Fake News Challenge

Prajwal Rao (5176504) & Julian Blair (3463793)  
COMP9417 2018s1 – Assignment 3

# Background

* Brief description of challenge
* Baseline provided
  + Explain how it works

# Implementation

## Problem structure hypothesis

* Divide and conquer approach – break up multi-class problem into several two-class problems

## Data Munging

* Common word filter:
  + Looking for words with high variance and low frequency (avoid common, homogeneous words)
  + Related to stop words in English, but specialised for entropy of words in articles

## Related vs. Unrelated

* Simplified baseline to two features (word overlap, headline-in-body word counts), added common word filter

# Results

* Some tables containing baseline results + divide-and-conquer results

# References

Galbraith, B., Iqbal, H., van Veen, H., Rao, D., Thorne, J. and Pan, Y. (2017) *Baseline FNC implementation*, 25 July, [Online], Available: <https://github.com/FakeNewsChallenge/fnc-1-baseline> [28 May 2018].

Pomerleau, D. and Rao, D. (2017) *Fake News Challenge*, [Online], Available: <http://www.fakenewschallenge.org/> [28 May 2018].