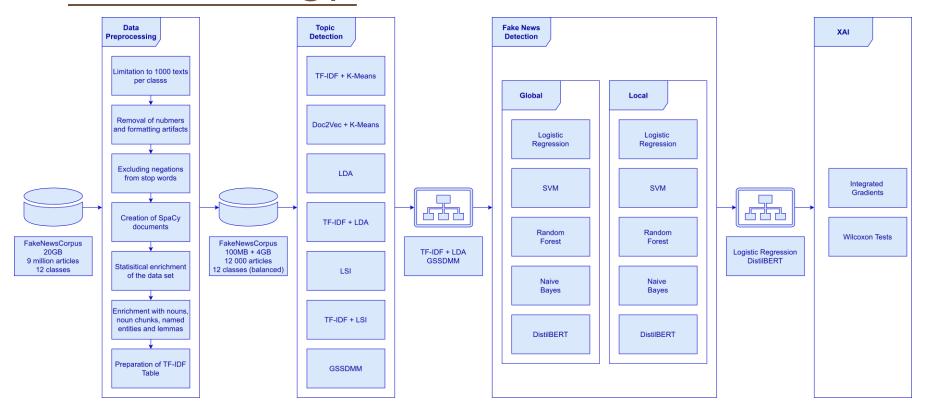
The Comparison of Local and Global Early Fake News Detection Methods

Maciej Pawlikowski, Hubert Ruczyński, Bartosz Siński, Adrian Stańdo

Main goals

- 1. Comparison of different topic detection models.
- 2. Comparison of fake news detection methods.
- 3. Introduction of **local** fake news detection methods.
- 4. Evaluation of the local approach, and comparison to corresponding global solutions.
- 5. Exploration of models **differences** between the two strategies with the usage of **XAI**.

Methodology



Topic detection

Clustoring Algorithm	Lo	emmas	Noun chunks		
Clustering Algorithm	Silhouette Score (1)	ore (↑) Calinski-Harabasz Score (↑) Silhouette Score (↑) Calinski-Haraba		Calinski-Harabasz Score (↑)	
TF-IDF + K-Means	0.038	293.9	0.067	314.1	
Doc2Vec + K-Means	0.134	0.134 449.6 0.386		3473.24	
LDA	0.607	18668.8	0.883	110432.1	
TF-IDF + LDA	0.873	91490.7	0.929	323993.32	
LSI	-0.320	49.5	-0.512	132.0	
TF-IDF + LSI	0.469	1655.9	-0.290	393.5	
GSSDMM	0.714	529.4	0.867	15681.4	

Fake News Detection

Model	No clustering			
Model	Accuracy (1)	Accuracy w/o politics (↑)		
Logistic Regression	0.581	0.628		
SVM	0.518	0.577		
Random Forest	0.491	0.540		
Naive Bayes	0.444	0.501		
DistilBERT	0.710	0.790		

Global and local models

Model	No clustering (Global Model)	GSSDMM (Local Models)	LDA (Local Models)
Logistic Regression	LR	GSSDMM + LR	LDA + LR
SVM	SVM	GSSDMM + SVM	LDA + SVM
Random Forest	RF	GSSDMM + RF	LDA + RF
Naive Bayes	NB	GSSDMM + NB	LDA + NB
DistilBERT	D-BERT	GSSDMM+D-BERT	LDA + D-BERT

Global vs local - performance

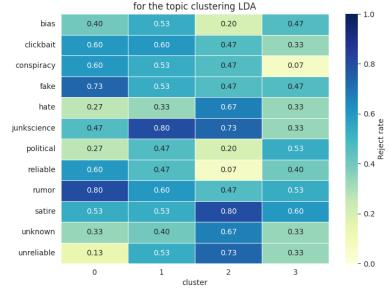
Model	N	o clustering	GSSDMM	LDA	
	Accuracy (↑)	Accuracy w/o politics (↑)	Weighted Accuracy (↑)	Weighted Accuracy (↑)	
Logistic Regression	0.581	0.628	0.540	0.530	
SVM	0.518	0.577	0.460	0.480	
Random Forest	0.491	0.540	0.470	0.460	
Naive Bayes	nive Bayes 0.444		0.440	0.450	
DistilBERT	T 0.710 0.790		0.600	0.620	

Global vs local – XAI (IG)

Heatmap of reject rate ($\alpha = 0.05$) of the hypothesis of the equality of explanations for the topic clustering GSS

Tor the topic clustering GSS							_	- 1.0	
bias	0.53	0.93	0.67	0.47	0.33	0.53	0.40		1.0
clickbait	0.73	0.67	0.60	0.67	0.87	0.40	0.47		
conspiracy	0.60	0.60	0.47	0.47	0.40	0.67	0.33		- 0.8
fake	0.60	0.80	0.53	0.33	0.33	0.27	0.53		
hate	0.40	0.60	0.47	0.73	0.67	0.47	0.27		- 0.6
junkscience	0.60	0.33	0.33	0.73	0.60	0.53	0.73		rate
political	0.53	0.60	0.67	0.60	0.73	0.27	0.40		Reject rate
reliable	0.73	0.33	0.53	0.47	0.53	0.60	0.40		- 0.4
rumor	0.67	0.53	0.47	0.47	0.53	0.53	1.00		
satire	0.53	0.33	0.67	0.33	0.53	0.60	0.80		- 0.2
unknown	0.60	0.47	0.67	0.40	0.80	0.67	0.53		
unreliable	0.40	0.53	0.47	0.60	0.60	0.67	0.27		
	0	1	2	3 cluster	4	5	6		- 0.0

Heatmap of reject rate ($\alpha = 0.05$) of the hypothesis of the equality of explanations



Limitations

- 1. Computational resources only private PCs.
- 2. Consideration of only 12,000 observations (out of 9,000,000 accessible!).
- 3. Consideration of 12 classess in such a small dataset.
- 4. Additional split into 4 and 7 clusters (results in ~250 and ~150 per class in clusters).
- 5. Employment of relatively simple models, only one transformer.

What could we improve?

- 1. Task simplification to 2 or 4 classes instead of 12.
- 2. Double the amount of observations to 24,000.
- 3. For local approaches with 4 and 7 clusters it results in ~3000/1500 and ~1500/750. osbervations per class, which is 12/6 times more than before.
- 4. Consider more transformer models, such as BERTopic.

Further works with more resources

- 1. Expansion to even greater numer of articles (100,000 observations).
- 2. Consideration of more/all classes.
- 3. Consideration of more topic detection methods (not only clustering models).
- 4. Testing even more advanced models as Fake News detectors (transformers, LLMs).
- 5. Deployment of the framework (best combination) as an online learning system.
- Consideration of online learning clustering methods, e.g. TextClust.

