

HYPERPARTISAN NEWS DETECTION

(SEMEVAL-2019 TASK 4)

By:

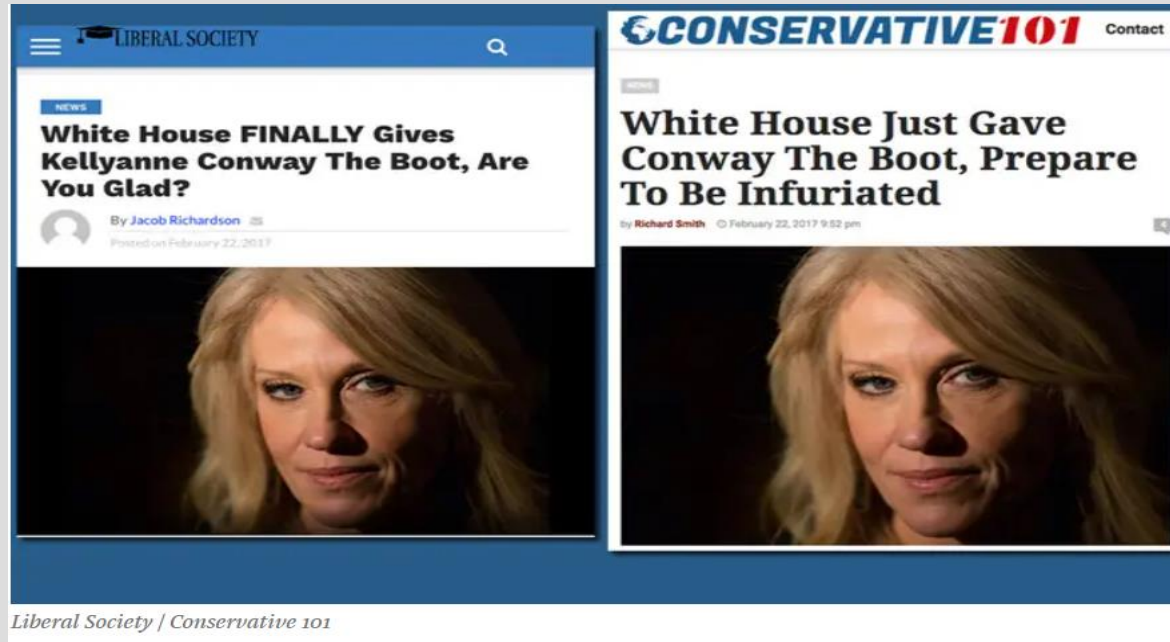
Maisha Binte Moin (bintemoi@ualberta.ca)



INTRODUCTION

- The use of social media has caused a shift in our manner of news consumption
- But social media can be an outlet for misinformation
- One subtle form of misinformation is hyperpartisan news, which exhibits extreme bias towards a single side in a political spectrum
- Spread at an alarming speed

MOTIVATION



‘If you want to know any fact about politics, you must read at least three different papers, compare at least three different versions of the same fact, and come in the end to your own conclusion.’ - Virginia Woolf

RELATED WORK

- Previously, fake news detection had been widely studied in comparison to hyperpartisan news detection
- Study (Potthast et al., 2017) that analyzed the writing style of hyperpartisan news to distinguish them from mainstream news
- Various deep learning and machine learning approaches used in SemEval-2019 Task 4
- Machine learning: SVM, Logistic Regression, Random Forest, XGBOOST, Naïve Bayes
- Deep learning: CNN (Jiang et al., 2019), BiLSTM (Zhang et al., 2019), BERT (Drissi et al., 2019)

DATA

Hyperpartisan

Donald Trump ran on many braggadocios and largely unrealistic campaign promises. One of those promises was to be the best, the hugest, the most competent infrastructure president the United States has ever seen. Trump was going to fix every infrastructure problem in the country and Make America Great Again in the process. That is, unless you're a brown American. In that case, you're on your own, even after a massive natural disaster like Hurricane Maria.

Not Hyperpartisan

President Donald Trump once left his son, Donald Trump Jr., stranded on a tarmac after he was five minutes late for a flight, the president's first wife Ivana Trump wrote in her new book "Raising Trump." Ivana referenced the anecdote when describing a key feature of Donald and the rest of the Trump family that she learned when she and the future president first began dating. The entire Trump clan arrived exactly on time," she wrote. "I learned early on that they were punctual to an extreme. For them, 'on time' meant five minutes early."

Two types of dataset have been made available for the task:

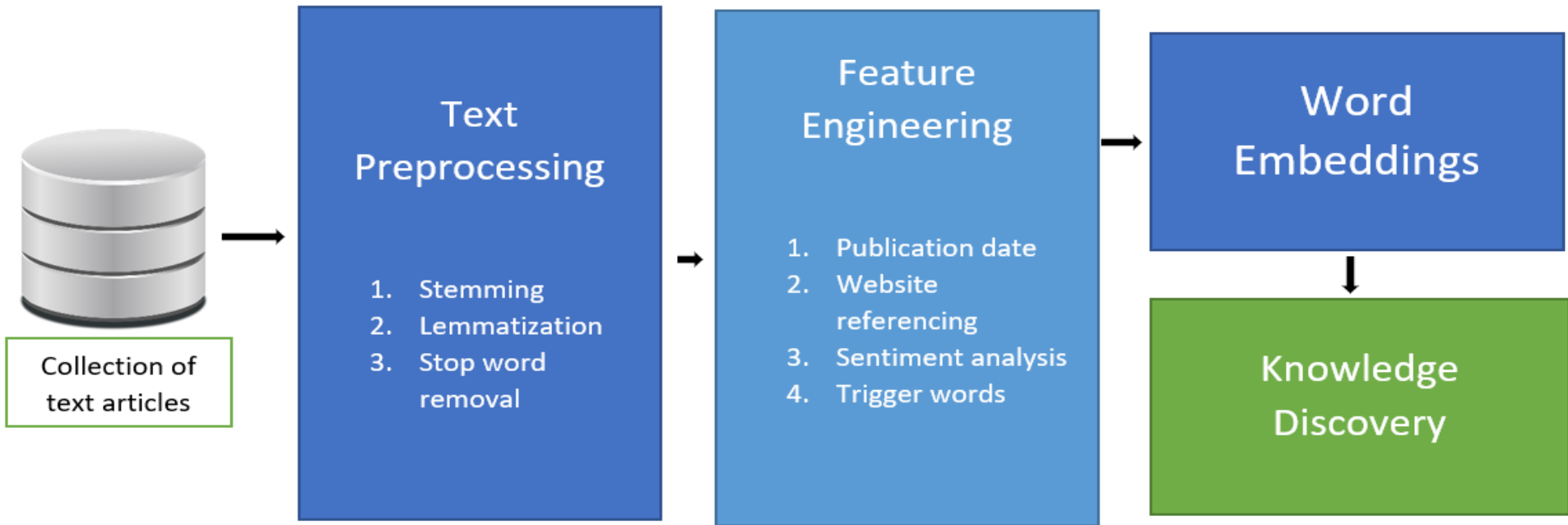
(1) By article

Public set: 645 articles; hidden test set: 628 articles, balanced

(2) By publisher

Public set: 750 000 articles, balanced; hidden test set: 4 000 articles

METHODOLOGY



RESULTS

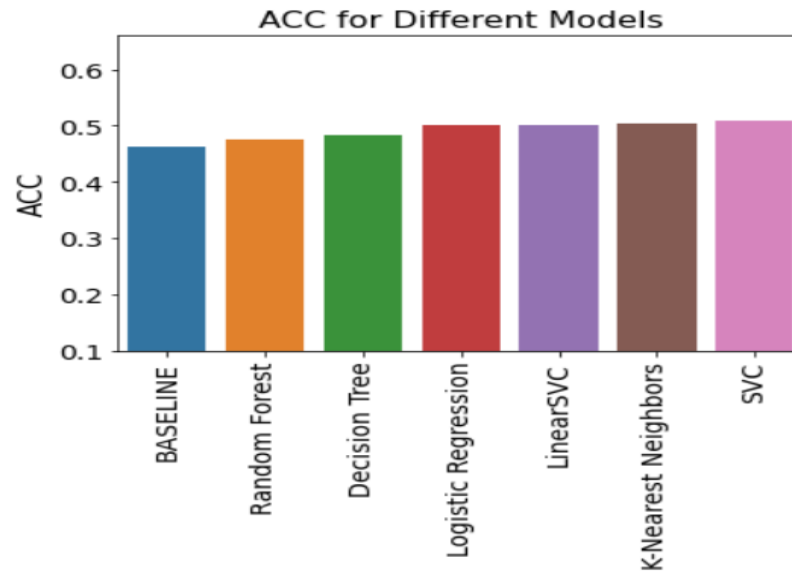


Figure 1: Accuracy comparison

Table 1: Classifier Results

No.	Model	ACC
1	Decision Tree	0.476
2	Random Forest	0.479
3	Logistic Regression	0.500
4	Linear SVC	0.501
5	K-Nearest Neighbors	0.504
6	SVC	0.530

DISCUSSION

Rank	publisher	freq.
1	thegatewaypundit.com	17
2	opslens.com	14
3	realclearpolitics.com	13
4	nypost.com	10
5	salon.com	8
6	express.co.uk	7
7	opednews.com	7
8	dcclothesline.com	6
9	turtleboysports.com	6
10	pjmedia.com	6
11	trueactivist.com	5
12	video.gq.com	4
13	bearingarms.com	4
14	breakingisraelnews.com	4
15	washingtonexaminer.com	4
Total		115

Table 6: The frequency distribution of the Top 20 most frequent publishers of hyperpartisan news in the by-article training set.

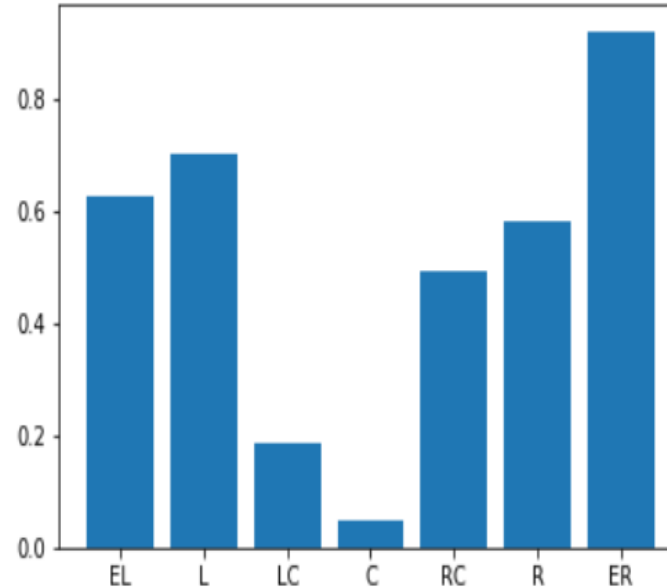


Figure 2: Percentage of hyperpartisan articles by publisher bias

Rank	publisher	freq.
1	cbsnews.com	24
2	circa.com	21
3	express.co.uk	13
4	heavy.com	12
5	snopes.com	11
6	nbcnews.com	8
7	nfl.com	8
8	insider.foxnews.com	8
9	nytimes.com	7
10	businessinsider.com.au	7
11	king5.au	6
12	kiro7.com	6
13	nypost.com	5
14	kcra.com	5
15	bradenton.com	4
Total		145

Table 7: The frequency distribution of the Top 20 most frequent publishers of mainstream news in the by-article training set.

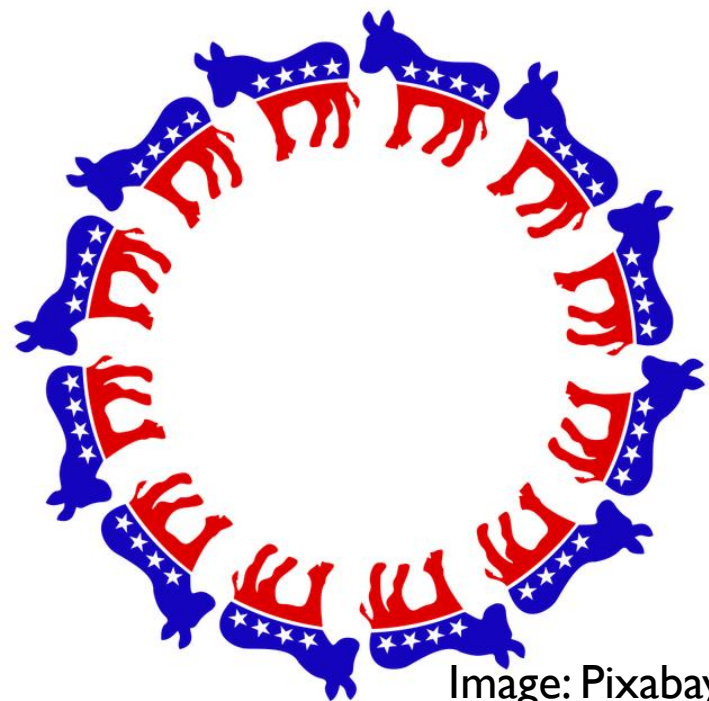
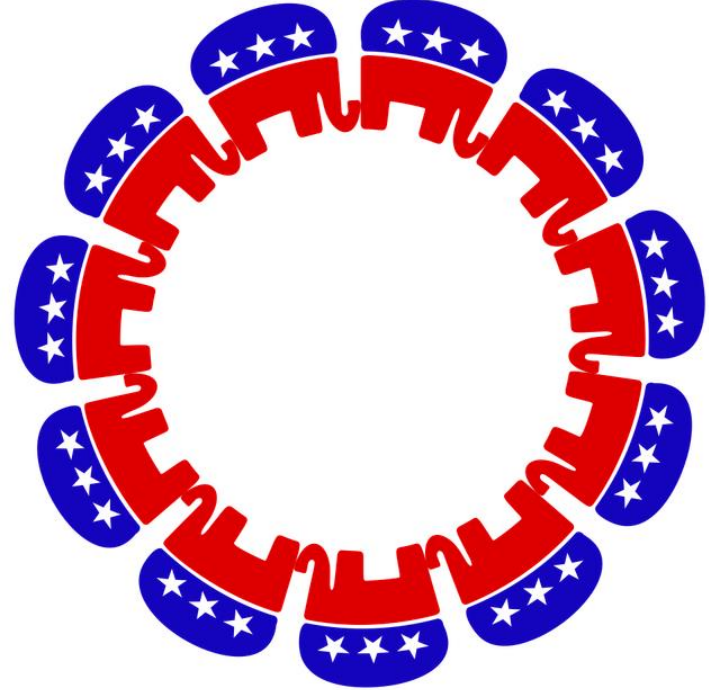
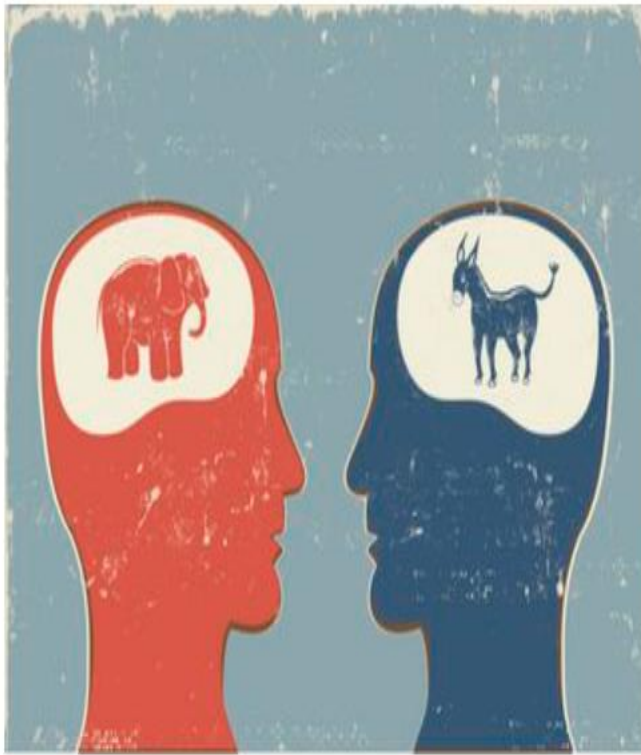


Image: Pixabay

FUTURE WORK

- Expanding the set of explainable features
- Experimenting with more recent transformer architectures:
 - (1) Longformer (Beltagy et al., 2020)
 - (2) BigBird (Zaheer et al., 2020)
- Both can handle longer sequences of text compared to BERT (512 tokens)

CONCLUSION



- Study and characterization of useful features for model explainability
- The usage and comparison of several machine learning algorithms for the task
- Proposal of possible ways of extending this work by using deep learning based approaches

REFERENCES

- [1] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638.
- [2] Ye Jiang, Johann Petrak, Xingyi Song, Kalina Bontcheva, and Diana Maynard. 2019. Team Bertha von Suttner at Semeval-2019 Task 4: Hyperpartisan News Detection using ELMo Sentence Representation Convolutional Network. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 840–844.
- [3] Chiyu Zhang, Arun Rajendran, and Muhammad Abdul-Mageed. 2019. UBC-NLP at SemEval-2019 Task 4: Hyperpartisan News Detection With Attention-Based Bi-LSTMs. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 1072–1077.
- [4] Mehdi Drissi, Pedro Sandoval, Vivaswat Ojha, and Julie Medero. 2019. Harvey MUDD college at SemEval-2019 Task 4: The Clint Buchanan hyperpartisan news detector. arXiv preprint arXiv:1905.01962.
- [5] Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
- [6] Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. Advances in Neural Information Processing Systems, 33:17283–17297.

THANK YOU!