Filtering Out Fake News

by

Milad Shirani

June 2022

Contents

- Project Overview
- Data
- Modeling and Results
- Q & A

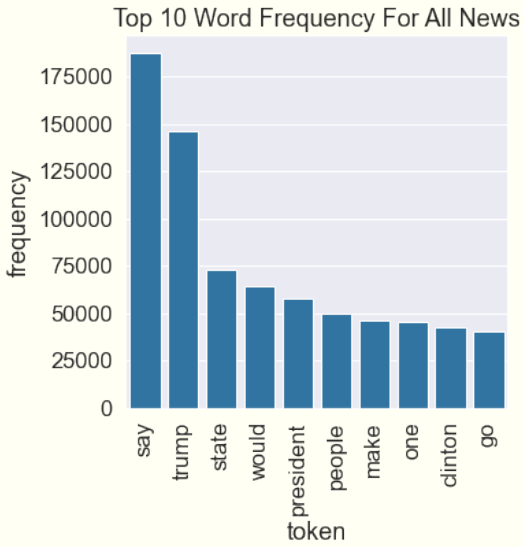
Project Overview

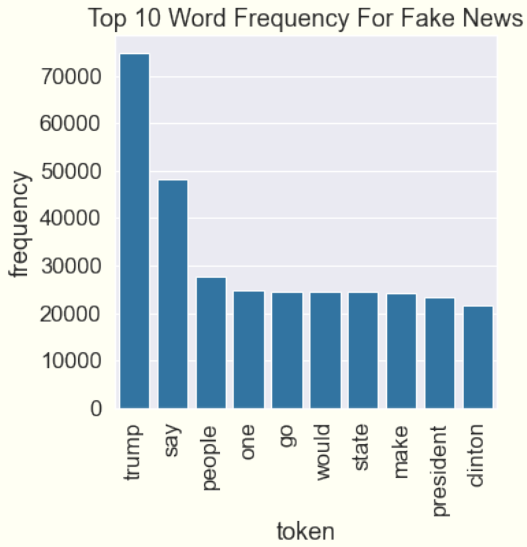
1. One of the main challenges that social medias and news sources face is detecting fake news.
2. In this work, we want to introduce a new model to filter out fake news.
3. We used several categorical models such as Logistic Regression, Decision Tree, Random Forese, XGBoost, LightGBM, Neural Network, Transfer Learning.

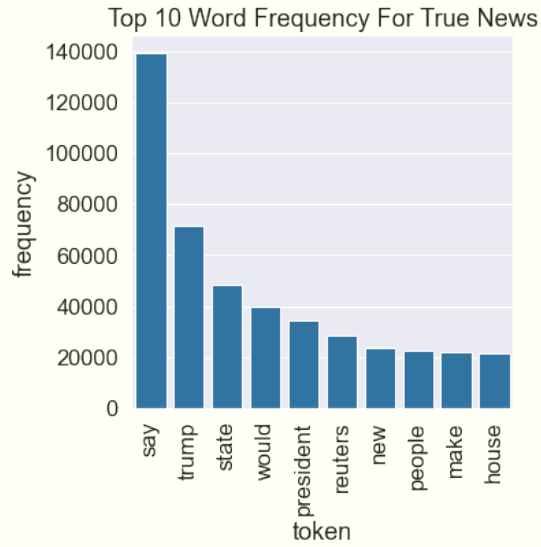
Data

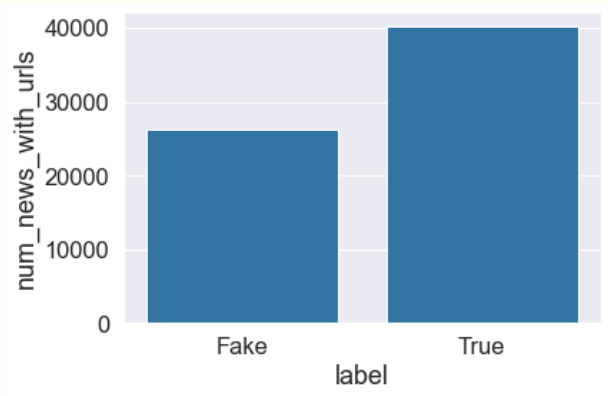
- We use 76525 data points for our analysis from the following sources
  1. Fake and real news dataset and it contains 44898 data points.
  2. Source based Fake News Classification and it contains 2096 data points.
  3. REAL and FAKE news dataset and it contains 6335 data points.
  4. GitHub Repo and it contains 6335 data points.

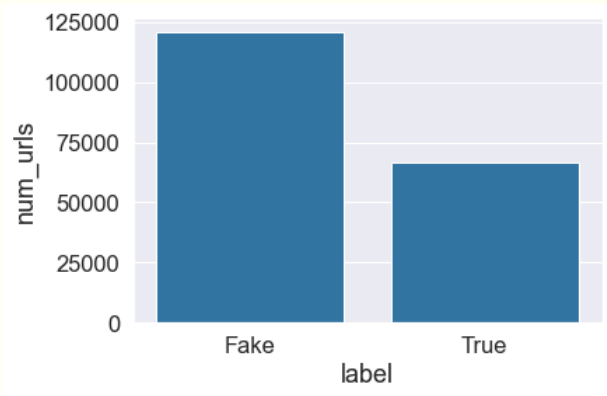
Modeling And Results











Modeling and Results

1. The final model is obtained by using the embedding layer "nnlm-en-dim128"published by Google on TensorFlow Hub.
2. We considered recall-score as well as accuracy of the model as our metrics. The results of this model are:

data	index	Fake	True	accuracy	macro avg	weighted avg
TEST	precision	0.89	0.90	0.90	0.90	0.90
	recall	0.85	0.93	0.90	0.89	0.90
	f1-score	0.87	0.92	0.90	0.89	0.90
	support	8719.00	13194.00	0.90	21913.00	21913.00
TRAIN	precision	0.92	0.92	0.92	0.92	0.92
	recall	0.87	0.95	0.92	0.91	0.92
	f1-score	0.89	0.93	0.92	0.91	0.92
	support	17502.00	26987.00	0.92	44489.00	44489.00

Conclusion

1. Filtering out fake news is one of the main challenges that social medias and news sources face.
2. The final model we introduce is obtained by fine tuning the embedding layer "nnlm-en-dim128"published by Google on TensorFlow Hub.

Next Steps

1. Gathering more data points for training.
2. Adding LSTM or Conv1d layers may improve the model's performance.
3. Trying tuning hyper-parameters of the model by using optuna may improve the performance of the model.

Q and A