COVID19 Fake News Detection and Model Explanation

SIT723 - Student ID 219384532 (OSCAR WU)

Link to Video

SIT723 - Research Presentation

https://cdnapisec.kaltura.com/p/2006242/sp/200624200/embedIframeJs/uiconf_id/32025882/partner_id/2006242?iframeembed=true&playerId=kaltura_player&entry_id=1_r5r31h1d&flashvars%5BstreamerType%5D=auto&flashvars%5BlocalizationCode%5D=en&flashvars%5BleadWithHTML5%5D=true&flashvars%5BsideBarContainer.plugin%5D=true&flashvars%5BsideBarContainer.position%5D=left&flashvars%5BsideBarContainer.clickToClose%5D=true&flashvars%5Bchapters.plugin%5D=true&flashvars%5Bchapters.layout%5D=vertical&flashvars%5Bchapters.thumbnailRotator%5D=false&flashvars%5BstreamSelector.plugin%5D=true&flashvars%5BEmbedPlayer.SpinnerTarget%5D=videoHolder&flashvars%5BdualScreen.plugin%5D=true&flashvars%5BKaltura.addCrossoriginToIframe%5D=true&&wid=1_u0t5emmt

https://cdnapisec.kaltura.com/p/2006242/sp/200624200/embedIframeJs/uiconf_id/32025882/partner_id/2006242?iframeembed=true&playerId=kaltura_player&entry_id=1_r5r31h1d&flashvars%5BstreamerType%5D=auto&flashvars%5BlocalizationCode%5D=en&flashvars%5BleadWithHTML5%5D=true&flashvars%5BsideBarContainer.plugin%5D=true&flashvars%5BsideBarContainer.position%5D=left&flashvars%5BsideBarContainer.clickToClose%5D=true&flashvars%5Bchapters.plugin%5D=true&flashvars%5Bchapters.layout%5D=vertical&flashvars%5Bchapters.thumbnailRotator%5D=false&flashvars%5BstreamSelector.plugin%5D=true&flashvars%5BEmbedPlayer.SpinnerTarget%5D=videoHolder&flashvars%5BdualScreen.plugin%5D=true&flashvars%5BKaltura.addCrossoriginToIframe%5D=true&&wid=1_u0t5emmt

Background & Motivation

The outbreak of COVID-19 in late 2019 has since then resulted in massive misinformation about the virus across social media platforms.

According WHO, Fake news, such as “eating garlic”, which result in uncertainty and negative effect in Community

WHO Post on Twiiter

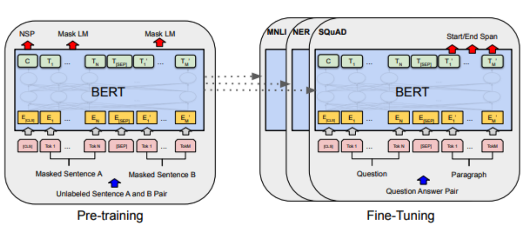
"Eating garlic can kill COVID-19"

Garlic: It may be good for general health, but it won't stop the coronavirus

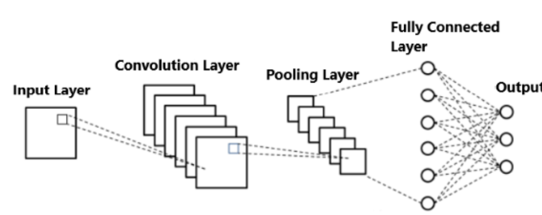
💡 Existing Studies used ML and DL to detect fake news.

Complex

Lack of Bench Marking Dataset to conduct fake news detection



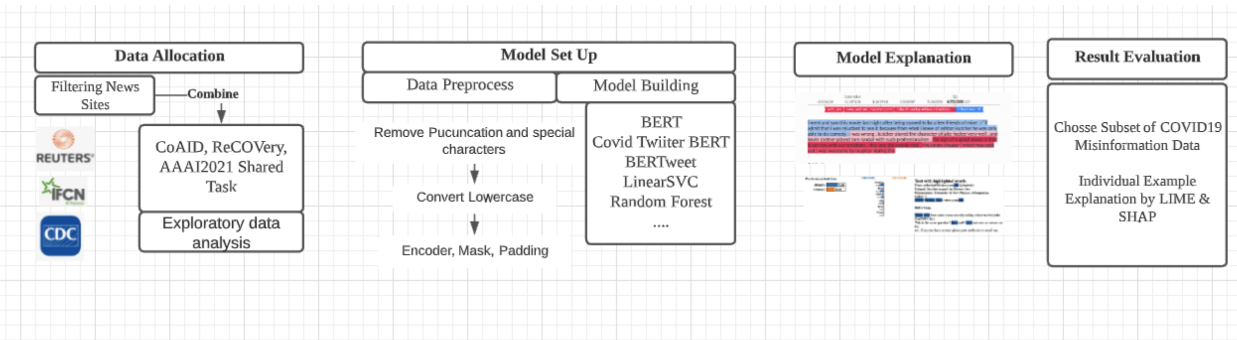
Sourced by: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding



CNN Architectures

These Black Models Lack of Explainability

Summary of Artefact



Objetives:

- To collect Multilingual COVID-19 Fake News Dataset
- A comparative of review ML models in detecting COVID-19
- To implement Model-Agnostic-Method (i.e., SHAP and LIME) to interpret model prediction

Result

- Dataset info

Table 3: Dataset Basic information

Covid19 fake news Dataset Info	
Attribute	**#**
Source Website	101
Country	116
Languge News Used	35
Unique Label	4
Dataset Shape	15041
Dataset Collected Date Range	2020-01-05 ~2022-01-15

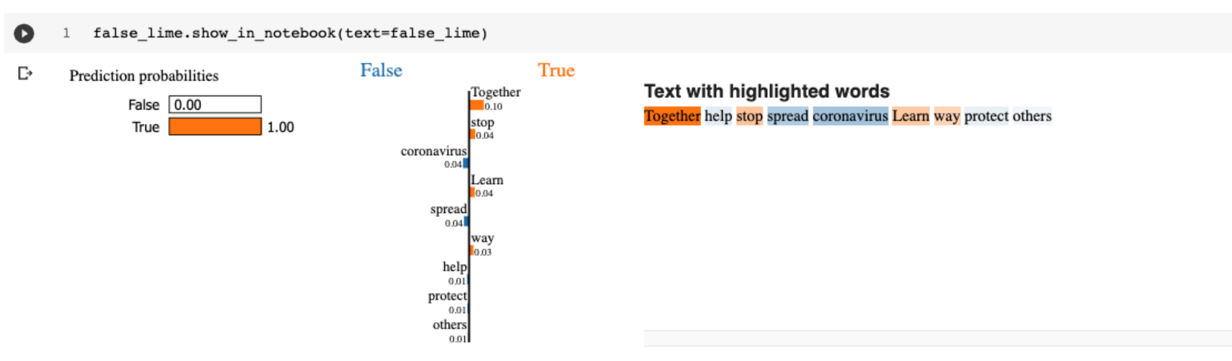
Table 5: Model Performance on COVID-19 Fake News Dataset

		Metrics				
	Model Name	F1-Score (False/True)	Recall (False/True)	Precision (False/True)	ACC	AUC
No Additional Data	LinearSVC	0.9188/0.7086	0.9125/0.7266	0.9252/0.6914	0.8730	0.8196
	LogisticRegression	0.9317/0.7295	0.9071/0.8172	0.9576/0.6588	0.8909	0.8621
	RandomForest	0.9109/0.6805	0.9049/0.6971	0.9170/0.6647	0.8606	0.8010
	CT-BERT-v2	0.9791/0.9291	0.9780/0.9325	0.9802/0.9256	0.9677	0.9553
	BERTweet	0.9740/0.9049	0.9846/0.8704	0.9636/0.9422	0.9591	0.9275
	Bert-large	0.9668/0.8918	0.9573/0.9218	0.9766/0.8636	0.9492	0.9400
	RoBERTa-large	0.9718/0.9106	0.9514/0.9773	0.9932/0.8525	0.9572	0.9643
	DistilBERT	0.9661/0.8787	0.9721/0.8595	0.9601/0.8987	0.9470	0.9158
Added Extra Data	LinearSVC	0.8416/0.8302	0.8409/0.8310	0.8423/0.8295	0.8361	0.8359
	LogisticRegression	0.8604/0.8498	0.8277/0.8152	0.8580/0.8523	0.8553	0.8552
	RandomForest	0.8327/0.8189	0.8628/0.8472	0.8282/0.8238	0.8261	0.8260
	CT-BERT-v2	0.9642/0.9613	0.9696/0.9555	0.9589/0.9671	0.9628	0.9625
	BERTweet	0.9379/0.9300	0.9621/0.9044	0.9150/0.9570	0.9342	0.9332
	Bert-large	0.9676/0.9645	0.9786/0.9528	0.9569/0.9766	0.9661	0.9657
	RoBERTa-large	0.9693/0.9676	0.9615/0.9759	0.9771/0.9595	0.9685	0.9687
	DistilBERT	0.9288/0.9172	0.9657/0.8782	0.8946/0.9599	0.9234.	0.9219

News Example: Together help stop spread coronavirus Learn way protect others

SHAP

LIME



Conclusion and Further Work

The study's findings provide three contributions: it describes the entire data gathering process and compares fake news detection models, including regular ML and BERT-based models. Furthermore, model-agnostic methods (i.e., SHAP and LIME) were presented in this study to show explainability in BERT-based models to promote public trust in ML.