea. I hate coffee

Word	Type of word	Score	Updated Score	Remarks
I	Ignored	0	0	Single letter words are ignored
Love	Positive	3.2	3.2+1=4.2	To compensate addition of 1 for neutral words, add one to positive words
Tea	Neutral	0	0+1=1	Neutral word will be given minimum score 1
I	Ignored	0	0	
Hate	Negative	-2.7	-2.7-1=abs(-3.7) = 3.7	Only absolute value are taken, since we are going to find percentage of negative score
Coffee	Neutral	0	0+1=1	
		Total	=4.2+1+3.7+1=9.9	

$\%\ \text{of}\ positive\ score = \frac{4.2}{9.9} = 0.424$ $\%\ \text{of}\ \ negative\ score = \frac{3.7}{9.9} = 0.374$ $\%\ \text{of}\ \ neutral\ score = \frac{2}{9.9} = 0.202$



```
senti.polarity_scores('I love tea. I hate coffee')
```

```
{'neg': 0.374, 'neu': 0.202, 'pos': 0.424, 'compound': 0.128}
```

Compound Score

Problem Statement: Calculate compound score for the following sentence

Input Sentence: I love tea. I hate coffee

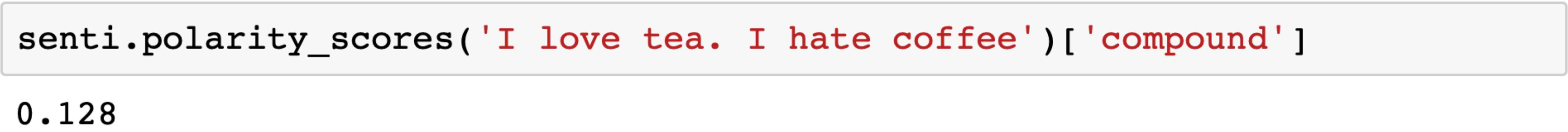
$$Total\ positive\ score = 3.2 \qquad Total\ negative\ score = -2.7$$

$$Overall\ score = Total\ positive\ score + Total\ negative\ score$$

$$Overall\ score = 3.2 + (-2.7) = 0.5$$

$$Compound\ score = \frac{score}{sqrt(score\ * score + alpha)}$$

$$Compound\ score = \frac{0.5}{sqrt(0.5\ * 0.5 + 15)} = 0.128$$



```
senti.polarity_scores('I love tea. I hate coffee')['compound']
```

```
0.128
```