Project Proposal

Fact.err - Fake News Detection

Inderpartap Cheema **Najeeb Qazi** **Ishan Sahay** **Sachin Kumar**

OVERVIEW

Fake news has been a hot topic in recent years. As more and more people rely on online news sources, the threat increases. Sometimes they are satirical, and sometimes they aim at steering public opinion. During 2016 presidential elections in the United States, fake news started popping up as a propaganda to mislead readers. In Feb 2019, WhatsApp announced that it was deleting 2 million accounts per month to prevent the spread of fake news. The Indian government said these fake news played a major role in triggering up to 30 mob lynchings. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. In this project, we aim at classifying fake news from subtle yet consistent differences in the two classes of articles.

The problem of fake news detection is not easy to address because of the following reasons -

1. Multiple linguistic patterns
2. Numerous news categories
3. Unconventional news vs Fake news
4. Testing on a category of news that the system has never trained upon

We would try to compare several state-of-the-art models, and try to find the best-fit for the same. Our end goal is to make a system that would be constantly evolving and learning on the real-time news feeds and increasing the number of news categories it is able to classify.

PROJECT SPECIFICATIONS

Constraints	Goals
• Legal issues with data scraping • Fake news detection is not an 'easy' problem (*see Overview*) • Limitations on the accuracy of Machine Learning models • Difficulties with new technologies used.	1. Collecting news samples by either scraping or using existing APIs from various known websites such as The Guardian, New York Times and various other sources. We would also try to mix this with some pre-existing datasets and essentially make a Data Warehouse. 2. Using RNN as the Deep learning model to train and test the data. 3. Create a web-application to display the information. 4. Make a Real-Time model that can filter news articles on a recurrent basis.

DATA PIPELINE