Fake News Detection

Classifying Unreliable and Inflammatory Stories with NLP

Problem

Social media sites are flooded with misinformation or “fake news” and this poses real problems for both our democracy and the platforms brands.

This is a difficult problem with few options:

1) Flag such news stories
2) Remove or demote the presence of the articles in people’s feeds

All options have issues, but we need scalable automated or semi-automated solutions

What Can a Model Do?

1) Recognize similarities between fake or real stories that it was trained on

2) Recognize formatting and coherence issues in articles

3) Recognize contextual features of certain sources known to spread false information

4) Use a known evidence base and determine reliability relative to evidential support

This project will mainly use the first approach

The Data

FakeNewsNet

Consists of ~23,000 news articles marked by PolitiFact and GossipCop as either fake or real scraped between 3/2/22 - 3/4/22

Kaggle: Fake and real news dataset

17903 Fake stories marked by PolitiFact and 20826 Real stories from Reuters

Kaggle: Source Based Fake News Classification

2006 stories marked as fake or real by the BS Detector Chrome Extension

A Closer Look at the Data

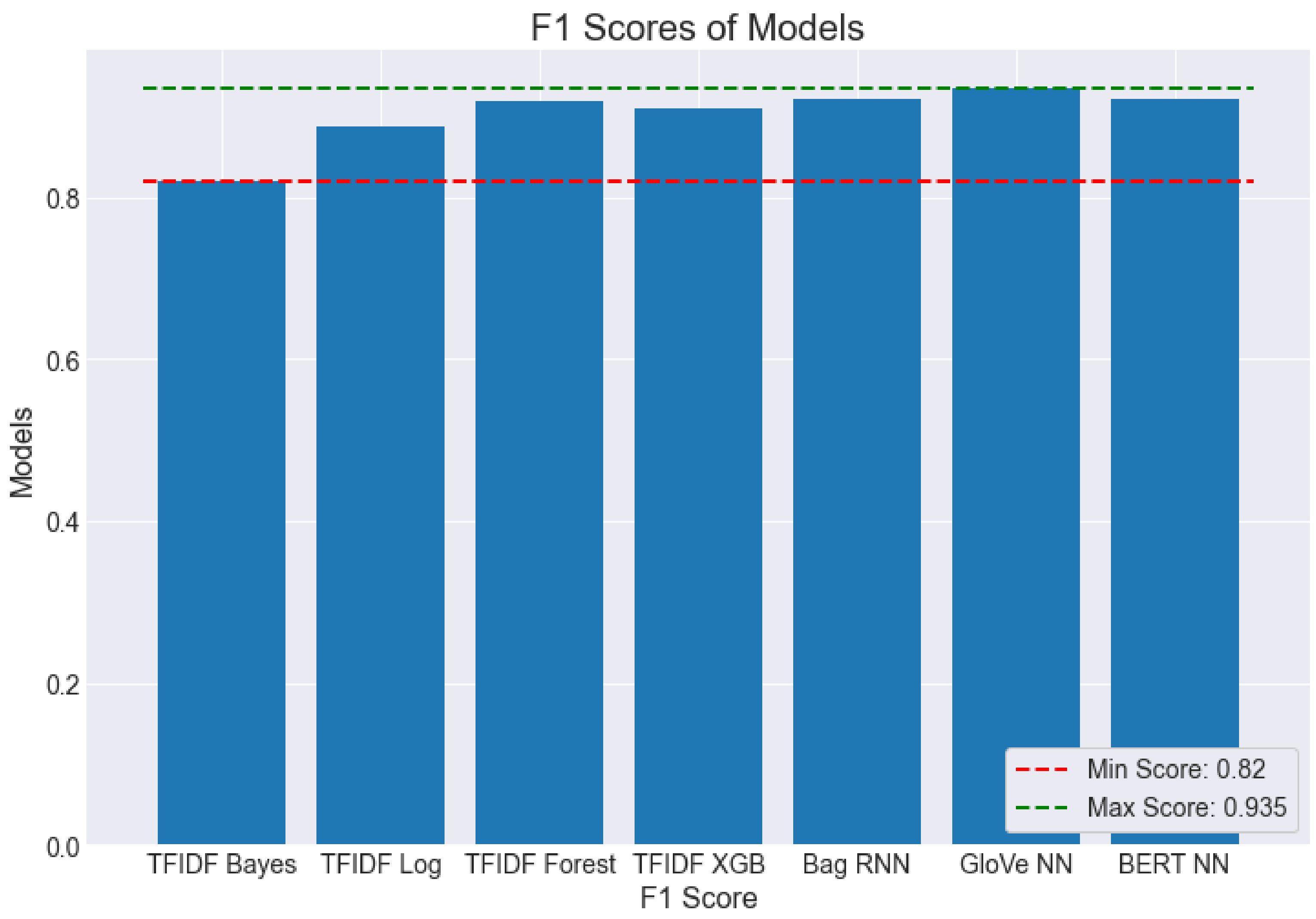
Most Frequently Occurring Bigrams in Real News

Most Frequently Occurring Bigrams in Fake News

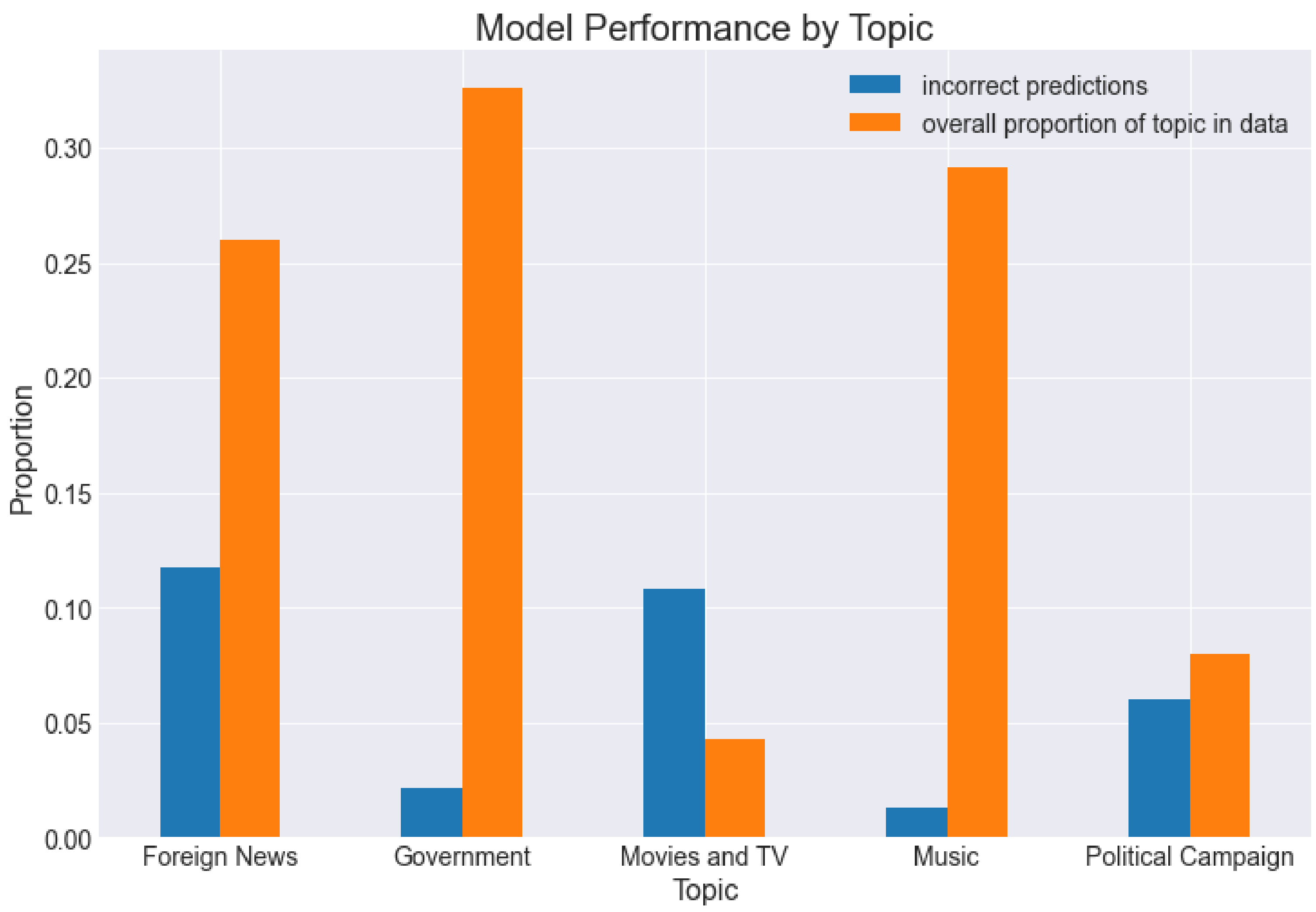
Methods of Modeling

- Naive Bayesian with TF-IDF vectorization
- Logistic with TF-IDF vectorization
- Random Forest with TF-IDF vectorization
- XGBoost with TF-IDF vectorization
- Neural net with bag of words embedding
- Neural net with GloVe embedding
- Neural net with Small BERT embedding

Model Results



Final (GloVe) Model Performance by Topic



LDA was used to cluster the data by topic and the five clusters were given rough names to indicate article content.

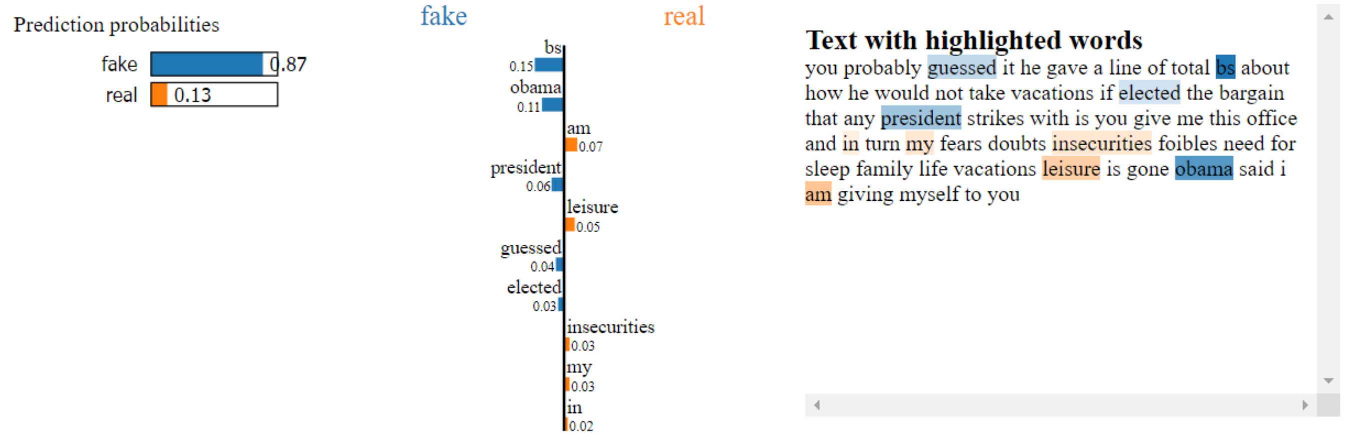
Interpreting and Using the Model

- Lime was used to examine how each mode performed on a sample of articles to gain insight into how the models arrived at their predictions

- A Streamlit app was made to allow use of the model and easier interpretation of the results with Lime

In [185]:
```
glove_pipe = make_pipeline(tokenizer, wrapped_glove)
exp_glove = explainer.explain_instance(X[fake_idx], glove_pipe.predict_proba, labels=[1])
exp_glove.show_in_notebook(text=X[fake_idx])
```

```
157/157 [==============================] - 46s 294ms/step
```



Limitations and Future Work

- Models performed quite well, but the test data and training data were quite similar in content, it may not generalize
- Since the model uses the mention of words related to topic, as topics change the performance of the model will degrade
- Future work could expand the data to include health related misinformation from COVID fake news data
- Work on coherence or evidence base models could be more generalizable.

Citations

- @article{turc2019, title={Well-Read Students Learn Better: On the Importance of Pre-training Compact Models}, author={Turc, Iulia and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina}, journal={arXiv preprint arXiv:1908.08962v2 }, year={2019} }
- @article{shu2018fakenewsnet, title={FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media}, author={Shu, Kai and Mahudeswaran, Deepak and Wang, Suhang and Lee, Dongwon and Liu, Huan}, journal={arXiv preprint arXiv:1809.01286}, year={2018} }
- @article{shu2017fake, title={Fake News Detection on Social Media: A Data Mining Perspective}, author={Shu, Kai and Sliva, Amy and Wang, Suhang and Tang, Jiliang and Liu, Huan}, journal={ACM SIGKDD Explorations Newsletter}, volume={19}, number={1}, pages={22--36}, year={2017}, publisher={ACM} }
- @article{shu2017exploiting, title={Exploiting Tri-Relationship for Fake News Detection}, author={Shu, Kai and Wang, Suhang and Liu, Huan}, journal={arXiv preprint arXiv:1712.07709}, year={2017} }
- @inproceedings{Ahmed2017DetectionOO, title={Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques}, author={Hadeer Ahmed and Issa Traor{'e} and Sherif Saad}, booktitle={ISDDC}, year={2017} }

Contact Info

Lia Elwonger

Email:

lelwonger@gmail.com

Github:

https://github.com/nonlocal-lia

Blog:

https://medium.com/@lelwonger