

A Machine Learning Approach for Detecting Propaganda in News Articles

**WITH CASE STUDY ON ISRAEL-PALESTINE
COVERAGE**

Reuben Lesser

MSc in Artificial Intelligence and Ethics

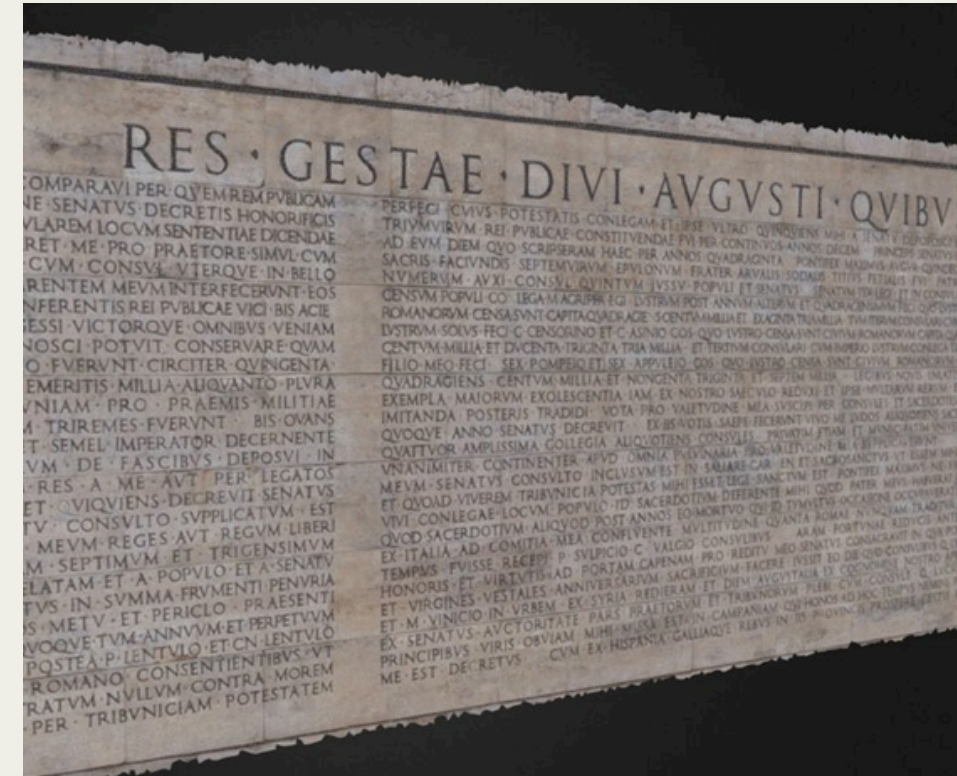
Dissertation viva

6 January, 2025



Background

- Defining propaganda and false information
 - Misinformation
 - Disinformation
 - Fake News
 - Propaganda
- History of Propaganda
- Automated propaganda detection
- Propaganda in Israel-Palestine news coverage

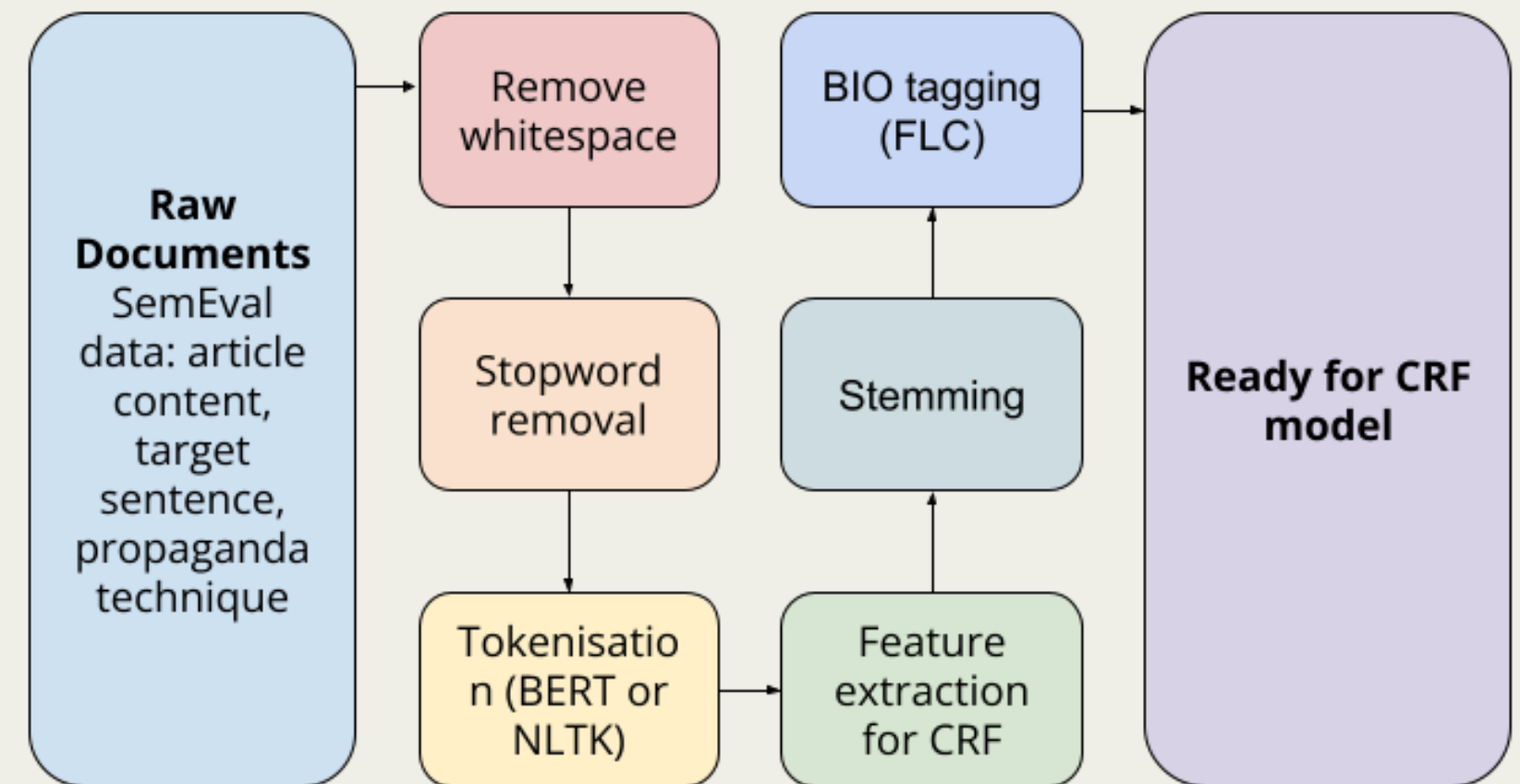


1949 Armistice lines



Methodology

- Data preprocessing and analysis
- Model architecture
 - CNN
 - CRF
- Training methodology
- Evaluation framework



Model Performance

- Initial model comparison
 - CNN vs CRF
- CRF performance
 - Sentence-level classification
 - Fragment-level classification

Accuracy	Predictions/ Classifications	$\frac{\text{Correct}}{\text{Correct} + \text{Incorrect}}$
Precision	Predictions/ Classifications	$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$
Recall	Predictions/ Classifications	$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$
F1	Predictions/ Classifications	$\frac{2 * \text{True Positive}}{\text{True Positive} + 0.5 (\text{False Positive} + \text{False Negative})}$

Model Performance Results

Metric	Pre-kfold macro	Post-kfold macro	Pre-kfold weighted	Post-kfold weighted
Accuracy	0.85	0.98	0.85	0.98
Precision	0.86	0.98	0.85	0.98
Recall	0.84	0.98	0.85	0.98
F1	0.84	0.98	0.85	0.98

Table 1: Sentence classification before and after K-fold

Class	Precision	Recall	F1-score	Support
Appeal to Authority	0.94	0.62	0.75	4,696
Bandwagon, Reductio ad hitlerum	0.91	0.71	0.80	5,622
Causal_Oversimplification	0.92	0.66	0.77	6,118
Doubt	0.81	0.69	0.75	5,520
Exaggeration, Minimisation	0.77	0.51	0.61	2,439
Loaded Language	0.59	0.41	0.49	1,076
Name Calling, Labeling	0.75	0.35	0.48	1,467
O	0.85	0.98	0.91	95,099
Slogans	0.96	0.53	0.69	1,529
Flag-Waving	0.80	0.52	0.63	3,360
Whataboutism, straw man, red herring	0.99	0.58	0.73	5,657
Black-and-White Fallacy	0.97	0.58	0.73	4,867
Prejudice	0.80	0.77	0.78	5,166
Thought-terminating Cliches	0.99	0.64	0.78	2,324
accuracy			0.86	144,940
macro avg	0.86	0.61	0.71	144,940
weighted avg	0.86	0.86	0.85	144,940

Table 3: Classification Report for Fragment-level Test Set

Israel-Palestine News Analysis

- Web scraping methodology
- Data cleaning and preprocessing
- LLM annotation
- Results of deployment on Israel-Palestine news articles

```
topics = {  
    'world/gaza': 'Gaza',  
    'world/israel': 'Israel',  
    'world/palestinian-territories': 'Palestine',  
    'world/middleeast': 'Middle East',  
    'world/west-bank': 'West Bank',  
    'world/hamas': 'Hamas',  
    'world/hezbollah': 'Hezbollah'  
}
```

```
topics = {  
    'https://www.bbc.co.uk/news/topics/c302m85q5ljt': 'Gaza',  
    'https://www.bbc.co.uk/news/topics/c8nq32jw5r5t': 'Israel',  
    'https://www.bbc.co.uk/news/topics/c207p54m4rqt': 'Palestine',  
    'https://www.bbc.co.uk/news/topics/cjnw18q4ggnt': 'Middle East',  
    'https://www.bbc.co.uk/news/topics/cxvmk3z8mn9t': 'West Bank',  
    'https://www.bbc.co.uk/news/topics/cnx753jen5zt': 'Hamas',  
    'https://www.bbc.co.uk/news/topics/c8nq32jwj2lt': 'Hezbollah'  
}
```

Israel-Palestine News Analysis

Results

Topic	BBC	Guardian	Average
Middle East	NaN	30.60	30.60
Hamas	26.67	29.99	28.33
Palestine	27.91	28.34	28.13
Gaza	26.85	29.21	28.03
West Bank	25.51	26.99	26.25
Israel	27.39	25.02	26.20
Hezbollah	26.04	22.96	24.50

Table 4: Coverage Analysis of Middle East Topics in BBC and Guardian

Technique	BBC	Guardian	Average
Loaded Language	66.60	71.96	69.28
Doubt	15.87	10.28	13.08
Exaggeration, Minimisation	8.45	5.61	7.03
Name Calling, Labeling	2.98	4.21	3.60
Appeal to fear/prejudice	2.49	4.21	3.35
Flag-waving	1.66	1.40	1.53
Appeal to Authority	0.14	1.40	0.77
Causal Oversimplification	0.49	0.93	0.71
Black-and-White Fallacy	0.55	0.00	0.28
Bandwagon, reductio ad hitlerum	0.28	0.00	0.14
Whataboutism, straw man, red herring	0.28	0.00	0.14
Thought-terminating cliches	0.21	0.00	0.11

Table 6: Propaganda Techniques Analysis in BBC and Guardian Coverage

Irrelevant	Relevant
24.86%	26.92%



Conclusion and Future Work

- Discussion of final CRF model for propaganda detection
- Drawbacks and limitations
- Ethical considerations
- Future research directions



Q&A