Solving the problem

1. Fake news exists
2. We can distinguish fake news
3. But we can also use this classifier to create fake news using a GAN

GAN: <https://becominghuman.ai/generative-adversarial-networks-for-text-generation-part-1-2b886c8cab10>

maybe mention this? Looks pretty involved…

<http://www.fakenewschallenge.org/>

* Data set about stance detection: whether the headlines match up to the article

Another possibility is to use a fake image dataset

Talk about some of the difficulties in defining fake news and how an article was determined to be fake.

Try to use different data from different sources to limit articles being written in the same format and being easy to classify as true.

Fake News Detection on Social Media

<https://www.kdd.org/exploration_files/19-1-Article2.pdf>

Fake News Detection Using Deep Learning

<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2710385.pdf>

DATA

Kaggle data will need to be cleaned quite a bit - look at discussions

<https://www.kaggle.com/c/fake-news/data>

<https://www.kaggle.com/mrisdal/fake-news>

<https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset>

true data : <https://research.signal-ai.com/newsir16/signal-dataset.html> can maybe add to false data

Can download data from PolitiFact which has a false meter, we can choose the dates to download from so will add a temporal element!

<https://twitter.com/PolitiFact>

* Different neural networks and the differences and why we chose our one
* How did we find the best performing non-neural network (linear SVM ect) compare this to the neural network?
* I’ve heard of Long short-term memory or recurrent neural network
* Try different embeddings for different classifiers – which works better than others and why
* How easy is it to detect fake news and how easy is it to produce fake news using Generative Adversarial Networks (GAN)?
* Transformer NLP (mentioned in assignment!) apparently the peak of the NLP

<https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>

* Look at GloVe embeddings - You can derive semantic relationships between words from the co-occurrence matrix. <https://towardsdatascience.com/light-on-math-ml-intuitive-guide-to-understanding-glove-embeddings-b13b4f19c010>

extra to read

<https://arxiv.org/pdf/2001.00623.pdf>

<https://arxiv.org/pdf/1805.08751.pdf>

inspiration for project

<https://www.kaggle.com/madz2000/nlp-using-glove-embeddings-99-87-accuracy>