

Task 8: Analysis Report

Data sources for task 4, 5, 6, 7 are the published news articles from recognised news sources, public tweets from twitter and credibility ratings as provided by expert reviewers.

- The task 4 plot is a simple bar chart depicting the average credibility rating of various news sources that had received 5 or more reviews.
- Task 5 is a bar chart depicting average number of tweets for each credibility rating category.

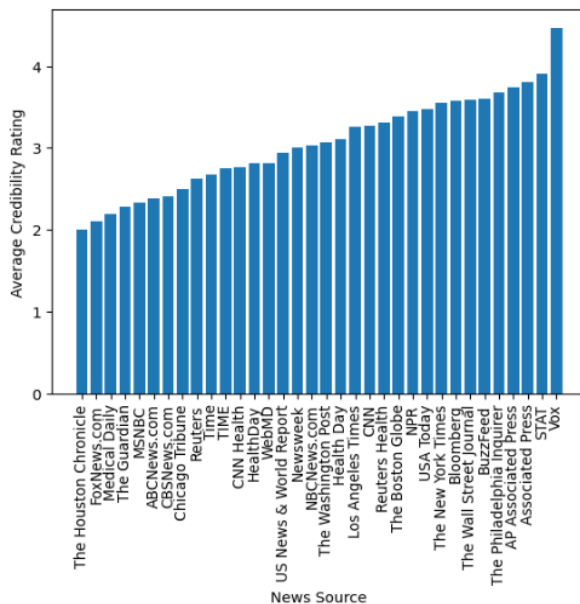


Figure 2: Task 4 - Average Credibility Rating

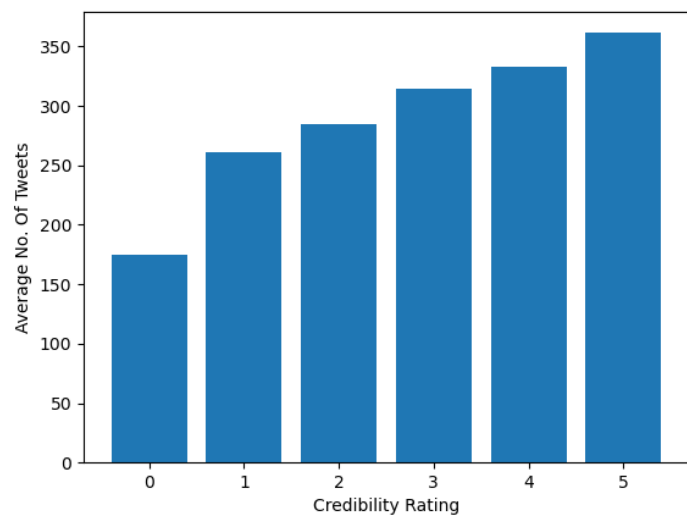


Figure 2: Task 5 - Average Number of Tweets

- Task 7 has 2 plots, the first being a distribution graph showing the distribution of log odds ratio, which is essentially an indicator of whether the word is likely to be used in a fake news article or a real news article. The second plot showcases the 30 extreme values that have the highest and lowest odds for fake news.

The range of the average credibility ratings is not very significant, being within 2 and 4.5. The most credible article according to reviewers is 'Vox' and the least credible is 'The Houston Chronicle', other notable sources are 'Fox', 'The Guardian' and most interesting is 'Medical Daily' who are all not very credible. Whereas 'Buzzfeed' is more credible than I expected.

I would say it is true, generally speaking there seems to be a positive relationship between the credibility of an article and the number of tweets it receives. However, it is still concerning the amount of tweets a low credibility article receives, especially those with a 0 rating.

LOG-ODDS-RATIO ANALYSIS. TASK 7

There is less variation between the data than I would've expecting, maybe using less credible or famous news sources would've led to more varied data. Also fake news has become more and more prevalent in recent years as it has been propagated by bots, the dataset ending in 2018 will likely reduce total volume of fake news and its propagation, also average numbers will be lower than relevant today, since articles written over a decade ago will still affect the numbers.