

# Spider-Jerusalem at SemEval-2019 Task 4 Hyperpartisan News Detection

**Amal Alabdkarim** aa4235@columbia.edu

**Tariq Alhindi** tariq@cs.columbia.edu

## Problem

Fake news effect public opinions on topics such as climate change and political elections. Computational methods for fighting fake news mainly focus on automatic fact-checking rather than looking at writing styles of news articles.

## Contribution

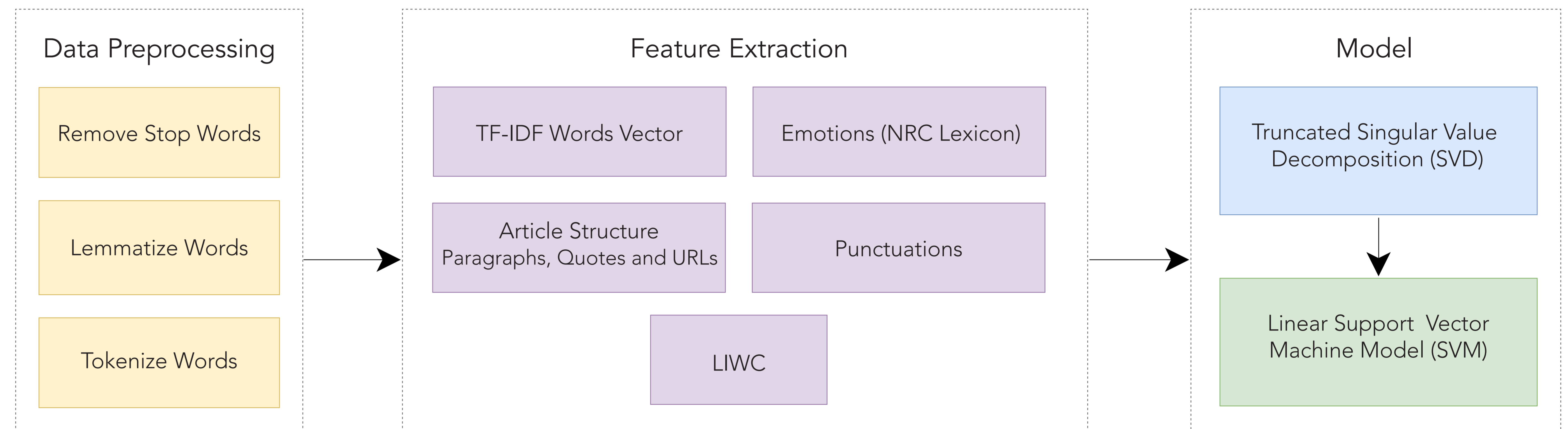
A system to detect hyperpartisan news articles as part of the shared task 4 in SemEval 2019.

<https://github.com/amal994/hyperpartisan-detection-task>

Our system was ranked seventeenth out of forty two participating teams in the binary classification task with an accuracy score of 0.742 on the blind test set.

## Method

We trained a Support Vector Machine model and a Truncated Singular Value Decomposition on a feature vector representing each article in our training dataset.



Hyperpartisan News Articles Classification System Overview

## Meta-Learning Experiment

We followed a class-based weighted majority approach, where the classifiers that are better in classification of one class were given a higher weight for that class predictions and lower weight for their predictions in the other class.

$$w(c, H) = \frac{\sum_i^n \mathbb{1}(y = H \text{ and } c(x) = H)}{\sum_i^n \mathbb{1}(y = H)}$$

Class Based Weights Equation

## Results

Measure	Validation Set	Test Set
Accuracy	0.767	0.742
F1-Score	0.767	0.709
Precision	0.767	0.814
Recall	0.767	0.627

Model performance on validation and test sets

As for the meta-learning, the classifier had a validation accuracy of 0.899 and the baseline majority vote classifier 0.884.

## Conclusion

- We presented a summary of our experiments and analysis of our results and prediction errors.
- This task was challenging due to the complexity in labeling such articles, and differences in writing styles across domains, publishers and individuals.
- The small size of the training data along with the class imbalance also contributed to the complexity, which made it harder for the model to learn.

## Error Analysis

**Article 1:** 2016 US presidential elections  
**Publisher:** Fox News  
**Gold Label:** Mainstream  
**Predicted Label:** Mainstream

**Article 2:** Article about Ivanka Trump  
**Publisher:** Yahoo! News  
**Gold Label:** Mainstream  
**Predicted Label:** Hyperpartisan

**Article 3:** Article about Social Justice  
**Publisher:** Online Athens  
**Gold Label:** Hyperpartisan  
**Predicted Label:** Hyperpartisan

**Article 4:** A Joke Made by Jimmy Kimmel  
**Publisher:** Real Clear Politics  
**Gold Label:** Hyperpartisan  
**Predicted Label:** Mainstream