

E08 - Sentiment Analysis (Python)

Çoban, Ömer Furkan

1. Setup and Data Loading

Loading the AFINN lexicon and stop words.

```
In [17]: import pandas as pd
import re
import os
import glob

# Paths
afinn_path = 'AFINN-111.txt'
stop_words_dir = 'stop-words-english'

# Load AFINN lexicon
afinn = {}
with open(afinn_path, 'r') as f:
    for line in f:
        parts = line.strip().split('\t')
        if len(parts) == 2:
            afinn[parts[0]] = int(parts[1])

print(f"Loaded {len(afinn)} terms from AFINN lexicon.")

# Load Stop Words
stop_words = set()
for sw_file in glob.glob(os.path.join(stop_words_dir, '*.txt')):
    with open(sw_file, 'r', encoding='utf-8', errors='ignore') as f:
        for line in f:
            word = line.strip().lower()
            if word:
                stop_words.add(word)

print(f"Loaded {len(stop_words)} stop words.")
```

Loaded 2477 terms from AFINN lexicon.

Loaded 855 stop words.

2. Task 1: Understanding Sentiment of an Input Text

Calculating sentiment scores for sentences based on the AFINN lexicon.

```
In [18]: def calculate_sentiment(text):
    # Simple tokenization: lowercase and alphanumeric words
    words = re.findall(r'\w+', text.lower())
    score = sum(afinn.get(word, 0) for word in words)

    if score > 0:
        sentiment = "Positive"
    elif score < 0:
        sentiment = "Negative"
    else:
        sentiment = "Neutral"

    return score, sentiment

# Sample Tests
test_sentences = [
    "I really like new book of that author.",
    "I hate new regulations about importing policies.",
    "Look at that door, it's still open."
]

results = []
for s in test_sentences:
    score, sent = calculate_sentiment(s)
    results.append({"Text": s, "Score": score, "Sentiment": sent})

df_task1 = pd.DataFrame(results)
df_task1
```

Out[18]:

	Text	Score	Sentiment
0	I really like new book of that author.	2	Positive
1	I hate new regulations about importing policies.	-3	Negative
2	Look at that door, it's still open.	0	Neutral

- "I really like new book of that author.": Contains the word 'like' which has a score of +2 in AFINN. Total score: +2 (Positive).
- "I hate new regulations about importing policies.": Contains the word 'hate' which has a score of -3 in AFINN. Total score: -3 (Negative).
- "Look at that door, it's still open.": Contains no words from the AFINN list. Total score: 0 (Neutral).

3. Task 2: Understanding Sentiment of New Terms

Estimating scores for terms not in AFINN based on the sentence sentiment.

```
In [19]: def estimate_new_terms(sentences):
    new_term_scores = {}
    term_counts = {}

    for s in sentences:
        sentence_score, _ = calculate_sentiment(s)
        words = re.findall(r'\w+', s.lower())

        for word in words:
            # Only consider words NOT in AFINN and NOT in stop words:
            if word not in afinn and word not in stop_words:
                new_term_scores[word] = new_term_scores.get(word, 0) + sentence_score
                term_counts[word] = term_counts.get(word, 0) + 1

    # Calculate average score for each new term
    final_scores = []
    for word, total_score in new_term_scores.items():
        avg_score = total_score / term_counts[word]
        final_scores.append({"Term": word, "Sentiment Score": avg_score})

    return pd.DataFrame(final_scores)

# Example with the test sentence
df_new_terms = estimate_new_terms(test_sentences)
df_new_terms.to_csv('sentiments_new.csv', index=False)
print("New term sentiments saved to sentiments_new.csv")
df_new_terms
```

New term sentiments saved to sentiments_new.csv

Out[19]:

	Term	Sentiment Score
0	book	2.0
1	author	2.0
2	regulations	-3.0
3	importing	-3.0
4	policies	-3.0
5	door	0.0
6	open	0.0

New terms are scored based on the total sentiment score of the sentences they appear in.

- 'book' and 'author': Appear in the first sentence (Score: +2). Average score: **2.0**.
- 'regulations', 'importing', 'policies': Appear in the second sentence (Score: -3). Average score: **-3.0**.
- 'door' and 'open': Appear in the third sentence (Score: 0). Average score: **0.0**.
- Note: Stop words like "new", "about", "of", "that" are excluded from this list as per the instructions.

4. Test: Analysis of Custom Sentences

Enter your own sentences below to analyze their sentiment.

```
In [20]: # Add your own sentences to this list
my_custom_sentences = [
    "This is a wonderful day!",
    "The weather is terrible and I feel sad.",
    "The movie was okay, nothing special."
]

test_results = []
for s in my_custom_sentences:
    score, sent = calculate_sentiment(s)
    test_results.append({"Text": s, "Calculated Score": score, "Sentiment result": sent})

df_test = pd.DataFrame(test_results)
df_test
```

Out[20]:

	Text	Calculated Score	Sentiment result
0	This is a wonderful day!	4	Positive
1	The weather is terrible and I feel sad.	-5	Negative
2	The movie was okay, nothing special.	0	Neutral

In [21]: # Add your own sentences to this list

```
sentences = [
    "The Trump administration has insisted it has numerous military options to deploy against Iran if the regime uses force again",
    "But that menu is far more limited than it was even a year ago.",
    "The U.S. troops and ships that were once at the president's disposal have shifted to the Caribbean.",
    "A major American defense system sent to the Middle East last year has returned to South Korea.",
    "And administration officials say there are no plans for the movement of major assets.",
    "The president can still order airstrikes that target Iranian leadership or military installations.",
    "But his choices are even more reduced than June, when the U.S. took out Iran's nuclear sites.",
    "And he also must contend with lawmakers who, just over a week after Trump ordered the capture of Venezuelan leader Nicolás Ma",
    "What's the objective?",
    "How does military force get you to that objective?" asked Rhode Island Sen. Jack Reed, the top Democrat on the Armed Service
```

```

    "They're certainly repressing their people, but the president has yet to make the case that a military strike will either aid
    "The Trump administration also has been eating away at dwindling U.S. weapons stockpiles with the fast pace of military operati
    "The bottleneck has become particularly stark for air defense that protects U.S. forces within range of Iran's weapons.",
    "If the administration strikes and Iranians retaliate forcefully, the U.S. may have a limited stockpile of interceptors to def
    "The Pentagon stations 10,000 U.S. troops at Al-Udeid Air Base in Qatar and smaller groupings in Iraq, Syria and Jordan.",
    "If it does become a longer-term volley of strikes, then your interceptor capacity becomes all the more important," said a fo
    "We could get in a sticky situation very quickly on that front."",
    "The White House insisted the president had plenty of choices.",
    "President Trump has a full menu of options at his disposal with regard to Iran," said spokesperson Anna Kelly.",
    "Senior officials met Tuesday to talk about the U.S. response, White House press secretary Karoline Leavitt told reporters, bu
    "Iranian protests against staggering inflation and government policies began in December and have spread across the country.",
    "The regime's security forces have cracked down against demonstrators, killing as many as 2,000 people, according to rights gr
    "As Tehran's response has escalated, so has Trump's.",
    "In a Truth Social post Tuesday, Trump said that Iran's "killers and abusers" within the regime "will pay a big price."",
    "I have cancelled all meetings with Iranian Officials until the senseless killing of protesters STOPS," Trump said.,
    "HELP IS ON ITS WAY.",
    "But an administration official told POLITICO on Monday that no major moves of U.S. troops or assets were in the works.",
    "The USS Ford, which was rerouted from the Middle East last year, remains in the Caribbean in the aftermath of the Venezuela c
    "The USS Vinson and USS Nimitz, the two U.S. aircraft carriers that Trump ordered to the Middle East in June, departed the reg
]

```

```

test_results = []
for s in sentences:
    score, sent = calculate_sentiment(s)
    test_results.append({"Text": s, "Calculated Score": score, "Sentiment result": sent})

df_test = pd.DataFrame(test_results)
df_test

```

Out[21]:

	Text	Calculated Score	Sentiment result
0	The Trump administration has insisted it has n...	0	Neutral
1	But that menu is far more limited than it was ...	-1	Negative
2	The U.S. troops and ships that were once at th...	0	Neutral
3	A major American defense system sent to the Mi...	0	Neutral
4	And administration officials say there are no ...	1	Positive
5	The president can still order airstrikes that ...	0	Neutral
6	But his choices are even more reduced than Jun...	0	Neutral
7	And he also must contend with lawmakers who, j...	-5	Negative
8	"What's the objective?	0	Neutral
9	"How does military force get you to that objec...	2	Positive
10	"They're certainly repressing their people, bu...	-1	Negative
11	The Trump administration also has been eating ...	0	Neutral
12	The bottleneck has become particularly stark f...	1	Positive
13	If the administration strikes and Iranians ret...	-2	Negative
14	The Pentagon stations 10,000 U.S. troops at Al...	0	Neutral
15	"If it does become a longer-term volley of str...	5	Positive
16	"We could get in a sticky situation very quick...	0	Neutral
17	The White House insisted the president had ple...	0	Neutral
18	"President Trump has a full menu of options at...	0	Neutral
19	Senior officials met Tuesday to talk about the...	0	Neutral
20	Iranian protests against staggering inflation ...	-2	Negative
21	The regime's security forces have cracked down...	-3	Negative
22	As Tehran's response has escalated, so has Tru...	0	Neutral
23	In a Truth Social post Tuesday, Trump said tha...	0	Neutral
24	"I have cancelled all meetings with Iranian Of...	-7	Negative
25	"HELP IS ON ITS WAY."	2	Positive
26	But an administration official told POLITICO o...	1	Positive
27	The USS Ford, which was rerouted from the Midd...	0	Neutral
28	The USS Vinson and USS Nimitz, the two U.S. ai...	0	Neutral