Université de Montréal

University of Montreal

Fake News

Dorsaf SALLAMI
Supervised by professor Esma Aïmeur

Plan

1

Context

- What is fake news?
- Fake news propagation
- Fake news related terms

2

EXMULF

- Related work
- The proposed approach
- Experiments and results

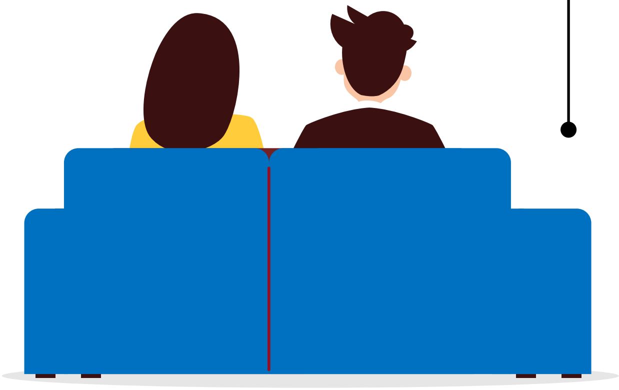
3

Current research

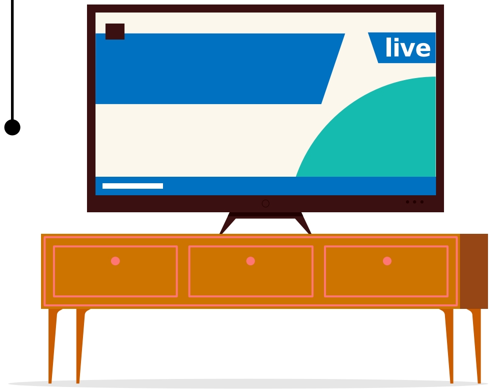
- Recommender System and fake news
- Fake news spreader personality

Fake news

An issue without a clear or universally accepted definition



From an age-old problem to a contemporary problem



Fake news Propagation

1

Circulation Process

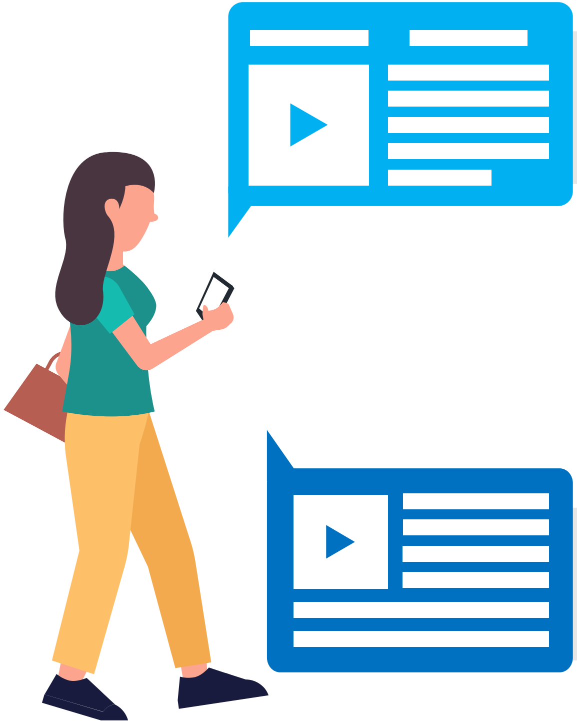
2

Social Networks

3

Platforms

Fake news Detection: Related works



Multimodal Data

- Correlation Semantic analysis
- Sentiment analysis
- Web scraping ….

1

Explainable Fake News Detection

- SHAP.
- Tsetlin Machine (TM).
- MIMIC, ATTN, PERT…

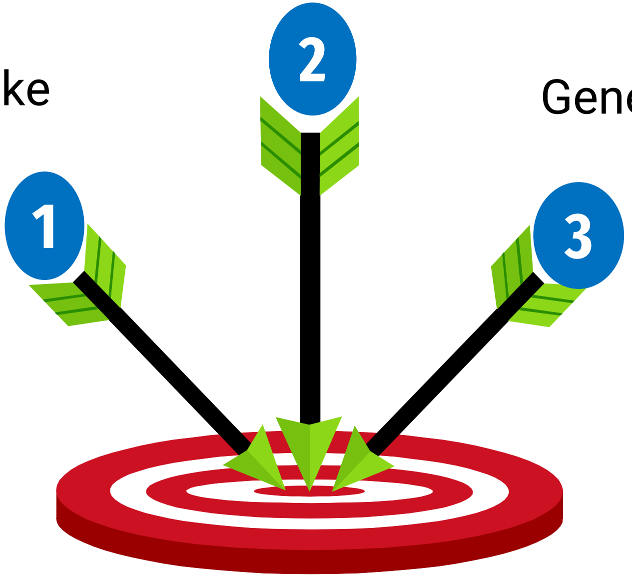
2

Main Contributions

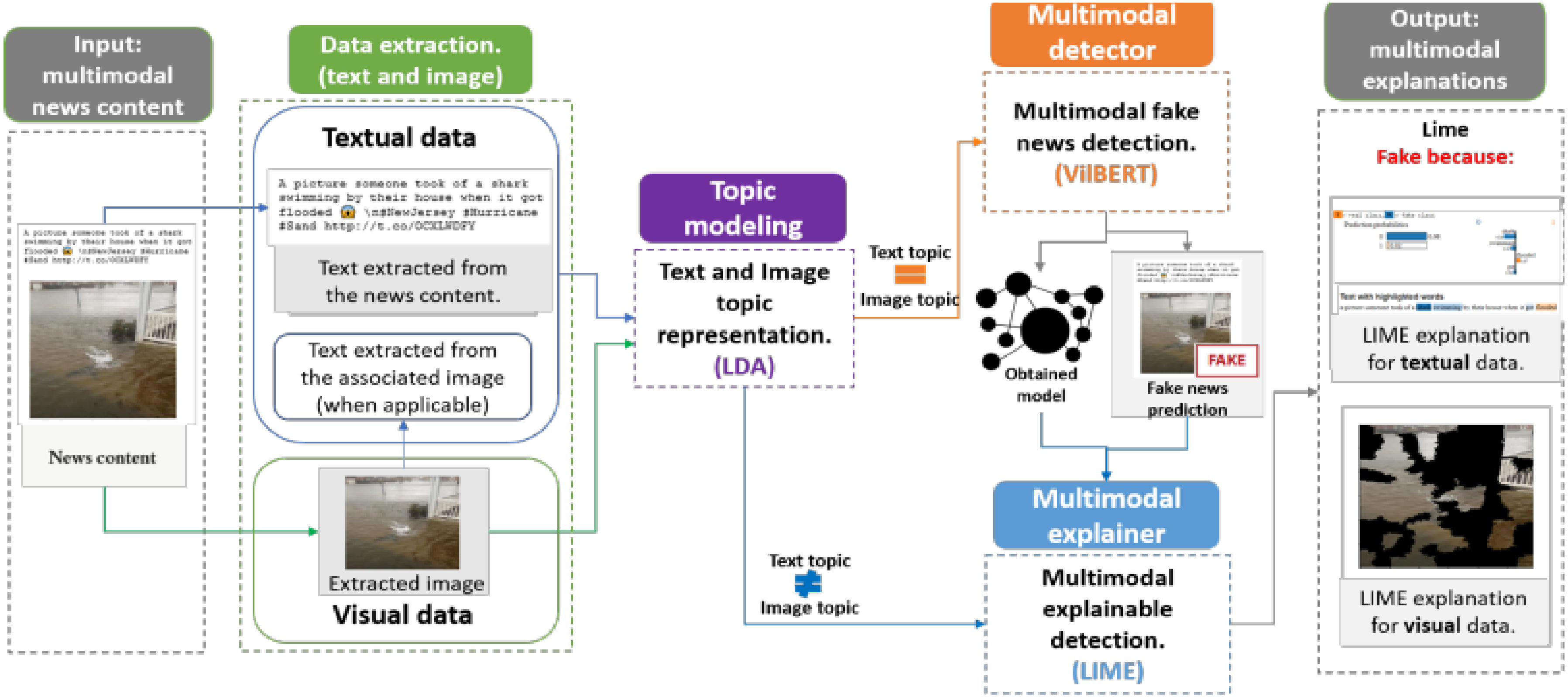
Multimodal topic modeling analysis to measure the topic similarity.

Multimodal data to detect fake (ViLBERT).

Generate appropriate multimodal explanations (LIME).



EXMULF : An Explainable Multimodal Content-based Fake News Detection System



EXMULF methodology overview

EXMULF

LIME

- Accessibility and simplicity.
- Model agnosticism.
- Local explanations.
- Interpretable.

VilBERT

EXMULF

- Model for learning task-agnostic.
- Image text alignment prediction.

Topic Modeling

- Incoherence between text and image

Experiments and Results (1/5)

DATASETS

Twitter
Weibo

Data preprocessing

- Removal of single modality instances
- Preprocessing of textual data:
- Preprocessing of images:

Experiments and Results (2/5)

Dataset	Model		Accuracy	Fake News Precision	Fake News Recall	Fake News F1	Real News Precision	Real News Recall	Real News F1
Twitter	Text only	$BERT_T$	0.572	0.602	0.586	0.597	0.543	0.553	0.544
		$BERT_{T+IT}$	0.577	0.612	0.574	0.598	0.551	0.564	0.556
	Image only	ResNet-34	0.624	0.712	0.567	0.6	0.558	0.72	0.62
		VGG-19	0.596	0.698	0.522	0.593	0.531	0.698	0.597
	Multi-modal	Fusion	0.7695	0.820	0.726	0.779	0.719	0.798	0.748
		SpotFake [22]	0.7777	0.751	0.900	0.82	0.832	0.606	0.701
		AMFB [8]	0.883	0.89	**0.95**	0.92	**0.87**	0.76	0.741
		HMCAN [15]	0.897	**0.971**	0.801	0.878	0.853	**0.979**	**0.912**
		BDANN [30]	0.830	0.810	0.630	0.710	0.830	0.930	0.880
		VilBERT	**0.898**	0.934	0.92	**0.926**	0.859	0.88	0.869
Weibo	Text only	$BERT_T$	0.680	0.731	0.715	0.709	0.667	0.676	0.669
		$BERT_{T+IT}$	0.682	0.739	0.72	0.71	0.672	0.684	0.673
	Image only	ResNet-34	0.694	0.701	0.634	0.698	0.698	0.711	0.699
		VGG-19	0.633	0.640	0.635	0.637	0.637	0.641	0.639
	Multi-modal	Fusion	0.8152	0.865	0.734	0.88	0.764	0.889	0.74
		SpotFake [22]	0.8923	0.902	**0.964**	0.932	0.847	0.656	0.739
		AMFB [8]	0.832	0.82	0.86	0.84	0.85	0.81	0.83
		FND-SCTI [29]	0.834	0.863	0.780	0.824	0.815	0.892	0.835
		HMCAN [15]	0.885	0.920	0.845	0.881	0.856	0.926	**0.890**
		BDANN [30]	0.842	0.830	0.870	0.850	0.850	0.820	0.830
		VilBERT	**0.9204**	**0.946**	0.948	**0.946**	**0.879**	**0.893**	0.885

Results

Experiments and Results (3/5)

A picture someone took of a shark swimming by their house when it got flooded 😱 \n#NewJersey #Hurricane #Sand http://t.co/OCXLWDFY

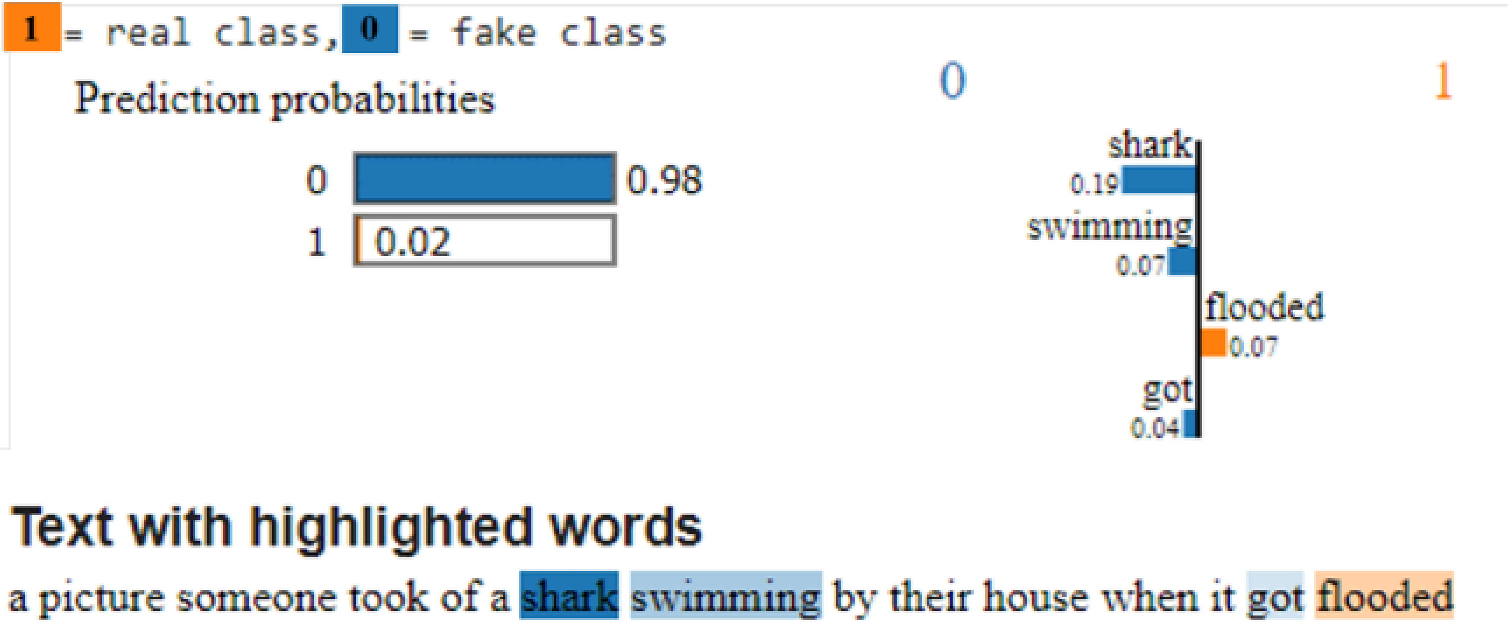
Input tweet example

Experiments and Results (4/5)

(a) (b) (c)

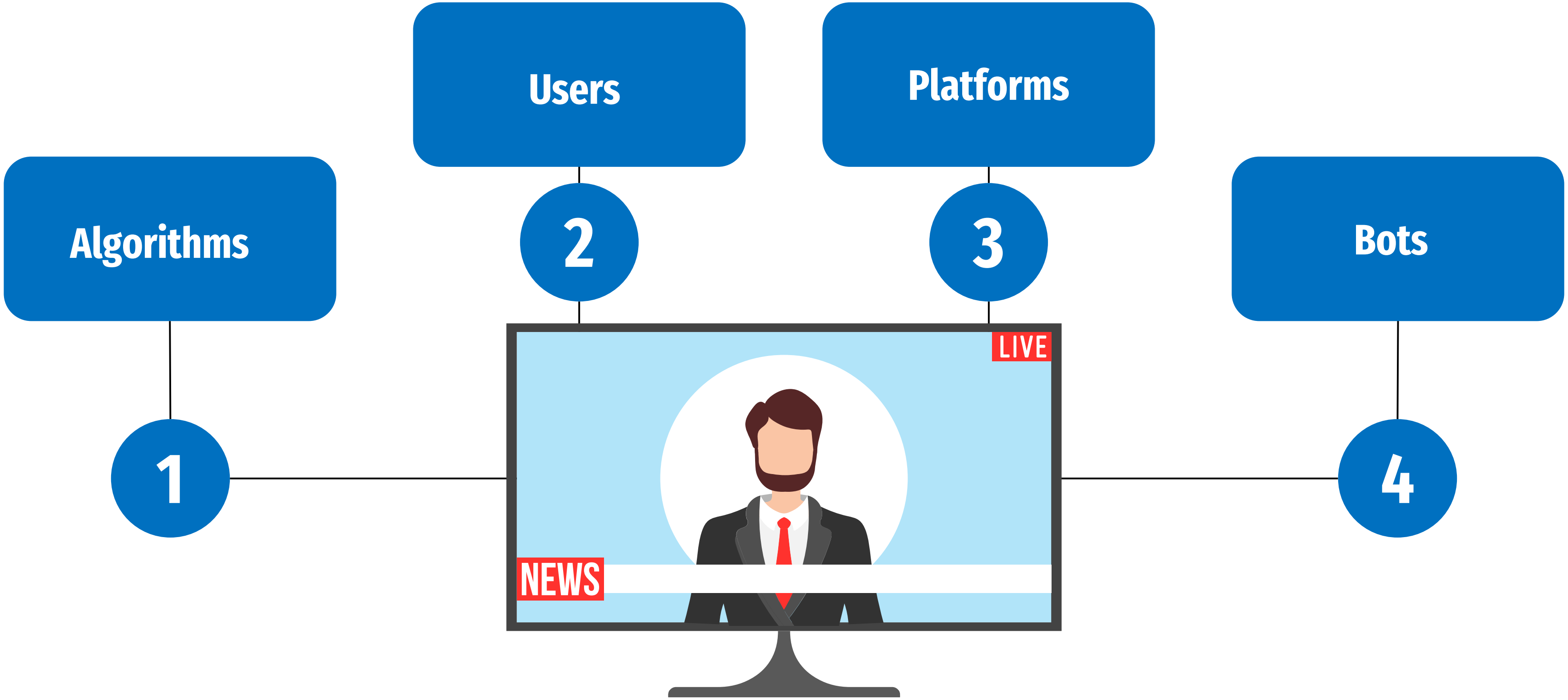
LIME explanations for image data. (a) presents the original fake tweet (b) shows the superpixels that are generated using the quickshift segmentation algorithm (c) shows the area of the image that produced the prediction of the class (fake, in our case)

Experiments and Results (5/5)



LIME explanations for textual data.

Fake news: Ongoing projects



Thank you –

Have questions or want to connect?

dorsaf.sallami@umontreal.ca
aimeur@iro.umontreal.ca