List of Codes:

1. read\_tweets.py: code runs over the “all-rnr-annotated-threads” file and reads data into a data frame. (PHEME dataset)

The “all-rnr-annotated-threads” can be retrieved in the following link:

<https://figshare.com/articles/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078>

1. convert\_veracity\_annotations.py: code returns the labels of the data 1) as rumour and non rumour 2) as veracity is true, false, unverified.
2. dataextraction.py: code utilised to extract data from Twitter with search queries on M&A. It has been ran on the terminal. (M&A dataset)
3. tweethydration.py: code used the tweet\_ids from the M&A dataset to obtain all the information required related to users and tweet information.
4. get\_text\_features.py: code has been used to extract linguistic features from the tweets on both the PHEME and M&A datasets.
5. The rest of the feature selection has been done over the excel spreadsheets.
6. model\_gbm.py: modelling of the 2-label classifier. It reads modellingdata.csv and scoring\_data.csv
7. model\_3labels\_gbm.py: modelling of the 3-label classifier.

Both (7) and (8) have been trained and tested on the PHEME dataset. In the codes there are included performance metrics calculations. In addition, there is a section of the predictive labelling of the M&A dataset.

1. FTSEanalysis.py: Code includes Granger Causality tests to determine any predictive relationship between the proportion of Fake News on Mergers and Acquisitions in the period and the close price and % volume of FTSE100. In addition, there are calculations of Pearson Correlation Coefficient.