



Fake News: How to Sort Fact from Fiction

Ryan Mackie

Along with the presidential election in 2016 came the emergence of fake news, which has become a major problem in our connected world.

The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection has become an important topic of research in recent years.

Workflow

01

Data Acquisition

02

EDA/
Visualizations

03

Modeling:
LSTM Neural Net

04

Building an App

01.
**Data
Collection**



Data Acquisition

Politifact

- Updates from the PolitiFact can be scraped daily
- PolitiFact contains over 17,000 stories that have been fact checked, dating back to November of 2007
- For this project, I scraped the entire collection fact checked articles from PolitiFact - 17,730 articles in total

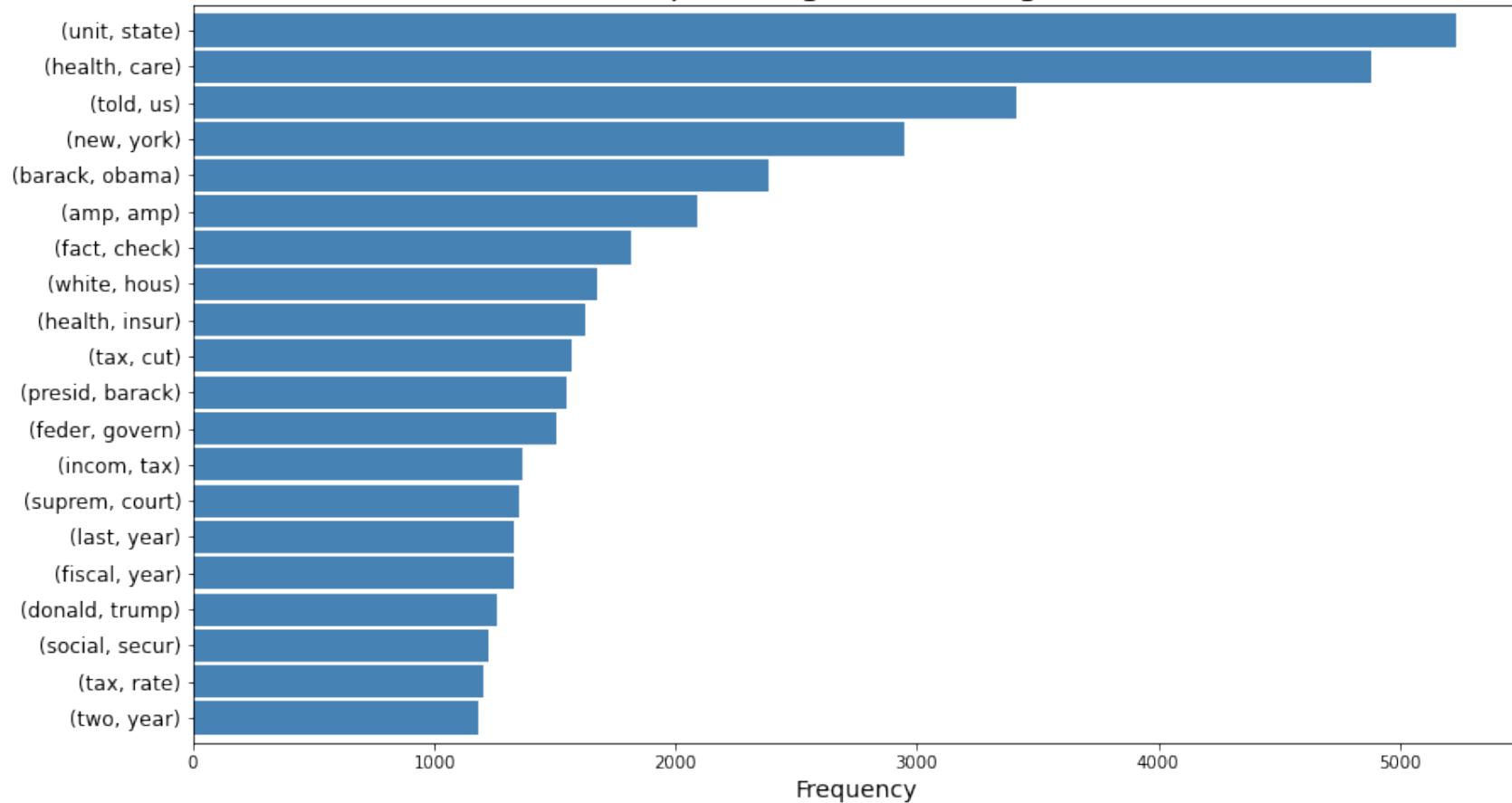
Kaggle

- Additional data was collected from Kaggle - Fake News
- The dataset used contains 20,387 unique stories that have been fact checked from between 2016 and 2018

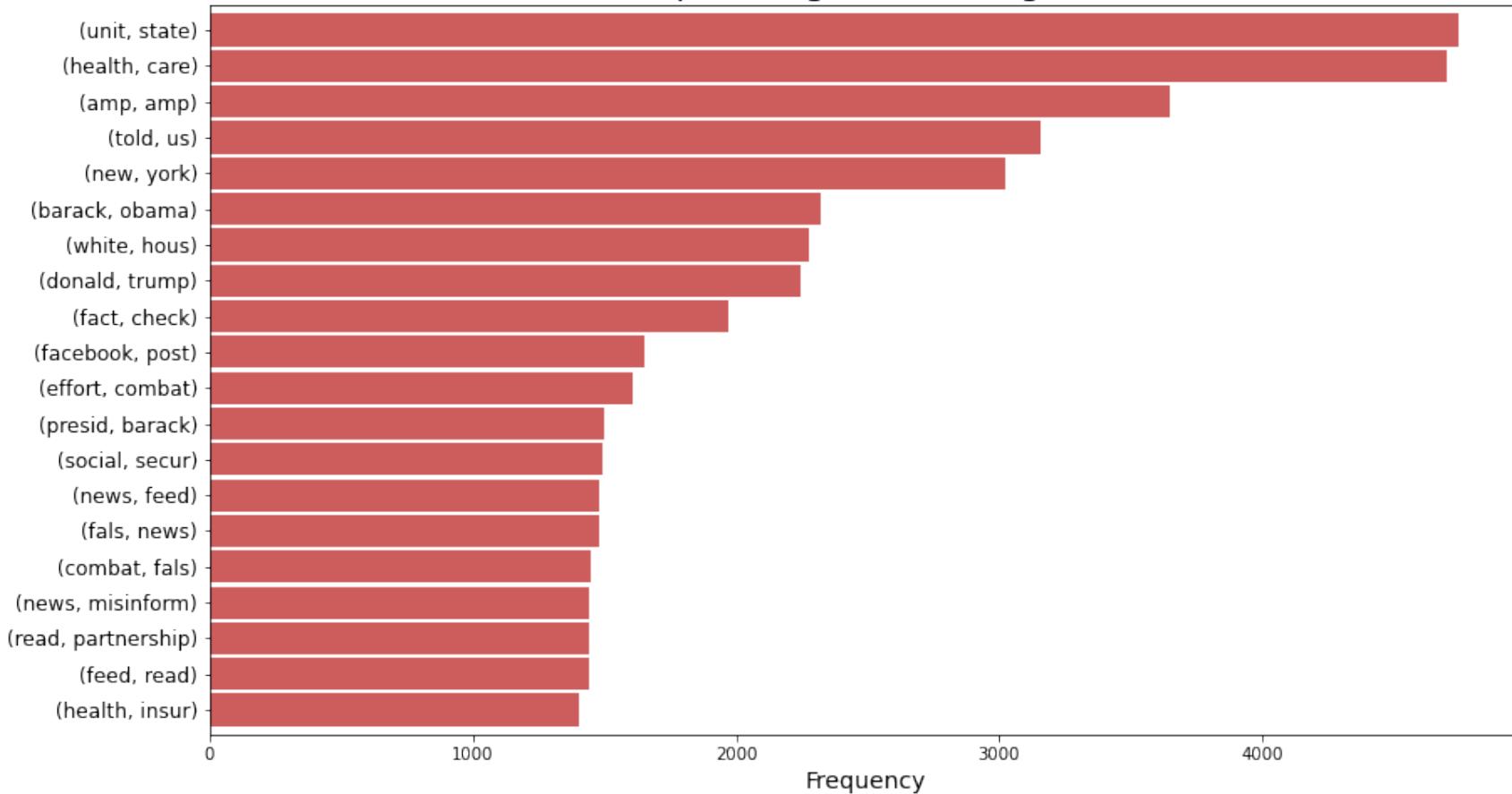
02. EDA/ Visualizations



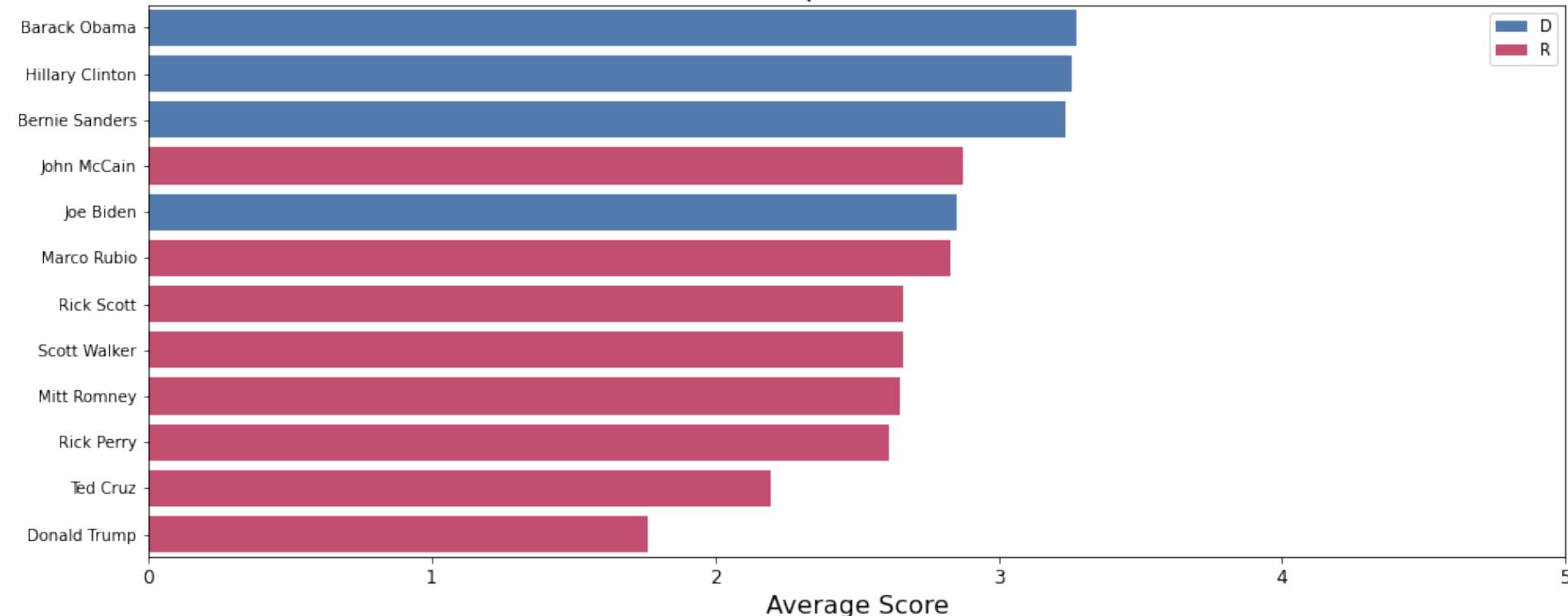
20 Most Frequent Bigrams Among True Stories



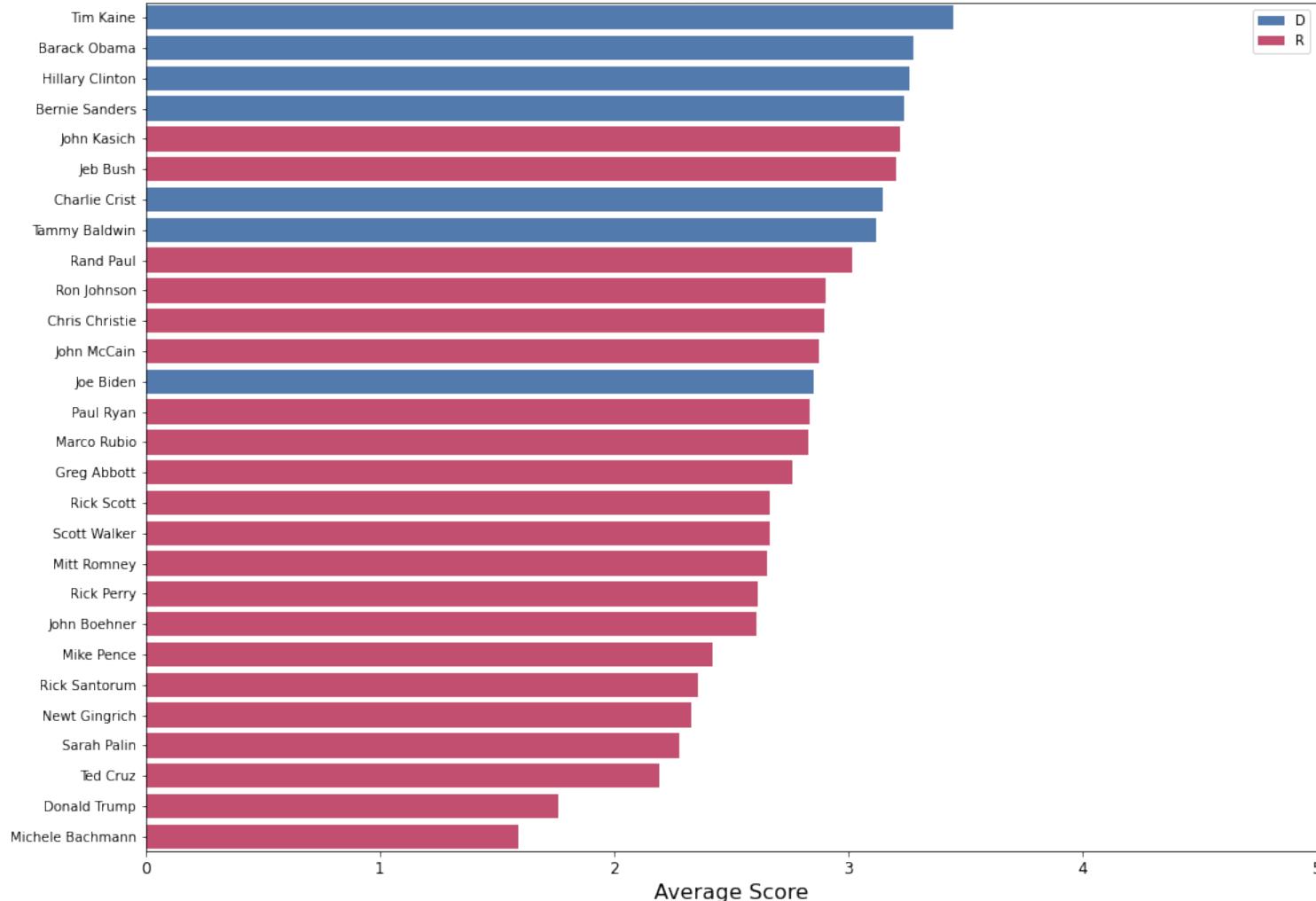
20 Most Frequent Bigrams Among False Stories



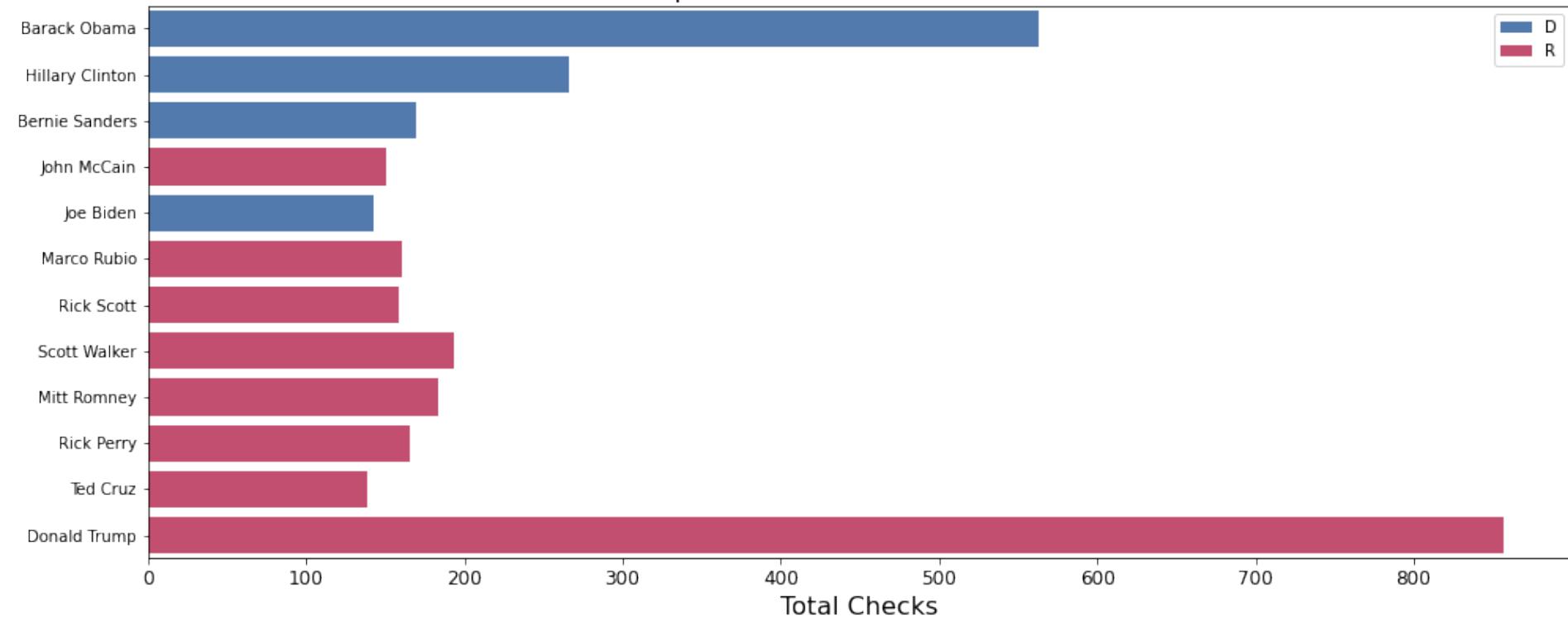
Politifact Scores: Top Fact Checked Politicians



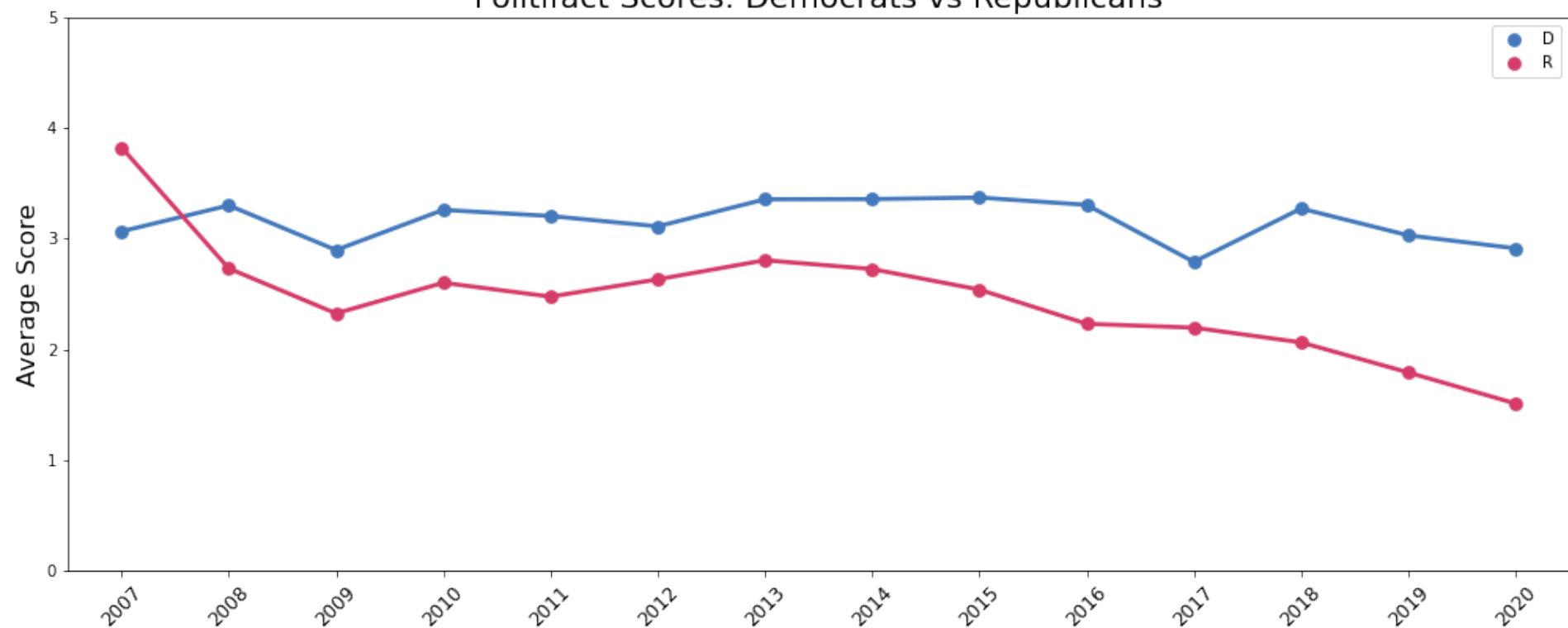
Politifact Scores: Top Fact Checked Politicians



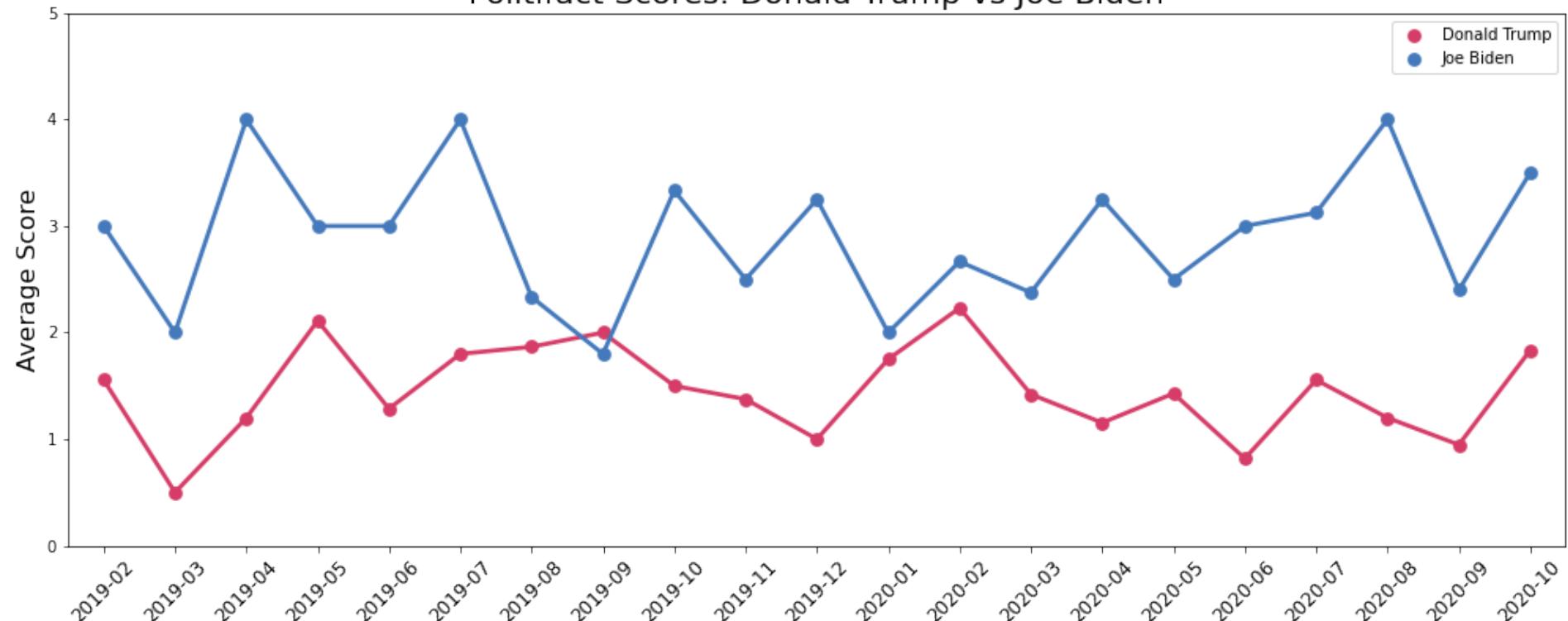
Top Fact Checked Politicians

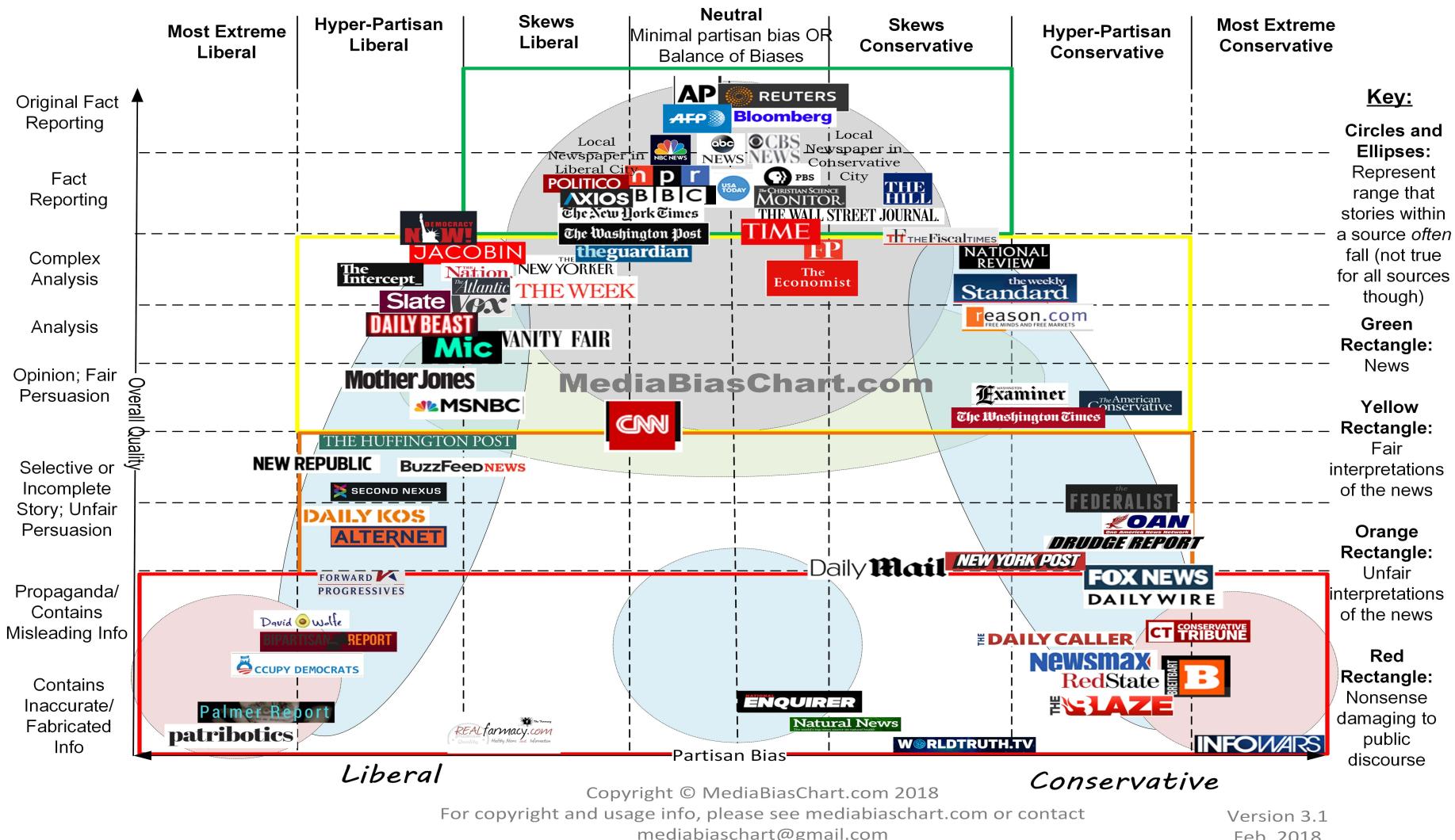


Politifact Scores: Democrats vs Republicans

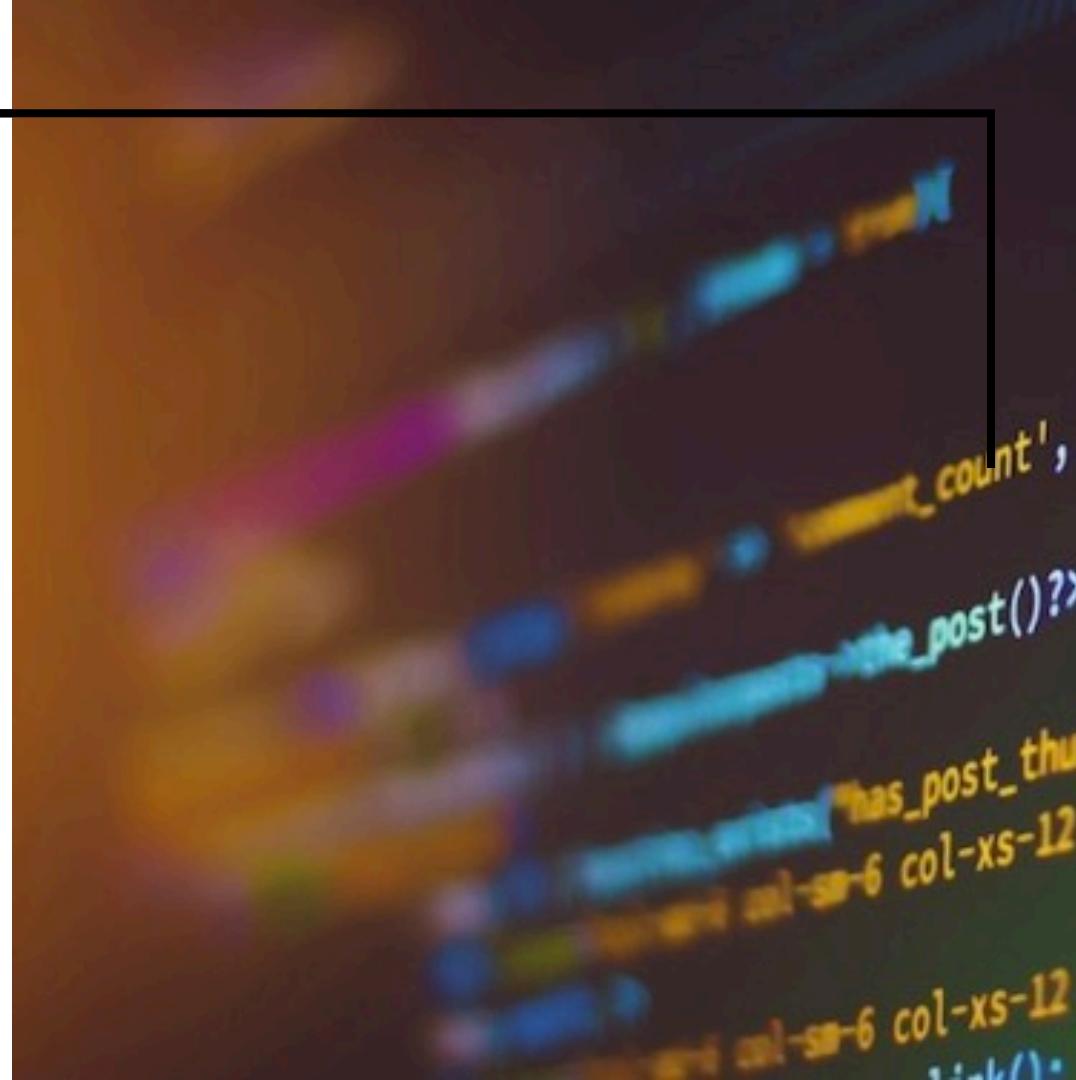


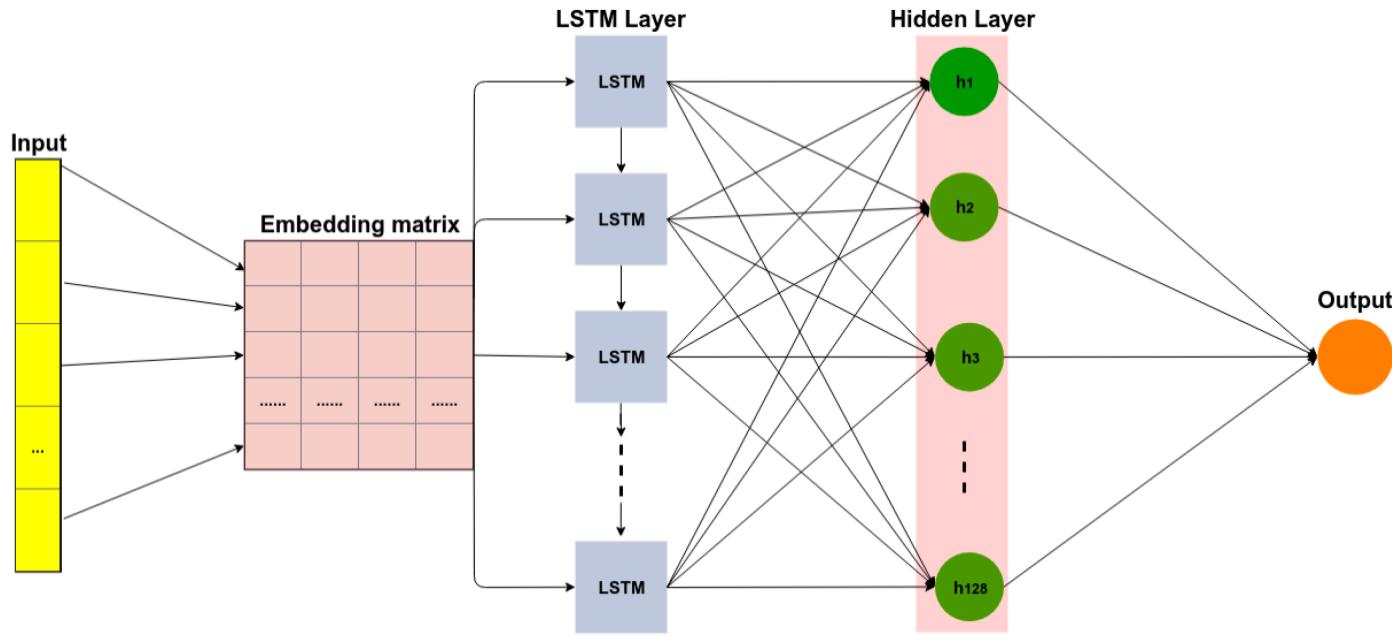
Politifact Scores: Donald Trump vs Joe Biden





03. Modeling: LSTM Neural Network





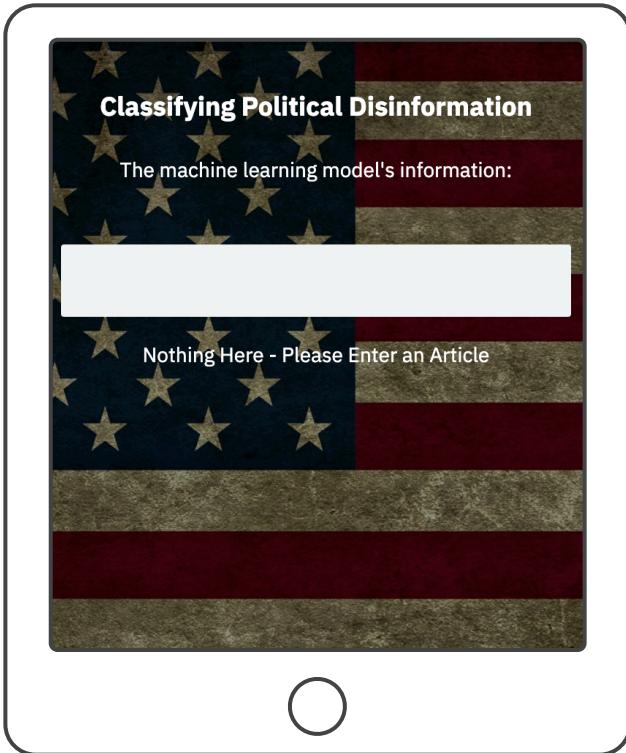
LSTM Model - Results

Politifact

- Using only data that was scraped from PolitiFact, the model was able to outperform the baseline accuracy of 51% by almost 18%.
- Decent results, but ran into trouble when using the app - not many reliable predictions.

Kaggle

- Updated the saved model to include the Kaggle dataset and this greatly improved the overall accuracy - 92%
- Some issues I could foresee by using this data:
 - Not sure how reliable the dataset is - where are the articles sourced from?
 - How are the articles classified - who is doing the classification?



04. App

THANKS!

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#).

RESOURCES

PHOTOS

- [Politifact Graphic](#)
- [538 Graphic](#)