איחזור מידע 3

README: Sentiment Analysis of News Articles

Project Contributors

- **Name:** Rachelli Adler **ID**: 213836687
- **Name:** Shifra Scheiner **ID:** 328949714
- **Course number**: 157125.5785

Project Overview

This project aims to determine the sentiment orientation of articles published by various newspapers towards Israel and Palestine. The classification identifies whether the content of a news article is pro-Israeli, pro-Palestinian, or neutral

---

2. Workflow

Step 1: Sentence Extraction

Articles were processed to identify sentences containing only pro-Israel or pro-Palestinian keywords. Each sentence was tagged with its sentiment orientation (I for Israel, P for Palestine). We took only the sentences that had one categories words.

Step 2: Sentiment Classification

Each extracted sentence was analyzed using multiple sentiment classifiers available on Hugging Face.

Sentiment Models Used:

1. Cardiff NLP - Twitter RoBERTa Base Sentiment

2. NLPTown - BERT Base Multilingual Uncased Sentiment
3. Cardiff NLP - Twitter XLM RoBERTa Base Sentiment
4. Siebert - Sentiment RoBERTa Large English
5. LXYuan - DistilBERT Multilingual Cased Sentiments
6. FiniteAutomata - BERTweet Sentiment
7. J-Hartmann - Sentiment RoBERTa Large 3-Classes

Step 4: Consensus Mechanism

- We used multiple models to classify each sentence. Each model's script is saved as a separate file (e.g., modelX.py).
- The final sentiment score for each sentence was calculated as the average of the scores from all models.

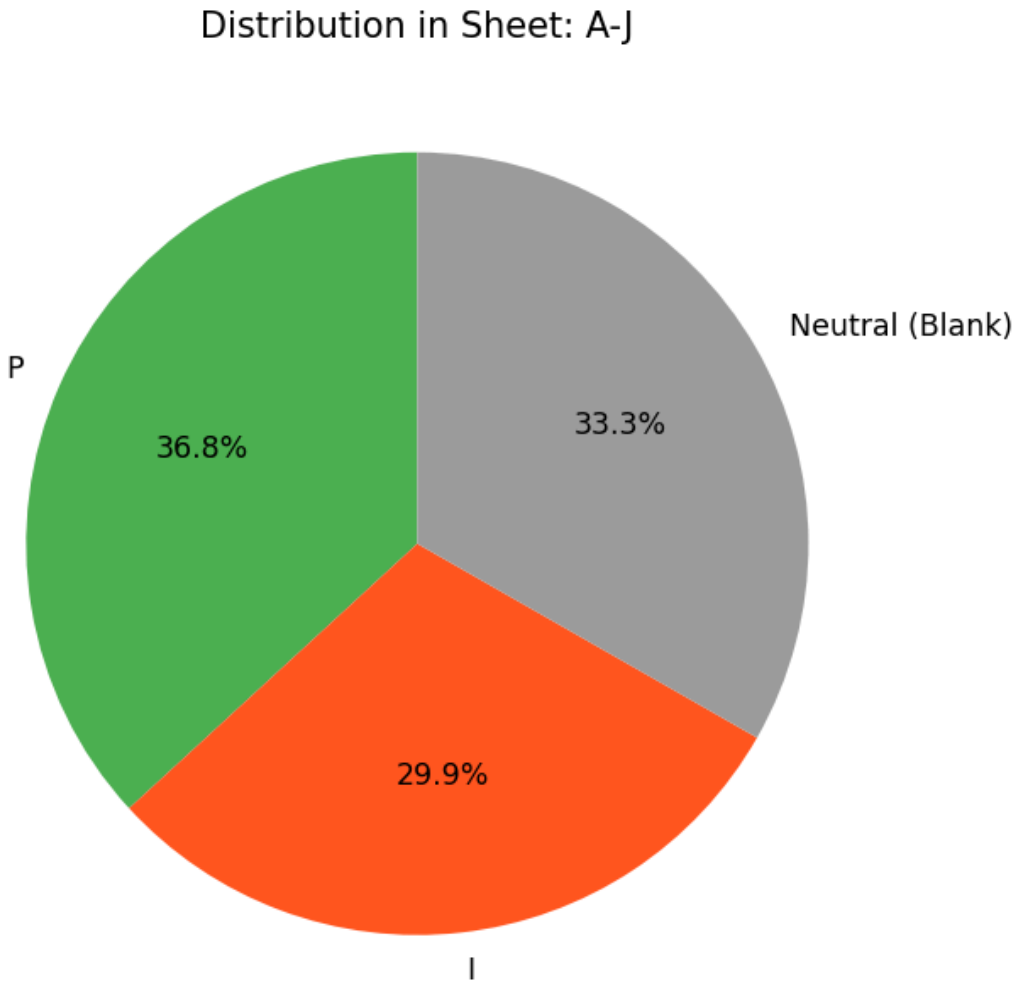
Step 5: Article Sentiment Aggregation

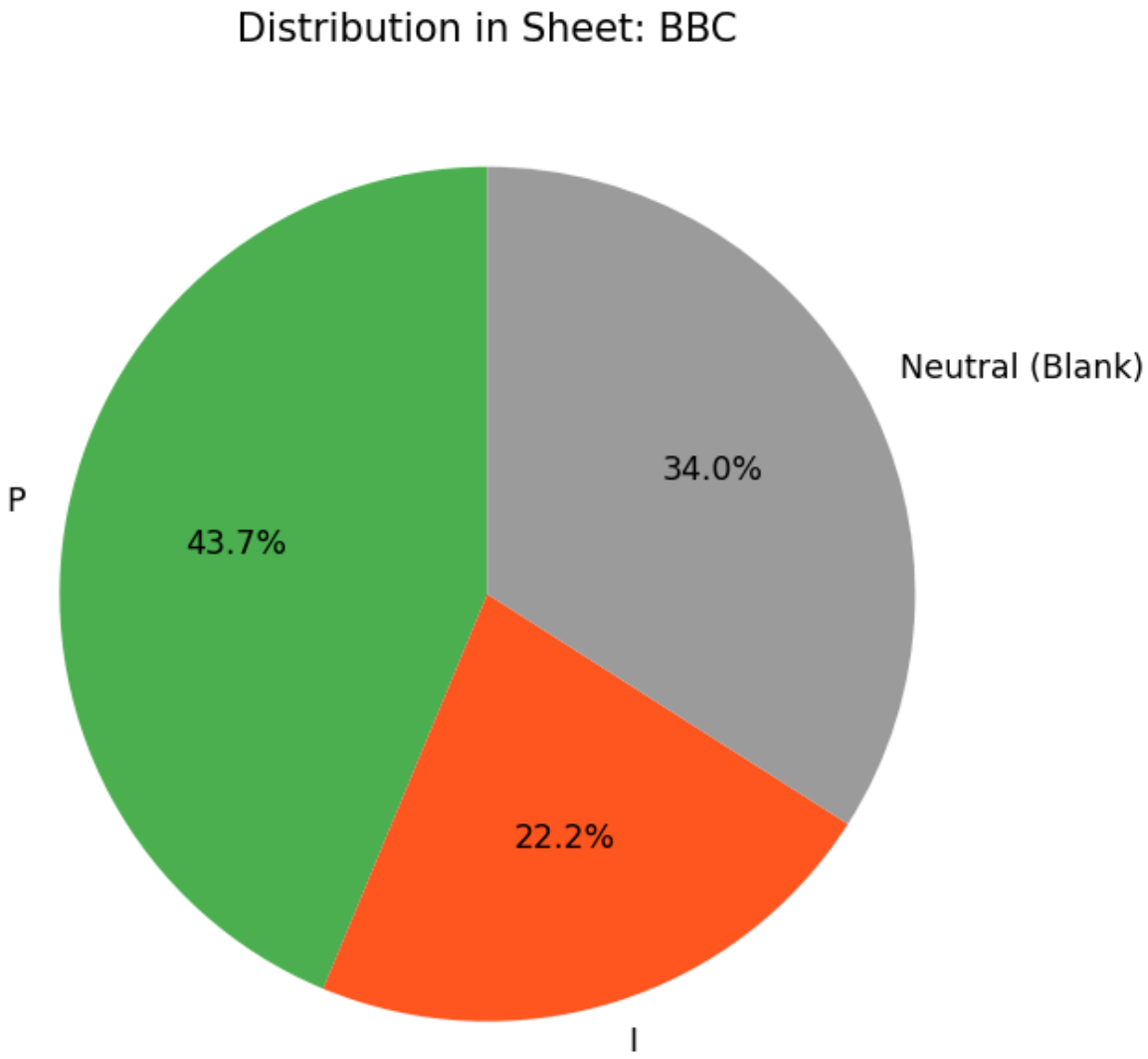
- Based on the overall score, each article was labeled as pro-Israeli, pro-Palestinian, or neutral.

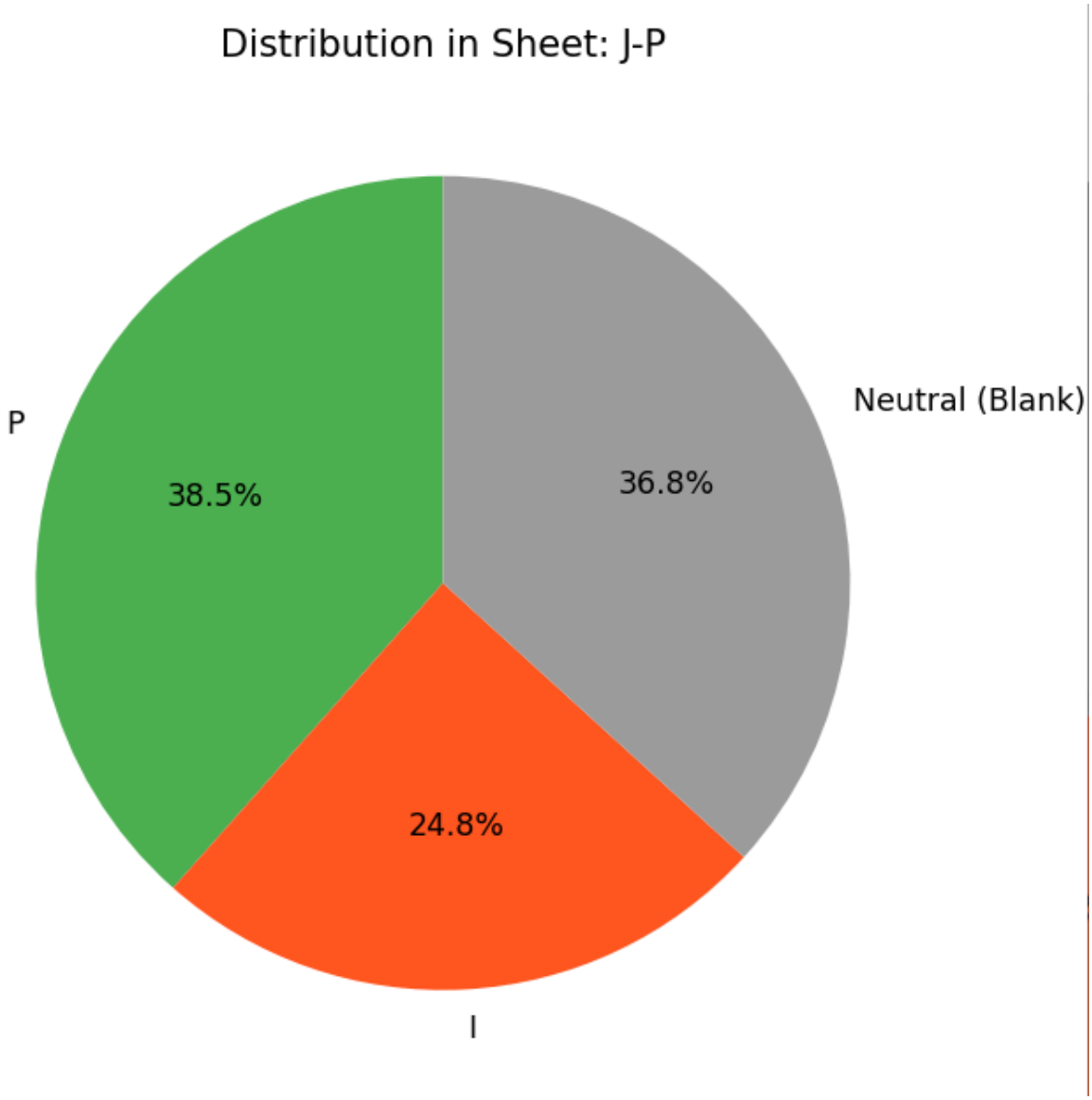
Step 6: Newspaper Sentiment Analysis

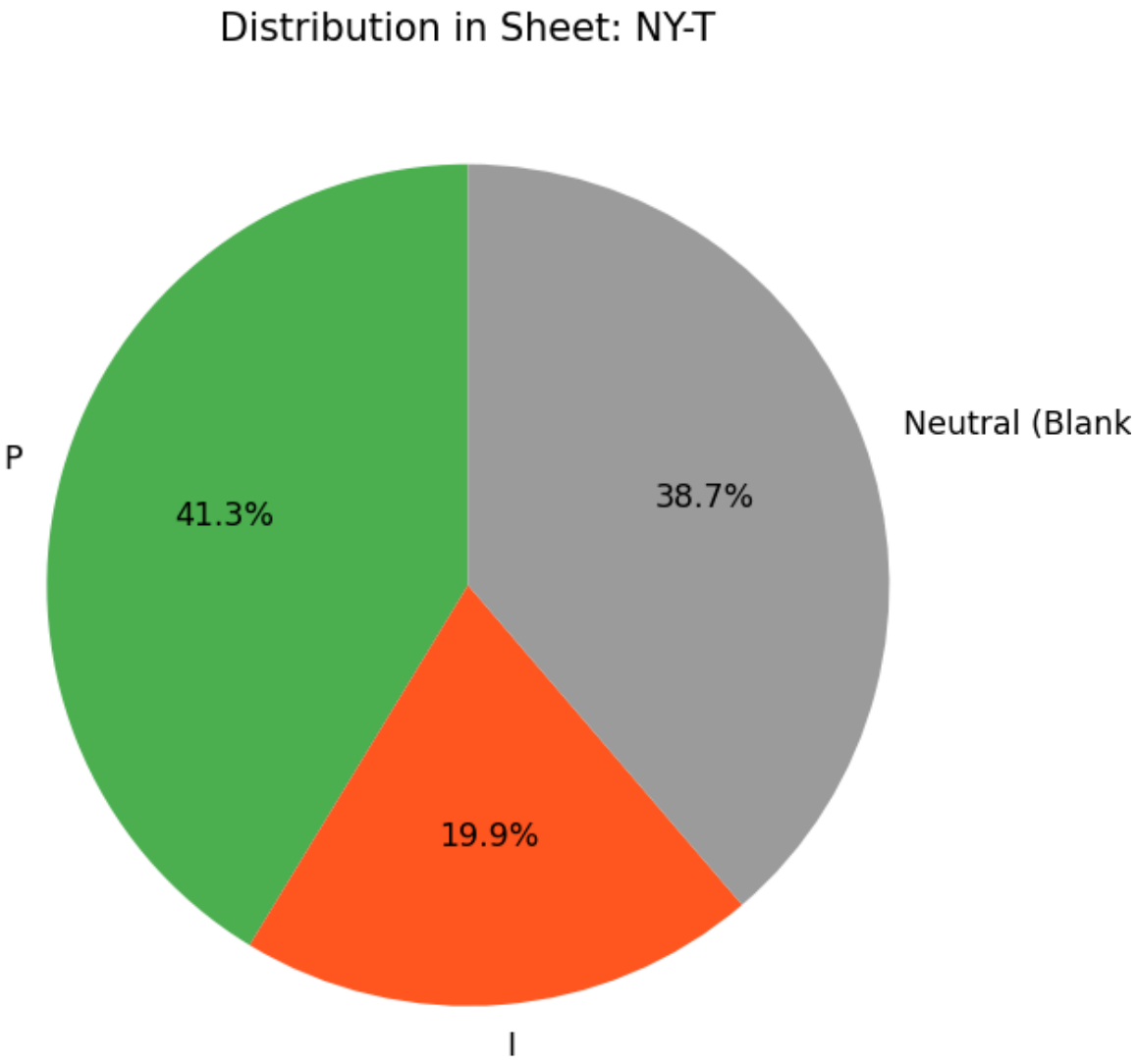
- We summarized the sentiment labels of all articles in each newspaper.
- Each newspaper received a final sentiment score based on the articles it published.

Results


Distribution in Sheet: A-J
P
36.8%
Neutral (Blank)
33.3%
29.9%
I


Distribution in Sheet: BBC
Neutral (Blank)
34.0%
P
43.7%
22.2%
I


Distribution in Sheet: J-P
P
38.5%
Neutral (Blank)
36.8%
24.8%
I



C:\Users\rache\AppData\Local\Programs\Python\Python312\python.exe "C:\Users\rache\Desktop\איחזור מידע\hw3\statistics.py"

Statistics for sheet: A-J

	I/P	majority sentiment	count
0	I	NEG	142
1	I	NEU	47
2	I	POS	4
3	P	NEG	122

```
4   P              NEU     90
5   P              POS     13

Statistics for sheet: BBC
  I/P majority sentiment  count
0   I              NEG   2809
1   I              NEU   1600
2   I              POS    238
3   P              NEG   1230
4   P              NEU    645
5   P              POS     79

Statistics for sheet: J-P
  I/P majority sentiment  count
0   I              NEG   1634
1   I              NEU   1369
2   I              POS    545
3   P              NEG    553
4   P              NEU    257
5   P              POS     71

Statistics for sheet: NY-T
  I/P majority sentiment  count
0   I              NEG    622
1   I              NEU    447
2   I              POS     48
3   P              NEG    259
4   P              NEU    146
5   P              POS     14

Process finished with exit code 0
```

Example Output (Excel)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Newspape	ument Nur	tence Nun	Sentence	I/P	del1 senti	model1 sco	del2 senti	model2 sco	del3 senti	model3 sco	del4 senti	model4 sco	el 5 senti	odel 5 sco	el 6 senti	odel 6 sco	el 7 senti	odel 7 sco	rity senti	erage sco
2	al-j	1	1	pope rene	P	NEU	0.79149	NEG	0.46812	NEG	0.72418	POS	0.99821	NEG	0.4727	NEU	0.92095	NEU	0.99832	NEG	0.76771
3	al-j	2	1	biden is st	I	POS	0.9688	POS	0.7156	POS	0.57747	POS	0.99867	POS	0.81125	POS	0.98781	POS	0.99927	POS	0.86555
4	al-j	2	2	united sta	I	NEU	0.81167	POS	0.31472	NEG	0.73338	POS	0.98888	NEG	0.48503	NEU	0.84571	NEU	0.99592	NEU	0.73933
5	al-j	7	1	relatives c	P	NEG	0.67095	NEG	0.36642	NEG	0.77564	NEG	0.99554	NEG	0.69222	NEG	0.6946	NEU	0.93724	NEG	0.73323
6	al-j	8	1	icj rules isi	I	NEG	0.79872	NEG	0.51167	NEG	0.78151	NEG	0.99824	NEG	0.49476	NEU	0.52884	NEU	0.99916	NEG	0.73042
7	al-j	10	1	uncoverin	P	NEU	0.82049	POS	0.39195	NEU	0.50992	POS	0.99501	NEG	0.57926	NEU	0.79	NEU	0.99659	NEU	0.72617
8	al-j	10	2	the discov	P	NEG	0.83326	POS	0.50728	NEU	0.88371	NEG	0.99508	NEG	0.89267	NEG	0.87207	NEG	0.9898	NEG	0.85341
9	al-j	11	1	uk will not	I	NEU	0.90519	NEG	0.53778	NEG	0.46302	POS	0.98624	NEU	0.42121	NEU	0.90342	NEU	0.9988	NEU	0.7451
10	al-j	11	2	the united	I	NEU	0.94021	NEG	0.52187	NEG	0.48187	POS	0.98782	NEG	0.36897	NEU	0.94328	NEU	0.99863	NEG	0.74895
11	al-j	13	1	pakistani ;	P	NEU	0.50868	NEG	0.52848	NEG	0.82918	NEG	0.99897	NEG	0.51461	NEU	0.57166	NEG	0.95625	NEG	0.70112
12	al-j	14	1	how israel	I	NEU	0.61026	NEG	0.30052	NEG	0.61381	NEG	0.99298	NEG	0.62944	NEG	0.58755	NEU	0.99777	NEG	0.67605
13	al-j	16	1	gaza cease	P	NEG	0.66426	NEG	0.6025	NEG	0.74616	NEG	0.99947	NEU	0.54715	NEG	0.90856	NEG	0.90859	NEG	0.7681
14	al-j	17	1	timeline: t	P	NEU	0.92949	POS	0.25776	NEU	0.58736	POS	0.9923	POS	0.45157	NEU	0.97091	NEU	0.99913	NEU	0.74122
15	al-j	18	1	what happ	I	NEU	0.61054	NEG	0.45144	NEG	0.71107	NEG	0.99533	NEG	0.6813	NEG	0.51307	NEU	0.99917	NEG	0.70885

File Structure:

- Excel file with four sheets (one per newspaper).
- Each sheet includes:
    - Article and sentence number
    - text.
    - Sentiment orientation (I/P).
    - Sentiment scores and labels for each model.
    - Final sentiment classification.

Sentiment Summary for Newspapers

For each newspaper:

- Pie charts showing the percentage of articles classified as pro-Israeli, pro-Palestinian, or neutral.
- Bar charts summarizing sentiment scores across all articles.

Challenges and Insights

1. **Difficulty in Identifying Sentiment:**
    - Sentences with ambiguous wording or balanced mentions of both sides posed challenges.
2. **Limitations of Sentiment Models:**
    - Differences between models highlighted the importance of the consensus approach.
3. **Behavioral Patterns of Newspapers:**
    - Because the newspapers are reporting on the war, the language they use is often negative and biased.
